AISM (2006) 58: 235–259
DOI 10.1007/s10463-005-0015-7

M.N.M. van Lieshout

A J-function for marked point patterns

Received: 5 July 2004 / Revised: 7 February 2005 / Published online: 30 May 2006

Abstract We propose a new summary statistic for marked point patterns. The underlying principle is to compare the distance from a marked point to the nearest other marked point in the pattern to the same distance seen from an arbitrary point in space. Information about the range of interaction can be inferred, and the statistic is well-behaved under random mark allocation. We develop a range of Hanisch style kernel estimators to tackle the problems of exploding tail variance earlier associated with J-function plug-in estimators, and carry out an exploratory analysis of a forestry data set.

Keywords Empty space function · J-function · Marked point pattern · Mark correlation function · Nearest neighbour distance distribution function · Product density · Random labelling · Reduced second moment measure · Spatial interaction · Spatial statistics

1 Introduction

Marked point patterns are spatial point configurations with a mark attached to each point (Stoyan and Stoyan, 1994). The points could represent the locations in (Euclidean) space of objects, while the marks capture additional information. The latter could be a type label, in which case we also speak of a multivariate point pattern (Cox and Lewis, 1972), a continuous measurement or shape descriptor, or a combination of these.

The statistical analysis of such a pattern in general begins by plotting a few summary statistics. Which statistic is used depends on taste and the type of mark.

M.N.M. van Lieshout
Department PNA4, Centrum voor Wiskunde en Informatica,
P.O. Box 94079, 1090 GB, Amsterdam, The Netherlands
E-mail: Marie-Colettte.van.Lieshout@cwi.nl

For discrete marks, cross versions of the classic nearest neighbour distance distribution function G (Diggle, 1983) or the second order K-function (Ripley, 1988) are popular. Alternatively, multivariate J-functions (Van Lieshout and Baddeley, 1999) may be used, or the mark correlation functions advocated by Penttinen and Stoyan (1989) and Stoyan and Stoyan (1994). For real-valued marks, mark correlation functions are typically used, at least if the marks are not binned so as to reduce the situation to the case of discrete marks. It is often a good idea to plot a range of summary statistics, as they tend to capture different aspects of the pattern and thus provide complementary information.

In this paper, we focus on generalisations of the J-function introduced by Van Lieshout and Baddeley (1996) to point processes with real-valued marks. The underlying idea of this summary statistic is to compare distances seen from an arbitrary fixed point 0 to the nearest point of the pattern—measured by the empty space function F—to those seen from a typical point of the pattern, as captured by the nearest neighbour distance function G. Thus, for (unmarked) point processes,

$$J(t) = \frac{1 - G(t)}{1 - F(t)}$$

defined for all $t \geq 0$ for which $F(t) \neq 1$. For a Poisson process, $J \equiv 1$; values $J(t) > 1$ indicate repulsion at range t, for clustered patterns the J-values tend to be less than 1.

The power of the J-function in hypothesis testing was assessed in Baddeley et al. (2000), Chen et al. (2001), and Thőnnes and Van Lieshout (1999). Extensions to multivariate point processes were proposed by Van Lieshout and Baddeley (1999), to germ grain models by Last and Holtmann (1999), and window based J-functions were suggested by Baddeley et al. (2000) and Chen et al. (2001). For applications in agriculture, astronomy, forestry and geology, see Foxall and Baddeley (2002), Kerscher (1998), Kerscher et al. (1998), Kerscher et al. (1999), Stein et al. (2001), or the recent theses of Chen (2003) and Paulo (2002).

The plan of this paper is as follows. In Sect. 2 we recall fundamental concepts from Palm theory, including the Nguyen–Zessin formula. In Sect. 3 we define a J-function for stationary marked point processes and discuss its behaviour under a Poisson or random labelling assumption. The next section concerns representation theorems for the J-function in terms of the conditional intensity and product densities, and explains the relationship with the reduced second moment measure. We turn to estimation in Sect. 5, and discuss Hanisch style ratio-estimators for the marked J-function, as well as kernel estimators for its derivative statistics. An application of our statistic to a forestry data set of pine saplings is the topic of Sect. 6. The paper is concluded with a summary and discussion.

2 Preliminaries and notation

In this section, we recall some fundamental concepts from Palm theory. Further details can be found for example in the textbooks by Daley and Vere-Jones (1988); or Stoyan et al. (1987).

Throughout this paper, let Y be a stationary marked point process on $\mathbb{R}^d$ with marks in a complete, separable metric space M (typically a subset of an Euclidean space) and write P for its distribution. We assume that the (first order) moment measure, denoted by μ_1, exists and is locally finite. Since Y is stationary, $\mu_1 = \lambda\ell \times \nu_M$, where λ is the intensity (assumed to be strictly positive to exclude trivial cases), ℓ is Lebesgue measure, and ν_M the mark distribution. If μ_1 exists, so does a reduced Palm distribution $P^{!(x,m)}(F)$, where $x \in \mathbb{R}^d$, $m \in M$, and $F \in \mathcal{N}$, the σ-algebra generated by the requirement that for all bounded Borel sets the number of points with marks in Borel subsets of M is a finite random variable. For any Borel set $B \subseteq M$ for which $\nu_M(B) > 0$, define a reduced Palm distribution $P_B^{!x}$ of Y with respect to B at the point $x \in \mathbb{R}^d$ by

$$\nu_M(B)\, P_B^{!x}(F) := \int_B P^{!(x,m)}(F)\, \mathrm{d}\nu_M(m). \tag{1}$$

The stationarity of Y implies that $P_B^{!x}(T_x F)$ takes the same value for ℓ-almost all x, where $T_x F = \{\mathbf{y}+x : \mathbf{y} \in F\}$ is the translation of F over x. In fact, we can choose the reduced Palm distributions to be translates of a single probability distribution, denoted by $P_B^{!0}(\cdot)$.

We shall use the notation $\lambda((x, m); \mathbf{y})$ for a Papangelou conditional intensity at $(x, m) \in \mathbb{R}^d \times M$ given the locally finite marked point pattern $\mathbf{y}$ elsewhere—provided it exists—and write $\lambda_B(x; \mathbf{y}) = \int_B \lambda((x, m); \mathbf{y})\, \mathrm{d}\nu_M(m)$ for its integral over a Borel set $B \subseteq M$. By stationarity, $\lambda_B(x; Y + x)$ is a.s. constant in x except on a null set, and we shall use the notation $\lambda_B(0, Y)$ as we did for the reduced Palm distribution.

Under the above assumptions, the following basic formulae

$$\mathbb{E}\left[\sum_{y \in Y} g(y, Y \setminus y)\right] = \lambda \int\int \mathbb{E}^{!(0,m)}\left[g((x, m), Y + x)\right]\, \mathrm{d}\ell(x)\, \mathrm{d}\nu_M(m) \tag{2}$$

$$= \int\int \mathbb{E}\left[g((x, m), Y)\, \lambda((x, m); Y)\right]\, \mathrm{d}\ell(x)\, \mathrm{d}\nu_M(m) \tag{3}$$

hold for all non-negative, measurable functions g (in the sense that the left hand side is finite if and only if the right hand side is). The first equation expresses the Campbell–Mecke formula [see e.g. Daley and Vere-Jones (1988), Proposition 12.1.IV] for stationary marked point processes, see equation (4.4.10) in Stoyan et al. (1987). The last equation due to Nguyen and Zessin (1979) requires the existence of a conditional intensity. In particular, for any Borel mark set B for which $\nu_M(B) > 0$,

$$\mathbb{E}\left[g(Y)\, \lambda_B(0; Y)\right] = \lambda\, \nu_M(B)\, \mathbb{E}_B^{!0}\left[g(Y)\right]. \tag{4}$$

Second and higher order behaviour of Y may be captured by the n-th order factorial moment measures $\mu^{(n)}$ and their densities with respect to the n-fold product measure of $\ell \times \nu_M$ with itself, the product densities denoted by $\rho^{(n)}$. For $n = 1$, $\mu^{(1)} = \mu_1$. Like the moment measure, $\mu^{(n)}$ for $n \geq 2$ is not necessarily locally finite, nor guaranteed to have a density. If it has, $\rho^{(n)}$ is permutation invariant and defined by the integral equations

$$\mathbb{E}\left[\sum_{y_1,\dots,y_n\in Y}^{\neq} g(y_1,\dots,y_n)\right] = \int\cdots\int g(y_1,\dots,y_n)\,\rho^{(n)}(y_1,\dots,y_n)\,\mathrm{d}\ell\times\nu_M(y_1)\cdots\mathrm{d}\ell\times\nu_M(y_n) \quad (5)$$

for all non-negative, measurable functions $g \geq 0$. In the physics literature, the related n-point correlation functions are commonly used. They are defined recursively in terms of product densities as follows:

$$\begin{aligned} \xi_1 &\equiv 1; \\ \rho^{(n)}(y_1,\dots,y_n) &= \lambda^n \textstyle\sum_{k=1}^{n} \sum_{D_1,\dots,D_k} \xi_{n(D_1)}(\mathbf{y}_{D_1})\cdots\xi_{n(D_k)}(\mathbf{y}_{D_k}) \end{aligned} \quad (6)$$

where the last sum ranges over all $\{D_1,\dots,D_k \neq \emptyset\}$ partitions of $\{1,\dots,n\}$ in k non-empty, disjoint sets, and $\mathbf{y}_{D_j} = \{y_i : i \in D_j\}$ is the corresponding partition of marked points. For a stationary Poisson process, $\rho^{(n)} \equiv \lambda^n$, so that for $n > 1$, the ξ_n account for the excess due to n-tuples in comparison with the reference Poisson process. Thus, $\xi_n > 0$ suggests clustering, while $\xi_n < 0$ tends to correspond to n-th order inhibition.

To conclude this section, note that if $\mu^{(n)}$ exists, then so does an n-point mark distribution $M^{x_1,\dots,x_n}(B_1\times\cdots\times B_n)$ for $x_i \in \mathbb{R}^d$, Borel sets $B_i \subseteq M$, and $i = 1,\dots,n$. For further details, see the textbooks by Daley and Vere-Jones (1988) and Stoyan et al. (1987).

3 A *J*-function for marked point patterns

We begin this section with recalling two well-known summary statistics from spatial statistics.

The *empty space function* F of Y is the cumulative distribution function of the distance from an arbitrarily chosen origin to the nearest point of the process, that is

$$F(t) := P\{Y \cap (B(0,t)\times M) \neq \emptyset\}$$

for $t \geq 0$. Here we write $B(0,t)$ for the closed ball of radius t centred at 0. The *nearest neighbour distance distribution function* from a point with mark in B is defined by

$$G_B(t) := P_B^{!0}\{Y \cap (B(0,t)\times M) \neq \emptyset\}$$

for $t \geq 0$ and Borel sets $B \subseteq M$ of positive ν_M-mass, the cumulative distribution function of the distance from a typical point of the process with mark in B to the nearest other point of Y regardless of its mark.

The J-function compares F to G_B, as made precise in the following definition.

Definition 3.1 *Let B be a Borel subset of M with $\nu_M(B) > 0$. Then the J-function with respect to mark set B is given by*

$$J_B(t) = \frac{1 - G_B(t)}{1 - F(t)}$$

and defined for all $t \geq 0$ for which $F(t) < 1$.

To interpret $J_B(t)$, note that for an independently marked Poisson point process, Slivnyak's theorem implies that $G_B(t) = F(t)$ for all B and t, hence $J_B \equiv 1$. Values less than 1 occur when $G_B(t) > F(t)$, that is when nearest neighbour distances are smaller than distances from the origin. Intuitively, such cases suggest clustering. On the other hand, values larger than 1 occur when the empty spaces are small in comparison to the distance from a point with mark in B to its nearest neighbour, an indication of inhibition. Note though that Bedford and Van den Berg (1997) showed that a J-function that is 1 on its domain of definition does not imply that Y is a Poisson process.

If one recalls that $\nu_M(B)\, G_B^{!0}(t) = \nu_M(B)\, P_B^{!0}\{Y \cap (B(0,t) \times M) \neq \emptyset\}$ is the almost everywhere constant value of $\int_B P^{!(x,m)}\{Y \cap (B(x,t) \times M) \neq \emptyset\}\, \mathrm{d}\nu_M(m)$, the definition of $J_B(t)$ may be rewritten as a mixture

$$J_B(t) = \frac{1}{\nu_M(B)} \int_B \left[\frac{P^{!(x,m)}\{Y \cap (B(x,t) \times M) \neq \emptyset\}}{1 - F(t)} \right] \mathrm{d}\nu_M(m) \tag{7}$$

for ℓ-almost all $x \in \mathbb{R}^d$. Thus, $J_B(t)$ may be interpreted as an average over B of J-functions with respect to a point marked $m \in B$ at an arbitrarily chosen origin.

Definition 3.2 *The marked point process Y has the* random labelling property *if the marks of the points are conditionally i.i.d. given the point locations.*

For marked point processes with the random labelling property, the J-function is of a convenient form, as stated more precisely in the following result.

Proposition 3.1 *Let X be a stationary point process on $\mathbb{R}^d$ with finite positive intensity λ, randomly labelled with mark distribution ν_M, and write Y for the marked point process thus obtained. Then, for all $t \geq 0$ with $F(t) < 1$, the J-function of Y with respect to a Borel mark set $B \subseteq M$ with $\nu_M(B) > 0$ is given by $J_B(t) = J_X(t)$, the J-function of X.*

Proof One needs to prove that the nearest neighbour distance distribution function of X coincides with that of Y with respect to any mark set, i.e. that

$$1 - G_B(t) = 1 - G_X(t) \tag{8}$$

for all Borel sets $B \subseteq M$ with $\nu_M(B) > 0$. Here we use the notation G_X for the nearest neighbour distance distribution function of X, and shall use similar notations P_X and $\mathbb{E}_X$ for the distribution of X and its expectation below. To prove Eq. (8), fix such a B and let A be any bounded Borel set of positive d-dimensional volume. Consider the measurable function

$$g((x,m), Y) = \mathbf{1}_A(x)\, \mathbf{1}_B(m)\, \mathbf{1}\{Y \cap (B(x,t) \times M) \neq \emptyset\}.$$

Now, the number of marked points falling in A is finite almost surely, with a finite expectation. Hence the expected sum of g over the points in Y is finite and

$$\begin{aligned}
\mathbb{E}\left[\sum_{(x,m(x))=y\in Y} g(y, Y\setminus\{y\})\right] &= \mathbb{E}\left[\sum_{x\in X\cap A} \mathbf{1}\{(X\setminus\{x\})\cap B(x,t)\neq\emptyset\}\,\mathbf{1}_B(m(x))\right]\\
&= \nu_M(B)\,\mathbb{E}_X\left[\sum_{x\in X\cap A} \mathbf{1}\{(X\setminus\{x\})\cap B(x,t)\neq\emptyset\}\right]\\
&= \lambda\,\nu_M(B)\int_A P_X^{!0}\{(X+x)\cap B(x,t)\neq\emptyset\}\,\mathrm{d}\ell(x)\\
&= \lambda\,\nu_M(B)\,\ell(A)\,G_X(t)
\end{aligned}$$

because of the conditionally independent mark assignments and the Campbell–Mecke formula for X.

On the other hand, since Y is stationary with $\mu_1 = \lambda\,\ell\times\nu_M$, where λ is the intensity of X, by the Campbell–Mecke formula (Eq. 2)

$$\begin{aligned}
\mathbb{E}\left[\sum_{y\in Y} g(y, Y\setminus\{y\})\right] &= \lambda\int_A\int_B P^{!(0,m)}\{(Y+x)\cap(B(x,t)\times M)\neq\emptyset\}\mathrm{d}\ell(x)\mathrm{d}\nu_M(m)\\
&= \lambda\,\nu_M(B)\,\ell(A)\,G_B(t).
\end{aligned}$$

We conclude that $G_X(t) = G_B(t)$, and the desired result follows upon division by $1 - F(t)$ on both sides of the Eq. (8). □

4 Representation theorems

In this section, relationships between the J-function and fundamental marked point process descriptors, namely the Papangelou conditional intensity and n-point correlation functions, are explored. We shall obtain a connection with the widely used second order K-function, and prove that the J-function with respect to any mark set becomes flat beyond the joint range of interaction.

4.1 Representation in terms of Papangelou conditional intensity

The Nguyen–Zessin formula motivated the definition of the J-function, as it relates expectations under the reduced Palm distribution to those under the distribution of Y itself. Hence it should not come as a surprise that the J-function with respect to a mark set can be expressed explicitly in terms of conditional intensities.

Proposition 4.1 *Let Y be a stationary marked point process with finite positive intensity λ for which a regular version of the conditional intensity exists that satisfies the Nguyen–Zessin formula (Eq. 4), and B a Borel set in the mark space M with $\nu_M(B) > 0$. Then $G_B(t) < 1$ for some $t \geq 0$ implies $F(t) < 1$ and*

$$J_B(t) = \mathbb{E}\left[\frac{\lambda_B(0;Y)}{\lambda\,\nu_M(B)}\,\middle|\, Y\cap(B(0,t)\times M)=\emptyset\right] \tag{9}$$

$$= \left(\mathbb{E}_B^{!0}\left[\frac{\lambda\nu_M(B)}{\lambda_B(0;Y)}\,\middle|\, Y\cap(B(0,t)\times M)=\emptyset\right]\right)^{-1}. \tag{10}$$

Proof Let A be the event $\{Y \cap (B(0,t) \times M) = \emptyset\}$, so that $F(t) = 1$ if and only if $P(A) = 0$. Apply the Nguyen–Zessin formula (Eq. 4) for the measurable function $g(Y) = \mathbf{1}_A(Y) \geq 0$ to obtain

$$\mathbb{E}\,[\mathbf{1}_A(Y)\,\lambda_B(0;Y)] = \lambda\,\nu_M(B)\,\mathbb{E}_B^{!0}\,[\mathbf{1}_A(Y)] = \lambda\,\nu_M(B)\;(1 - G_B(t))\,.$$

If $P(A) = 0$, so is the expectation in the left hand side of the above equality. Since by assumption $\lambda\,\nu_M(B) > 0$, necessarily $G_B(t) = 1$. Hence $G_B(t) < 1$ implies $P(A) > 0$, or, equivalently, $F(t) < 1$. Thus, we may divide the left and right hand side by $\lambda\,\nu_M(B)\;(1 - F(t))$ to obtain Eq. (9).

Next, apply the Nguyen–Zessin formula to the function $g(Y) = \mathbf{1}_A(Y)/\lambda_B(0;Y)$ (we shall show below that the function is well-defined). Then, from

$$\lambda\,\nu_M(B)\,\mathbb{E}_B^{!0}\,[\mathbf{1}_A(Y)/\lambda_B(0;Y)] = \mathbb{E}\,[\mathbf{1}_A(Y)] = 1 - F(t),$$

we obtain Eq. (10) upon dividing both sides of the equation by $1 - G_B(t)$. Note that $\lambda_B(0;Y) > 0$ almost surely with respect to the reduced Palm distribution $P_B^{!0}$, since

$$\lambda\,\nu_M(B)\;P_B^{!0}\{\lambda_B(0;Y) = 0\} = \mathbb{E}\,[\mathbf{1}\{\lambda_B(0;Y) = 0\}\,\lambda_B(0;Y)] = 0,$$

hence the conditional expectation in Eq. (10) is well-defined. It follows that

$$\begin{aligned}\lambda\,\nu_M(B) &= \lambda\,\nu_M(B)\,\mathbb{E}_B^{!0}\,[\mathbf{1}\{\lambda_B(0;Y) > 0\}] = \mathbb{E}\,[\mathbf{1}\{\lambda_B(0;Y) > 0\}\,\lambda_B(0;Y)]\\ &\leq \mathbb{E}\,[\lambda_B(0;Y)] = \lambda\,\nu_M(B).\end{aligned}$$

Hence

$$\mathbf{1}\{\lambda_B(0;Y) > 0\}\,\lambda_B(0;Y) = \lambda_B(0;Y) \quad P-\text{a.s.},$$

or, in other words, $P\{\lambda_B(0;Y) = 0\} = 0$, and the function $g(Y)$ is well-defined. □

The following corollary gives a useful interpretation of the J-statistic.

Corollary 4.1 *Let Y be a stationary marked point process with finite positive intensity λ for which a regular version of the conditional intensity exists that satisfies the Nguyen–Zessin formula* (*Eq. 4*)*. Then $J_B(t) \geq 1$* (*respectively ≤ 1*) *if and only if*

$$\mathrm{Cov}\,(\lambda_B(0;Y), \mathbf{1}\{Y \cap (B(0,t) \times M) = \emptyset\}) \geq 0$$

(*respectively is non-positive*).

Another corollary states that $J_B(t)$ is constant beyond the joint range of interaction.

Definition 4.1 *A marked point process Y has* joint interaction range *s if for all Borel sets $B \subseteq M$ with $\nu_M(B) > 0$ its conditional intensity $\lambda_B(0;Y)$ is constant for all realisations which contain no points in $B(0,s)$.*

Thus, for t greater than the joint interaction range, given that $Y \cap (B(0,t) \times M) = \emptyset$, the conditional intensity $\lambda_B(0;Y) = \lambda_B(0;\emptyset)$.

Corollary 4.2 *If Y has joint interaction range s, $0 < s < \infty$, then $J_B(t)$ is constant for all $t \geq s$ for which it is defined, and*

$$J_B(t) \equiv \frac{\lambda_B(0; \emptyset)}{\lambda\, \nu_M(B)}.$$

Note that neither F nor G_B are constant beyond the joint interaction range. For example, for a stationary planar Poisson process, $\lambda_B(0; Y) \equiv \lambda \nu_M(B)$ does not depend on Y. Nevertheless, $F(t) = G_B(t) = 1 - \exp\left[-\lambda \pi t^2\right]$ is strictly increasing and depends on the intensity, a fact that makes interpretation of its graph harder than that of $J_B(t)$.

A widely used family of marked point process models with finite joint interaction range is that of the pairwise interaction models. Such models have a conditional intensity of the form

$$\lambda((u, m); \mathbf{y}) = \beta(m) \prod_{y_j \in \mathbf{y}} \gamma((u, m), y_j)$$

for $(u, m) \notin \mathbf{y}$, where β and γ are non-negative measurable functions (cf. Baddeley and Møller, 1989; Ripley and Kelly, 1977; Ripley, 1989 or Ogata and Tanemura, 1989). Note that these authors consider marked point processes in a bounded domain, but that, under conditions similar to those for existence of a conditional intensity [cf. Sect. 5.5.3 in Stoyan et al. (1987)], a stationary extension on $\mathbb{R}^d$ can be shown to exist. Typically β is bounded, and $\gamma \leq 1$. It is easily verified that

$$J_B(t) = \frac{1}{\lambda\, \nu_M(B)} \int_B \beta(m)\, \mathbb{E}\left[\prod_{y \in Y} \gamma((0, m), y) \,\middle|\, Y \cap (B(0, t) \times M) = \emptyset \right] \mathrm{d}\nu_M(m)$$

wherever defined. If $\gamma((u, m), (v, n)) \equiv 1$ for $||u - v|| > s$, the J-function with respect to B reduces to

$$J_B(t) = \frac{1}{\lambda\, \nu_M(B)} \int_B \beta(m)\, \mathrm{d}\nu_M(m)$$

for $t \geq s$ such that $F(t) < 1$.

4.2 Representation in terms of product densities

In astronomical folklore, the connection between the empty space function and the product densities, or equivalently the n-point correlation functions, is well-known. Indeed, White (1979) argues that the fact that F uses product densities of all orders makes it particularly appropriate to detect clustering in galaxy catalogues. Here, we shall give an expression of the J-function for marked point processes in terms of n-point correlation functions, and consider in detail how to obtain a classic second order analysis (Stoyan and Stoyan, 1994) by truncation.

We shall need the following lemma.

Lemma 4.1 (*White* (1979))
Let Y be a stationary marked point process. Suppose that all order factorial moment measures exist as locally finite measures, and have a Radon–Nikodym derivative $\rho^{(n)}$ with respect to the n-fold product of $\ell \times \nu_M$ with itself, $n \in \mathbb{N}$. Then the empty space function can be written as

$$F(t) = -\sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_{B(0,t)\times M} \cdots \int_{B(0,t)\times M} \rho^{(n)}(y_1, \dots, y_n)\, \mathrm{d}\ell \times \nu_M(y_1) \cdots \mathrm{d}\ell \times \nu_M(y_n)$$
$$= 1 - \exp\left[\sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!} \int_{B(0,t)\times M} \cdots \int_{B(0,t)\times M} \xi_n(y_1, \dots, y_n)\, \mathrm{d}\ell \times \nu_M(y_1) \cdots \mathrm{d}\ell \times \nu_M(y_n)\right]$$

where the ξ_n are given by Eq. (6).

The main result of this subsection is the following.

Proposition 4.2 *Let Y be a stationary marked point process. Suppose that all order factorial moment measures exist as locally finite measures, and have a Radon–Nikodym derivative $\rho^{(n)}$ with respect to the n-fold product of $\ell \times \nu_M$ with itself, $n \in \mathbb{N}$. Then the J-function with respect to any Borel mark set $B \subseteq M$ with $\nu_M(B) > 0$ can be written as*

$$J_B(t) = \frac{1}{\nu_M(B)} \left[\nu_M(B) + \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!} J_n^B(t)\right]$$

for all $t \geq 0$ for which $F(t) < 1$, where $J_n^B(t)$ is the common value of $\int_B \int_{B(0,t)\times M} \cdots \int_{B(0,t)\times M} \xi_{n+1}((a, m), y_1 + a, \dots, y_n + a)\, \mathrm{d}\nu_M(m)\, \mathrm{d}\ell \times \nu_M(y_1) \cdots \mathrm{d}\ell \times \nu_M(y_n)$ for ℓ-almost all $a \in \mathbb{R}^d$.

Proof By the Campbell–Mecke formula (Eq. 2), for any bounded Borel set $A \subset \mathbb{R}^d$ of strictly positive Lebesgue measure $\ell(A) > 0$,

$$\lambda\, \nu_M(B)\, \ell(A)\, (1 - G_B(t)) = \mathbb{E}\left[\sum_{(a,m)\in Y\cap(A\times B)} \mathbf{1}\{Y \setminus \{(a, m)\} \cap (B(a, t) \times M) = \emptyset\}\right].$$

The expectation on the right hand side is well-defined and finite, since the first order (factorial) moment measure exists as a locally finite measure. Now, by the inclusion-exclusion formula, the expectation may be rewritten as

$$\mathbb{E}\left[\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\sum_{(a,m),y_1,\ldots,y_n\in Y}^{\neq}\mathbf{1}_{A\times B}(a,m)\,\mathbf{1}_{B(a,t)\times M}(y_1)\cdots\mathbf{1}_{B(a,t)\times M}(y_n)\right]$$

which by Eq. (5) is equal to $\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\int_A f_n^B(a,t)\,\mathrm{d}\ell(a)$ with

$$f_n^B(a,t)=\int_B\int_{B(0,t)\times M}\cdots\int_{B(0,t)\times M}\rho^{(n+1)}((a,m),y_1+a,\ldots,y_n+a)\mathrm{d}\nu_M(m)$$
$$\times\mathrm{d}\ell\times\nu_M(y_1)\cdots\mathrm{d}\ell\times\nu_M(y_n).$$

By stationarity, the factorial moment measures are translation invariant, hence for all $n\in\mathbb{N}$, $f_n^B(\cdot,t)$ is almost everywhere constant, say $f_n^B(t)$. Of course, $f_0^B\equiv\lambda\nu_M(B)$. From the recursion relation (Eq. 7), it follows that $J_n^B(t)$ is well-defined too, with $J_0^B\equiv\nu_M(B)$. In summary

$$\lambda\,\nu_M(B)\,(1-G_B(t))=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}f_n^B(t). \tag{11}$$

In order to complete the proof, we need to show that the right hand side of Eq. (11) is equal to $\lambda\,(1-F(t))\sum_{n=0}^{\infty}\frac{(-\lambda)^n}{n!}J_n^B(t)$. To do so, use the definition of the correlation functions, and split the partition into those terms that contain the first marked point and those that do not. More precisely, Eq. (11) equals

$$\lambda\nu_M(B)+\lambda\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n!}\sum_{D\subseteq\{1,\ldots,n\}}J^B_{n(D)}(t)\sum_{k=1}^{n-n(D)}\sum_{\substack{D_1,\ldots,D_k\neq\emptyset\\ \cup D_j=\{1,\ldots,n\}\setminus D}}I_{n(D_1)}\cdots I_{n(D_k)}$$
$$=\left[\lambda\,J_0^B(t)+\lambda\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n!}J_n^B(t)\right]$$
$$\times\left[1+\sum_{m=1}^{\infty}\frac{(-\lambda)^m}{m!}\sum_{k=1}^{m}\sum_{\substack{D_1,\ldots,D_k\neq\emptyset\\ \cup D_j=\{1,\ldots,m\}}}I_{n(D_1)}\cdots I_{n(D_k)}\right]$$

where I_n is the n-fold integral over the Cartesian product of $B(0,t)\times M$ of the n-point correlation functions with respect to the appropriate $\ell\times\nu_M$ product measure. By convention, multiplication by an empty sum is set to 1. Note that the second term is $1-F(t)$, by White's lemma, and the desired representation holds. □

Proposition 4.2 requires the existence of factorial moment measures of all orders. If such an assumption is not valid, one could truncate the alternating series at some fixed n. Indeed, the approximation in terms of product densities up to second order reads

$$J_B(t) - 1 \approx \frac{-\lambda}{\nu_M(B)} \int_B \left[\int_{B(0,t)\times M} \xi_2((0,m),y)\, \mathrm{d}\ell \times \nu_M(y) \right] \mathrm{d}\nu_M(m)$$
$$= \frac{-1}{\lambda\, \nu_M(B)} \int_B \left[\int_{B(0,t)\times M} \left(\rho^{(2)}((0,m),y) - \lambda^2\right) \mathrm{d}\ell \times \nu_M(y) \right] \mathrm{d}\nu_M(m).$$

Thus,

$$J_B(t) - 1 \approx -\lambda \left[\mathcal{K}_B(t) - \ell(B(0,t))\right] \tag{12}$$

where $\mathcal{K}_B$ is the reduced second moment measure (see e.g. the textbooks by Ripley, 1988 and Stoyan and Stoyan, 1994) with respect to the mark set B. In other words, $\lambda\mathcal{K}_B(t)$ is the expected number of further points within a radius t of a typical point with mark in B (i.e. under $P_B^{!0}$). In terms of the 2-point mark distribution function, Eq. (12) reads

$$J_B(t) - 1 \approx -\lambda \int_{B(0,t)} \left[\frac{M^{0,x}(B\times M)}{\nu_M(B)}\, g(0,x) - 1 \right] \mathrm{d}\ell(x)$$
$$= -\lambda \int_{B(0,t)} \left[k_f(0,x)\, g(0,x) - 1\right] \mathrm{d}\ell(x)$$

where g is the pair correlation function of the unmarked point process associated with Y, which is proportional by a factor $1/\lambda^2$ to the second order product density, and the functional k_f is given by $k_f(0,x) = \int_M \int_M f(m_1,m_2)\mathrm{d}M^{0,x}(m_1,m_2)$ with $f(m_1,m_2) = \mathbf{1}_B(m_1)/\nu_M(B)$.

A typical second order analysis of marked point processes (Penttinen and Stoyan, 1989; Stoyan and Stoyan, 1994) plots estimates of the pair correlation function and $k_f(\cdot,\cdot)$ for a suitably chosen non-negative, Borel measurable and integrable function f. For a single positive real-valued mark, the function $f(m_1,m_2) = m_1 m_2/\mu^2$ may be used, where $\mu = \int_{\mathbb{R}_+} m \mathrm{d}\nu_M(m)$ is the mean mark. For discrete or binned labels, $f(m_1,m_2) = \mathbf{1}_A(m_1)\, 1_B(m_2)\, /(\nu_M(A)\, \nu_M(B))$ is a convenient choice. The latter amounts to a cross J-function analysis (see Sect. 7 or Van Lieshout and Baddeley, 1999) with truncation at second order product densities. At this point, it should be emphasised that, in spite of its popularity, the second order structure does not provide a full description of the spatial interaction structure (Baddeley and Silverman, 1984). Thus, it is wise to plot a range of summary statistics when exploring spatial data.

Example 4.1 Let X be a stationary, isotropic, planar point process for which a second order product density exists. Then

$$J(t) \approx 1 - \lambda \int_{B(0,t)} (g(||x||) - 1)\, \mathrm{d}\ell(x) = 1 - 2\pi\lambda \int_0^t r\, (g(r) - 1)\, \mathrm{d}r,$$

which is known as the 'Gaussian approximation' in astronomy (Kerscher, 1998).

5 Estimation

Throughout this section, assume that Y is a stationary marked point process on $\mathbb{R}^d$ with marks in a complete, separable metric space M with finite, positive intensity λ. Thus, the first order moment measure and hence Palm kernels exist. Below, we propose Hanisch style kernel estimators for the J-function and some associated characteristics when Y is observed within a compact set $W \subseteq \mathbb{R}^d$ of positive volume $\ell(W)$. We shall rely on the principle that Palm characteristics may be estimated by averages over points of the marked point process (Stoyan et al. 1987, p. 130), i.e. the estimator

$$\lambda \widehat{\nu_M(B)\mathbb{E}_B^{!0}} f(Y) = \sum_{y=(x,m)\in Y} \frac{\mathbf{1}_W(x)\,\mathbf{1}_B(m)\, f((Y-x)\setminus\{(0,m)\})}{\ell(W)}$$

is unbiased for any non-negative measurable function f by virtue of the Campbell–Mecke formula (Eq. 2). For example, for the mark distribution we have

$$\lambda\,\widehat{\nu_M(B)} = \sum_{(x,m)\in Y} \frac{\mathbf{1}_W(x)\,\mathbf{1}_B(m)}{\ell(W)},$$

for any Borel subset B of M and any compact set $W \subseteq \mathbb{R}^d$ of positive volume $\ell(W) > 0$. For many f, care has to be taken with regard to edge effects caused by the fact that not Y itself is observed, but rather $Y \cap (W \times M)$. For such f, the estimator described above cannot be computed based on the available data. This phenomenon known as the 'edge effect' is particularly irksome for irregularly shaped windows W, and in higher dimensions.

In the approach of Hanisch (1984) for functions based on inter-point distances such as the J-function, a solution for the edge effect problem lies in the observation that on the event $\{d(x,\partial W) \geq d(x,Y)\}$ the observed distance $d(x, Y \cap (W \times M))$ is equal to the true one $d(x,Y)$. Other types of edge-corrected estimators for F and G are reviewed in many textbooks, including Cressie (1991, chapter 8), Ripley (1988, chapter 3), Stoyan et al. (1987, pp. 122–131), as well as in Baddeley and Gill (1997). We have chosen the Hanisch approach, as it leads to estimators that—in contrast to other estimators—respect the monotonicity and continuity properties of F and G_B, and do not discard too much data. The plug-in principle then yields ratio-unbiased estimators for J_B. Unfortunately though, empirical evidence (see e.g. Van Lieshout and Baddeley, 1996, 1999) suggests that the variance of $\widehat{J_B(t)}$ will increase with t, causing a rather fluctuating tail behaviour of $\widehat{J_B(t)}$. To solve this problem, we propose to combine edge correction with smoothing (Silverman, 1986) to obtain more robust estimators.

5.1 Cumulative distribution functions

The first result of this section concerns an edge-corrected estimator for the nearest neighbour distance distribution function with respect to a mark set.

Proposition 5.1 *For any Borel set $B \subseteq M$ with $\nu_M(B) > 0$, define for $t \geq 0$ such that $\ell(W^{\ominus t}) > 0$,*

$$\sum_{y_k \in Y \cap (W \times B)} \left[\frac{\mathbf{1}\{b_k > s_k\}\, \mathbf{1}\{s_k \leq t\}}{\ell(W^{\ominus s_k})} \right] \tag{13}$$

with the convention that $0/0 = 0$. Here, $y_k = (x_k, m_k)$ is a marked point, and $s_k = d(x_k, (Y \cap (W \times M)) \setminus \{(x_k, m_k)\})$, respectively, $b_k = d(x_k, \partial W)$ are the Euclidean distances from x_k to the nearest other marked point in Y and to the boundary of W. The notation $W^{\ominus t}$ is used for the set $\{w \in W : d(w, \partial W) > t\}$. Then Eq. (13) is an unbiased estimator of $\lambda\, \nu_M(B)\, G_B(t)$ for all $t \geq 0$ for which is defined; it is a non-negative function that is increasing in t.

Proof First, note that whenever $b_k > s_k$, $\ell(W^{\ominus s_k}) > 0$, hence Eq. (13) is well-defined for any $t \geq 0$. Now, $\mathbf{1}\{d(x_k, \partial W) > d(x_k, (Y \cap (W \times M)) \setminus \{(x_k, m_k)\})\} = \mathbf{1}\{d(x_k, \partial W) > d(x_k, Y \setminus \{(x_k, m_k)\})\}$, and, moreover, when the indicator functions take the value 1, $s_k = d(x_k, Y \setminus \{(x_k, m_k)\})$. Hence the s_k in Eq. (13) may be replaced by $d(x_k, Y \setminus \{(x_k, m_k)\})$.

The mapping $(x, Y) \mapsto d(x, Y)$ on $\mathbb{R}^d \times N$, the product space of $\mathbb{R}^d$ and the configuration space N of locally finite marked point patterns $\mathbf{y}$ in $\mathbb{R}^d \times M$ is jointly measurable when restricted to W, and non-negative. Therefore,

$$\mathbb{E}\left[\sum_{(x,m) \in Y \cap (W \times B)} \frac{\mathbf{1}\{x \in W^{\ominus d(x, Y \setminus \{(x,m)\})}\}\, \mathbf{1}\{d(x, Y \setminus \{(x,m)\}) \leq t\}}{\ell(W^{\ominus d(x, Y \setminus \{(x,m)\})})} \right]$$
$$= \lambda\, \nu_M(B) \int_W \mathbb{E}_B^{!0} \left[\frac{\mathbf{1}\{x \in W^{\ominus d(0,Y)}\}}{\ell(W^{\ominus d(0,Y)})} \mathbf{1}\{d(0, Y) \leq t\} \right] \mathrm{d}\ell(x)$$

by stationarity and the Campbell–Mecke formula (Eq. 2). Since $\ell(W^{\ominus d(0,Y)}) > 0$ on the event $\{d(0, Y) \leq t\}$, an application of Fubini's theorem yields that the expectation of Eq. (13) reduces to

$$\lambda\, \nu_M(B)\, \mathbb{E}_B^{!0} [\mathbf{1}\{d(0, Y) \leq t\}] = \lambda\, \nu_M(B)\, G_B(t).$$

Clearly, Eq. (13) is non-negative and increasing in t in the range for which it is well-defined. □

For the rectangular or circular windows that are typically encountered in spatial statistics, the term $\ell(W^{\ominus s})$ can be evaluated explicitly.

The practitioner usually is interested in estimators of $G_B(t)$ rather than in those of $\lambda\, \nu_M(B)\, G_B(t)$. In the spirit of Stoyan and Stoyan (2000), who advocated to use intensity estimators similar to estimators of the numerator, we choose to divide Eq. (13) by

$$\sum_{y_k \in Y \cap (W \times B)} \left[\frac{\mathbf{1}\{b_k > s_k\}}{\ell(W^{\ominus s_k})} \right], \tag{14}$$

again with $0/0 = 0$. Note that Eq. (14) is well-defined with expectation $\lambda\, \nu_M(B)$ $G_B(T^-)$, where $T := \sup\{t \geq 0 : W^{\ominus t} \neq \emptyset\} = \sup\{t \geq 0 : \ell(W^{\ominus t}) > 0\}$.

The Hanisch estimator for the empty space function is well known (Chiu and Stoyan, 1998): for $t \leq T := \max\{s : \#\{i : b_i \geq s\} > 0\} = \max\{b_k : x_k \in L\}$, set

$$\widehat{F(t)} = \sum_{x_k \in L} \left[\frac{\mathbf{1}\{b_k \geq s_k\}\, \mathbf{1}\{s_k \leq t\}}{\#\{i : b_i \geq s_k\}} \right] \tag{15}$$

with the convention that $0/0 = 0$. The sum is over a finite lattice $L \neq \emptyset$ in W. It should be noted that there is no need to delete x_k from the point pattern in the computation of s_k, as it will almost surely be no part of a realisation of Y. Clearly, $\widehat{F}(t)$ is increasing; it is also unbiased. Unfortunately, though, Eq. (15) is not necessarily a distribution function, and one may have to normalise by dividing by $\widehat{F(T)}$.

In a range of papers (Baddeley et al. 2000; Chen 2003; Chen et al. 2001), uncorrected estimators of J-functions were considered. These may be seen as unbiased for 'window averaged J-functions', and can be surprisingly powerful as test statistic. As is our J-function, for Poisson processes the windowed J-function is identically equal to 1, but in general explicit evaluation seems to be more cumbersome, no representation theorems have been found, and the behaviour under random labelling is unknown. The theoretical windowed values are typically closer to 1 than the classic ones, which may be understood as 'Poissonisation due to window averaging'.

5.2 Densities and hazard rates

For exploratory purposes, densities and hazard rates often convey more information than cumulative statistics (Baddeley and Gill, 1994; Stoyan and Stoyan, 1994). Indeed, suppose a Papangelou conditional intensity exists so that the Nguyen–Zessin formula (Eq. 4) holds. Then, provided $G_B(t) < 1$,

$$J'_B(t) = \left[h_F(t) - h_{G_B}(t) \right] J_B(t)$$

where h_I denotes the hazard rate of statistic I. Thus, the relative derivative statistic $J'_B(t)/J_B(t)$ is a signed measure of spatial association. The hazard rates exist, whence $J_B(t)$ is differentiable, under the Nguyen–Zessin condition. Indeed, by theorems from Baddeley and Gill (1997) and Hansen et al. (1996), the empty space function of a stationary point process X is absolutely continuous with density and hazard rate given by

$$f_F(t) = \frac{\mathbb{E} h_{d-1}(\partial(X \oplus B(0,t)) \cap Z)}{\ell(Z)}; \quad h_F(t) = \frac{\mathbb{E}\left[h_{d-1}(\partial(X \oplus B(0,t)) \cap Z) \right]}{\mathbb{E}\left[\ell(Z \setminus (X \oplus B(0,t))) \right]} \tag{16}$$

for any non-empty compact regular set Z (i.e. $\mathrm{cl}(Z^{\mathrm{int}}) = Z$). Here h_{d-1} is the $d-1$ dimensional Hausdorff measure, and $Y \oplus B(0,t) = \cup_{(x,m) \in Y} B(x,t)$ the Minkowski sum of Y and a closed ball of radius t centered at the origin. Chiu and Stoyan

(1998) observed that Eq. (15) is a discretisation of

$$\tilde{F}(t) = \int_W \left[\frac{\mathbf{1}\{x \in W \ominus B(0, d(x, Y))\}\, \mathbf{1}\{d(x, Y) \leq t\}}{\ell(W \ominus B(0, d(x, Y)))} \right] \mathrm{d}\ell(x)$$

$$= \int_0^t \frac{h_{d-1}(W \ominus B(0, s) \cap \partial(Y \oplus B(0, s)))}{\ell(W \ominus B(0, s))}\, \mathrm{d}s$$

for all t for which it is defined. Here $W \ominus B(0, s) = (W^c \oplus B(0, s))^c = \{w \in W : d(w, \partial W) \geq s\}$. Provided W is regular, the integrand is an unbiased estimator of $f(s)$ based on the minus sampling principle, an alternative interpretation of Eq. (15).

In general, densities and hazard rates do not exist for G_B. Indeed, the nearest neighbour distance distribution function may be degenerate, for instance for randomly translated grids. However, if the Nguyen–Zessin identity holds, a density does exist.

Proposition 5.2 *Let Y be a stationary marked point process with finite positive intensity λ for which a regular version of the conditional intensity exists that satisfies the Nguyen–Zessin formula* (*Eq. 4*)*, and B a Borel mark set with $\nu_M(B) > 0$. Then the nearest neighbour distance $d(0, Y)$ from a point at* 0 *with mark in B is absolutely continuous with density*

$$g_B(t) = f_F(t)\, \mathbb{E}\left[\frac{\lambda_B(0; Y)}{\lambda\, \nu_M(B)} \,\middle|\, d(0, Y) = t \right]. \tag{17}$$

Proof Suppose $P\{d(0, Y) \in A\} = 0$ for any Borel subset A of the positive half line, i.e. $\mathbf{1}\{d(0, Y) \in A\} = 0$ P-almost surely. Consequently, $\lambda\, \nu_M(B)\, P_B^{!0}\{d(0, Y) \in A\} = \mathbb{E}\,[\lambda_B(0; Y)\, \mathbf{1}\{d(0, Y) \in A\}] = 0$ by Eq. (4). Therefore, the nearest neighbour distance distribution with respect to mark set B is absolutely continuous with respect to the distribution of $d(0, Y)$, so, by the Radon–Nikodym theorem

$$P_B^{!0}\{d(0, Y) \in A\} = \int_A f_{GF}(a)\, \mathrm{d}F(a)$$

for some measurable, integrable function f_{GF} on the positive half line, and in particular

$$G_B(t) = \int_0^t f_{GF}(s)\, \mathrm{d}F(s) = \int_0^t f_{GF}(s)\, f_F(s)\, \mathrm{d}s.$$

The Nguyen–Zessin formula further implies that

$$\begin{aligned} \lambda \nu_M(B)\, G_B(t) &= \mathbb{E}\,[\lambda_B(0; Y)\, \mathbf{1}\{Y \cap (B(0, t) \times M) \neq \emptyset\}] \\ &= \mathbb{E}\mathbb{E}\,[\lambda_B(0; Y)\, \mathbf{1}\{d(0, Y) \leq t\} \,|\, d(0, Y)] \\ &= \int_0^t \mathbb{E}\,[\lambda_B(0; Y) \,|\, d(0, Y) = s]\; \mathrm{d}F(s). \end{aligned}$$

Hence, $\lambda\, \nu_M(B)\, f_{GF}(s) = \mathbb{E}\,[\,\lambda_B(0; Y) |\, d(0, Y) = s]$ for almost all s, and Eq. (17) follows. □

It is instructive to note that beyond the joint range of interaction

$$g_B(t) = f_F(t)\, \lambda_B(0; \emptyset)/(\lambda\, \nu_M(B)) = f_F(t)\, J_B(t)$$

so that

$$\begin{aligned}(h_F(t) - h_{G_B(t)}) J_B(t) &= \left[\frac{f_F(t)}{1 - F(t)} - \frac{f_F(t) J_B(t)}{1 - G_B(t)}\right] J_B(t) \\ &= \left[\frac{f_F(t)}{1 - F(t)} - \frac{f_F(t)}{1 - F(t)}\right] J_B(t)\end{aligned}$$

vanishes as it should.

In the following proposition, we derive Hanisch-style kernel estimators (Silverman, 1986) of $f_F(t)$ and $\lambda\, \nu_M(B)\, g_B(t)$. Note that as the domain of definition is $t \geq 0$, one might consider to use adaptations such as setting the value of the density at 0 for negative t (as below) or reflection (Silverman, 1986, p. 30).

Proposition 5.3 *Let Y be a stationary marked point process with finite positive intensity λ for which a regular version of the conditional intensity exists that satisfies the Nguyen–Zessin formula (Eq. 4), B be a Borel mark set with $\nu_M(B) > 0$, $L \neq \emptyset$ a finite set of points in W, and $t \in [0, T]$ with $T = \max\{s : \#\{i : b_i \geq s\} > 0\}$. Given a symmetric, measurable, non-negative kernel $k_h(\cdot)$ on $\mathbb{R}$ with bandwidth h that integrates to unity, i.e. $k_h(x) = k(x/h)/h$,*

$$\widehat{f_F(t)} = \sum_{x_k \in L} \frac{k_h(t - s_k)\, \mathbf{1}\{b_k \geq s_k\}}{\#\{i : b_i \geq s_k\}}$$

with the convention that $0/0 = 0$ is an unbiased estimator of $\int_0^T k_h(t - s) f_F(s) \mathrm{d}s$. As before, $s_k = d(x_k, Y \cap (W \times M))$; and $b_k = d(x_k, \partial W)$. Furthermore, for $t \in [0, T)$ with $T = \sup\{s : W^{\ominus s} \neq \emptyset\}$,

$$\lambda \widehat{\nu_M(B)} g_B(t) = \sum_{y_k \in Y \cap (W \times B)} \frac{k_h(t - s_k)\, \mathbf{1}\{b_k > s_k\}}{\ell(W^{\ominus s_k})} \tag{18}$$

(again with $0/0 = 0$) is an unbiased estimator of $\lambda\, \nu_M(B) \int_0^T k_h(t - s)\, g_B(s) \mathrm{d}s$. Here, $y_k = (x_k, m_k)$ and $s_k = d(x_k, (Y \cap (W \times M)) \setminus \{(x_k, m_k)\})$.

Proof First, note that whenever $b_k > s_k$, $\ell(W^{\ominus s_k}) > 0$, hence Eq. (18) is well-defined for any $t \geq 0$. Moreover,

$$\begin{aligned}&\mathbf{1}\{d(x_k, \partial W) > d(x_k, (Y \cap (W \times M)) \setminus \{(x_k, m_k)\})\} \\ &\quad = \mathbf{1}\{d(x_k, \partial W) > d(x_k, Y \setminus \{(x_k, m_k)\})\},\end{aligned}$$

and, when the indicator functions take the value 1, $s_k = d(x_k, Y \setminus \{(x_k, m_k)\})$. Hence distances to the nearest other point in $Y \cap (W \times M)$ may be replaced by distances to Y in the formulae below. A similar remark is true for $\widehat{f_F(t)}$. Now,

$$\mathbb{E}\widehat{f_F(t)} = \sum_{x_k \in L} \mathbb{E}\left[\frac{k_h(t - d(x_k, Y))\,\mathbf{1}\{d(x_k, \partial W) \geq d(x_k, Y)\}}{\#\{i : d(x_i, \partial W) \geq d(x_k, Y)\}}\right]$$
$$= \sum_{x_k \in L} \int_0^\infty \left[\frac{k_h(t-s)\,\mathbf{1}\{d(x_k, \partial W) \geq s\}}{\#\{i : d(x_i, \partial W) \geq s\}}\right] f_F(s)\,\mathrm{d}s$$
$$= \int_0^T k_h(t-s)\, f_F(s)\,\mathrm{d}s.$$

Regarding the nearest neighbour distance distribution density $g_B(t)$,

$$\mathbb{E}\lambda\widehat{\nu_M(B) g_B}(t) = \lambda \int_W \int_B \mathbb{E}^{!(0,m)}\left[\frac{k_h(t - d(0, Y))\,\mathbf{1}\{x \in W^{\ominus d(0,Y)}\}}{\ell(W^{\ominus d(0,Y)})}\right] \mathrm{d}\ell(x)\,\mathrm{d}\nu_M(m)$$
$$= \lambda\, \nu_M(B) \int_W \mathbb{E}_B^{!0}\left[\frac{k_h(t - d(0, Y))\,\mathbf{1}\{x \in W^{\ominus d(0,Y)}\}}{\ell(W^{\ominus d(0,Y)})}\right] \mathrm{d}\ell(x)$$
$$= \lambda\, \nu_M(B) \int_0^T k_h(t-s)\, g_B(s)\mathrm{d}s$$

by stationarity, the Campbell–Mecke formula and Fubini's theorem. □

One is often interested in an estimator of $g_B(t)$ itself. As for the nearest neighbour distance distribution function with respect to the mark set B, we divide Eq. (18) by Eq. (14), cf. the discussion following Eq. (14).

Since M is a metric space, say equipped with the metric $\rho(\cdot, \cdot)$, we may consider the family $\tilde{k}_{\tilde{h}}(m, n) := \tilde{k}(\rho(m, n)/\tilde{h})/\tilde{h}$ of kernels with bandwidth $\tilde{h} > 0$ based on some fixed measurable function $\tilde{k} : \mathbb{R}^+ \to \mathbb{R}^+$ that integrates to unity, and apply ideas from kernel estimation theory to both the mark set and the range, as exemplified by the following proposition.

Proposition 5.4 *Let Y be a stationary marked point process with finite positive intensity λ and marks in the complete, separable metric space (M, ρ) for which a regular version of the conditional intensity exists that satisfies the Nguyen–Zessin formula (Eq. 4). Write $y_k = (x_k, m_k)$, $s_k = d(x_k, (Y \cap (W \times M)) \setminus \{y_k\})$ and $b_k = d(x_k, \partial W)$. Then, given symmetric, measurable, non-negative kernels k_h with bandwidth h on $\mathbb{R}$ and $\tilde{k}_{\tilde{h}}$ with bandwidth $\tilde{h}$ on $M \times M$ that integrate to unity, for $m \in M$ and $t \in [0, T)$ with $T = \sup\{s : \ell(W^{\ominus s}) > 0\}$,*

$$\widehat{h_m(t)} = \frac{\sum_{y_k \in Y \cap (W \times M)} \tilde{k}_{\tilde{h}}(m, m_k)\, k_h(t - s_k)\, \mathbf{1}\{b_k > s_k\}/\ell(W^{\ominus s_k})}{\sum_{y_k \in Y \cap (W \times M)} \tilde{k}_{\tilde{h}}(m, m_k)\, \mathbf{1}\{b_k > s_k\}\, \mathbf{1}\{s_k > t\}/\ell(W^{\ominus s_k})}$$

with the convention that $0/0 = 0$ *in both numerator and denominator is a ratio-unbiased estimator of*

$$\frac{\lambda \int\limits_M \int\limits_0^T \tilde{k}_{\tilde{h}}(m,n)\, k_h(t-s)\, g_n(s)\, \mathrm{d}\nu_M(n)\, \mathrm{d}s}{\lambda \int\limits_M \int\limits_t^T \tilde{k}_{\tilde{h}}(m,n)\, g_n(s)\, \mathrm{d}\nu_M(n)\, \mathrm{d}s},$$

where $g_n(s) = f_F(s)\, \mathbb{E}\left[\lambda((0,n);Y) \mid d(0,Y) = s\right]/\lambda$, *cf.* (*Eq. 14*).

Proof As in the proof of Proposition 5.3, we may replace s_k by $d(x_k, Y \setminus \{y_k\})$. Since

$$\begin{aligned}
&\mathbb{E}\left[\sum_{(x,n)\in Y\cap(W\times M)} \frac{\tilde{k}_{\tilde{h}}(m,m_k)\, k_h(t-d(x,Y\setminus\{(x,n)\}))\, \mathbf{1}\{x \in W^{\ominus d(x,Y\setminus\{(x,n)\})}\}}{\ell(W^{\ominus d(x,Y\setminus\{(x,n)\})})}\right] \\
&= \lambda \int\limits_W \int\limits_M \mathbb{E}^{!(0,n)}\left[\frac{\tilde{k}_{\tilde{h}}(m,n)\, k_h(t-d(0,Y))\, \mathbf{1}\{x \in W^{\ominus d(0,Y)}\}}{\ell(W^{\ominus d(0,Y)})}\right] \mathrm{d}\ell(x)\, \mathrm{d}\nu_M(n) \\
&= \lambda \int\limits_M \tilde{k}_{\tilde{h}}(m,n)\, \mathbb{E}^{!(0,n)}\left[k_h(t-d(0,Y))\, \mathbf{1}\{d(0,Y) < T\}\right] \mathrm{d}\nu_M(n),
\end{aligned}$$

by the Campbell–Mecke theorem for stationary processes and Fubini's theorem. Similarly, one may show that the expectation of the denominator is given by

$$\lambda \int\limits_M \tilde{k}_{\tilde{h}}(m,n)\, \mathbb{E}^{!(0,n)}\left[\mathbf{1}\{t < d(0,Y) < T\}\right] \mathrm{d}\nu_M(n).$$

We proceed to show that for any pair of non-negative measurable functions h on M, and k on $\mathbb{R}^+$,

$$\int\limits_M h(n)\, \mathbb{E}^{!(0,n)}\left[k(d(0,Y))\right] \mathrm{d}\nu_M(n) = \int\limits_M \int\limits_0^\infty h(n)\, k(s)\, g_n(s)\, \mathrm{d}\nu_M(n)\, \mathrm{d}s.$$

To do so, follow the route from indicator functions h via step functions by linearity, to general h by approximation and the monotone convergence theorem. Indeed, let $h(n) = \mathbf{1}_B(n)$. If $\nu_M(B) = 0$, the desired identity trivially holds. Otherwise, by Proposition 5.2

$$
\begin{aligned}
\int_B \mathbb{E}^{!(0,n)}\left[k(d(0,Y))\right] \mathrm{d}\nu_M(n) &= \nu_M(B)\, \mathbb{E}_B^{!0}\left[k(d(0,Y))\right] \\
&= \nu_M(B) \int_0^\infty k(s)\, f_F(s)\, \mathbb{E}\left[\frac{\lambda_B(0;Y)}{\lambda\, \nu_M(B)} \middle| \mathrm{d}(0,Y) = s \right] \mathrm{d}s \\
&= \frac{1}{\lambda} \int_0^\infty k(s)\, f_F(s)\, \mathbb{E}\left[\lambda_B(0;Y) \mid d(0,Y) = s\right] \mathrm{d}s \\
&= \int_B \int_0^\infty k(s)\, \frac{f_F(s)}{\lambda}\, \mathbb{E}\left[\lambda((0,n);Y) \mid d(0,Y) = s\right] d\nu_M(n)\, \mathrm{d}s.
\end{aligned}
$$

To finish the proof, apply the claim with $h(n) = \tilde{k}_{\tilde{h}}(m,n)$ and $k(d(0,Y)) = k_h(t - d(0,Y))\, \mathbf{1}\{T > d(0,Y)\}$. The proof of the second statement follows upon replacement of the kernel $k_h(t - d(0,Y))$ by the indicator function $\mathbf{1}\{d(0,Y) > t\}$. $\square$

Note that $\widehat{h_m(t)}$ may be interpreted as the hazard rate of the nearest neighbour distance distribution function with respect to a mark in the infinitesimal neighbourhood $d\nu_M(m)$ at range t. By letting m run through M, a wealth of information on Y is obtained. We shall use the function in Sect. 6 below to investigate a random labelling hypothesis.

6 Forestry example

Below, we illustrate the use of the marked J-function by means of a data set of pine saplings in Finland (cf. Fig. 1) collected by Professor S. Kellomaki from Joensuu, and kindly provided by Professor A. Penttinen. The observation window W is the

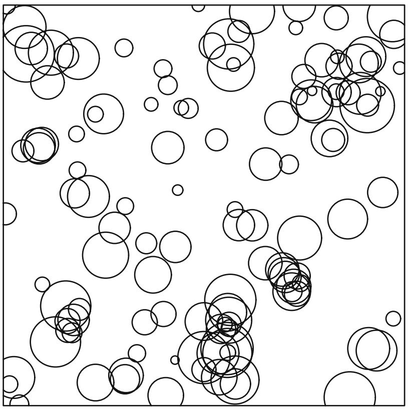

Fig. 1 Positions of 126 pine saplings (Kellomaki, Joensuu) within a $10 \times 10\,\mathrm{m}^2$ window. The marks record the height in meter and are represented by a disc

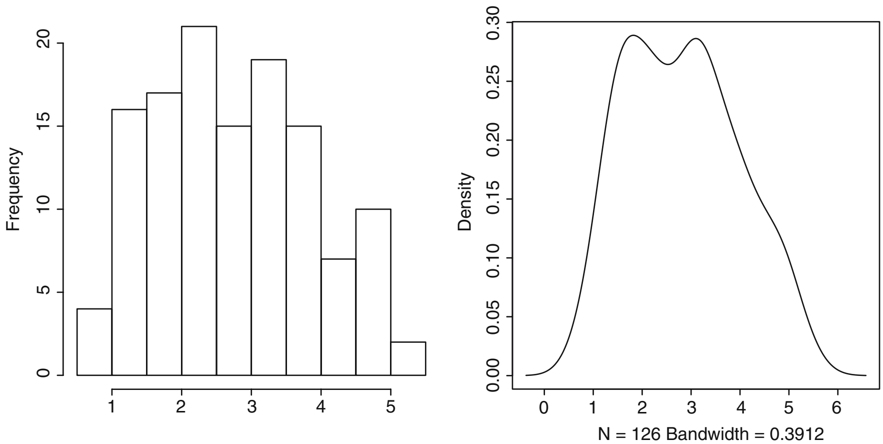

Fig. 2 Histogram and Gaussian kernel estimator of the mark distribution for the pine saplings data

square $[-5, 5] \times [-8, 2]$, but note that the data were originally recorded in a larger circular plot with polar coordinates. After transformation to Euclidean coordinates, to get rid of alignments at larger distances, some random rounding was done as a result of which there are a few close pairs of neighbours. For each of the 126 pines, the height and diameter at breast height were measured. The marks are strongly positively correlated. Moreover, a number of trees are broken resulting in zero diameter at breast height. For these reasons, we base our analysis on the height marks only, i.e. take $M = \mathbb{R}^+$.

We begin our analysis with first order characteristics. The intensity estimator is $\hat{\lambda} = 1.26$, and the sample mean of the mark distribution $\hat{\mu} = 2.83$. A histogram and kernel estimator are plotted in Fig. 2. Note that the histogram counts in bin B are unbiased estimators of $\lambda \, \nu_M(B) \, \ell(W)$, the kernel estimator is ratio-unbiased due to the fact that the number of points is random.

The Hanisch estimators for the nearest neighbour distance distribution function and empty space function introduced in Sect. 5 have been implemented in `R` using the package `spatstat` developed by Baddeley and Turner (2005). The t values at which we evaluated the estimators were separated by 0.025. For the lattice L, a regular 100 by 100 grid girting the boundary was used. Within the scope of this paper we restricted ourselves to exploratory data analysis. Of course a more formal test could be designed quite easily (Besag and Diggle, 1977). Empirical evidence (Baddeley et al., 2000; Chen, 2003; Thőnnes and Van Lieshout, 1999) suggests that the power of tests based on a J-function is comparable to that of the more powerful of the alternatives based on F or nearest neighbour distances.

Figure 3 shows the graph of the estimated J-function of the point process of locations. Note that it is less than 1 over the considered range, which suggests a clustering of trees. The impression is confirmed by a Monte Carlo test at the 1% level. Indeed, the J-function of the (unmarked) saplings data lies below the lower envelope computed over 99 independent simulations of a binomial point process with the same intensity up to $t \approx 0.55$. Note that since there are many small saplings in the field, one would not expect to see a hard core effect that is typically observed

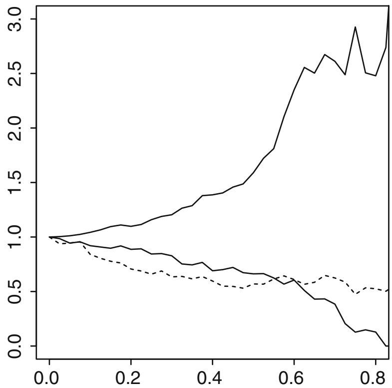

Fig. 3 Empirical $\widehat{J(t)}$-function (*dotted line*) for the pine saplings locations and envelopes of 99 simulations of a binomial process with 126 points (*solid lines*)

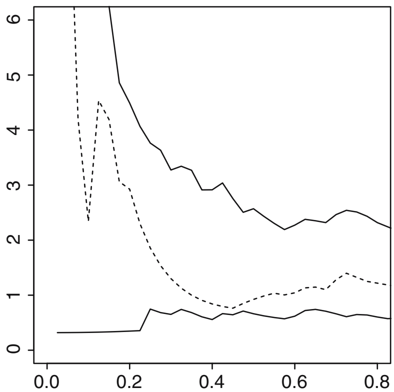

Fig. 4 Empirical mark pair correlation function $\widehat{g_{13}(t)}$-function (*dotted line*) for the pine saplings data and envelopes of 99 independent random re-labellings (*solid lines*)

in older, more established forests, a feature reinforced by the rounding involved in transforming the data from polar to Euclidean coordinates.

To investigate the mark dependence structure, we begin by partitioning the mark space in three parts: heights less than 2, heights in [2, 4], and heights larger than 4. Write λ_1, λ_2, and λ_3 for the respective intensities. The estimated mark pair correlation function $\widehat{g_{13}} = \rho_{13}^{(2)}/(\lambda_1\lambda_3)$ with respect to the smallest and largest height sets (cf. Stoyan and Stoyan, 1994, p. 266) is plotted in Fig. 4. Note that at this point we have to assume isotropy. As its values are larger than 1, it indicates positive correlation at smaller range. To assess statistical significance, we condition on the locations, and re-sample the labels without replacement. At the 1% level, i.e. for 99 independent samples, it can be seen from the upper and lower envelopes in Fig. 4 that the correlation is not significant. A similar message is given by the cross J_{13}- and J_{31}-functions, as illustrated in Fig. 5.

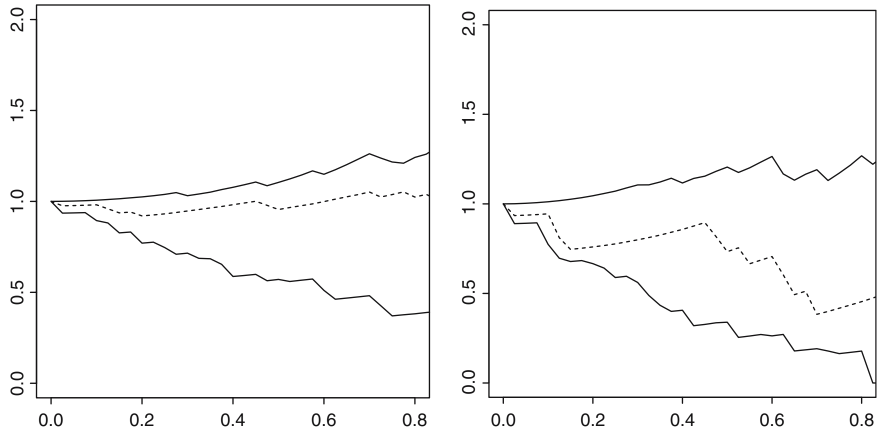

Fig. 5 Estimated cross J-function (*dotted line*) for the pine saplings data and envelopes of 99 independent random re-labellings (*solid lines*): $\widehat{J_{13}(t)}$ (*left*) and $\widehat{J_{31}(t)}$ (*right*)

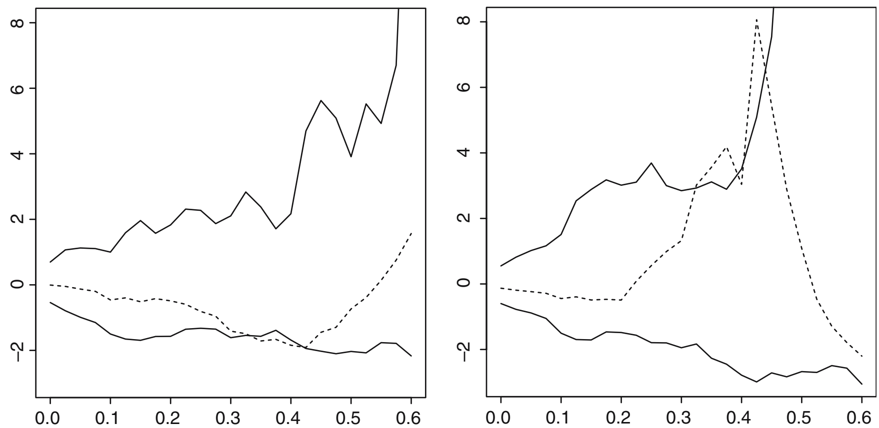

Fig. 6 Empirical difference of hazard rates $\widehat{h_m(t)} - \widehat{h_G(t)}$ (*dotted line*) for the pine saplings data and envelopes of 99 independent random re-labellings (*solid lines*) for $m = 1.5$ (*left*) and $m = 4.5$ (*right*)

To conclude this section, consider the hazard rate statistic of Proposition 5.4. Figure 6 presents graphs of $\widehat{h_m(t)}$ based on a box kernel for the mark with $\tilde{h} = 0.5$, and a Gaussian kernel with $h = 0.075$ for the nearest neighbour distances. In order to allow comparison with the random labelling null hypothesis, a plug-in Hanisch style estimator of the hazard rate of the underlying location process X was subtracted (cf. Proposition 3.1). Upper and lower envelopes based on 99 independent re-samplings of the marks without replacement are given as well. It can be seen that the deviation from the null hypothesis is significant both for large and small marks at intermediate range (from around $t = 0.3$). In the case $m = 4.5$, $h_m(t)$ is too large (many distances around 0.3), while for $m = 1.5$ there is a shortage of intermediate nearest neighbour distances due to a peak at larger inter-point distances.

7 Discussion and conclusion

In this paper, we introduced a summary statistic for stationary marked point patterns based on comparing the distance to the nearest (other) point of the pattern seen from an arbitrarily chosen origin to that from a typical marked point. This J-statistic captures both the type and strength of interaction, and reduces to a simple form under random labelling. We derived representations in terms of both Papangelou conditional intensity and correlation functions, thus relating the J-function to fundamental concepts in marked point process theory.

Further variations on the theme are possible. For example, the balls $B(0,t)$ used in the definition of $J_{\mathrm{A}}(t)$ may be replaced by any bounded Borel set. This would be particularly useful in a directional analysis of a non-isotropic marked point process. In another vein, only distances to points with a certain type of mark may be considered. More precisely, if $F_{\mathrm{B}}(t) = P\{Y \cap (B(0,t) \times B) \neq \emptyset\}$ is the empty space function of $Y \cap (B \times M)$ and $G_{\mathrm{AB}}(t) = P_A^{!0}\{Y \cap (B(0,t) \times B) \neq \emptyset\}$ the cross nearest neighbour distance distribution function from points with mark in A to those with mark in B, an A-to-B cross J-function for marked point processes is given by

$$J_{\mathrm{AB}}(t) = \frac{1 - G_{\mathrm{AB}}(t)}{1 - F_{\mathrm{B}}(t)}$$

for Borel subsets A and B of the mark space M with $\nu_M(A) > 0$, and all $t \geq 0$ for which $F_{\mathrm{B}}(t) < 1$. The function $J_{\mathrm{AB}}(t)$ measures the influence of the presence of a point with mark in the set A on the presence of points with a mark in B within distance t compared to the same event seen from an arbitrary origin. Values $J_{\mathrm{AB}} > 1$ can be interpreted as indicating inhibition of points with mark in B by those with mark in A. Similarly, values less than 1 mean that the presence of a point with a mark in the set A decreases the probability of finding points with a mark value in B nearby.

The definition of J_{AB} is not symmetric in the mark sets, which is sometimes an advantage as pointed out by Van Lieshout and Baddeley (1999) and illustrated in practice by Foxall and Baddeley (2002). Note that if $\nu_M(B) = 0$, the expected number of points with mark in B is zero. If $A_1, \ldots, A_n$ form a partition of M, we may assign label i to a point if its mark falls in A_i and write X_i for the locations of points with label i. In this sense, an analysis based on the cross J-function amounts to an analysis of the multivariate point pattern $Y = (X_1, \ldots, X_n)$ as in Van Lieshout and Baddeley (1999) or the thesis by Chen (2003). In general, though, A and B need not be disjoint. Indeed, for $B = M$, $G_{\mathrm{AM}} \equiv G_{\mathrm{A}}$, and we regain the J-function with respect to the mark set A.

From a statistical perspective, we discussed Hanisch style kernel estimators of densities and hazard rates of J_{A} that suppress the variance explosion at larger range encountered by plug-in estimators of J_{A} itself. Finally, the new statistic was used to explore the spatial structure of a forestry data set.

Acknowledgements The author is grateful to A. Penttinen for the data and helpful discussions, to C. Beisbart for posing a question that motivated this work, to A. Baddeley and R. Turner for access to ‘spatstat’, and to P. Gregori and two anonymous referees for their suggestions.

References

Baddeley, A.J., Gill, R.D. (1994). The empty space hazard of a spatial pattern, Research Report, No. 1994/3, Department of Mathematics, The University of Western Australia.

Baddeley, A.J., Gill, R.D. (1997). Kaplan–Meier estimators of distance distributions for spatial point processes. *Annals of Statistics*, *25*, 263–292.

Baddeley, A.J., Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *International Statistical Review*, *57*, 89–121.

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

Baddeley, A.J., and Turner, R. (2005). Spatstat; *An R library for spatial statistics*, `http://www.cran.r-project.org`.

Baddeley, A.J., Kerscher, M., Schladitz, K., Scott, B.T. (2000). Estimating the J function without edge correction. *Statistica Neerlandica*, *54*, 315–328.

Besag, J.E., Diggle, P.J. (1977). Simple Monte Carlo tests for spatial pattern. *Applied Statistics*, *26*, 327–333.

Bedford, T., Van den Berg, J. (1997). A remark on the Van Lieshout and Baddeley J-function for point processes. *Advances in Applied Probability*, *29*, 19–25.

Chen, J. (2003). *Summary statistics in point patterns and their applications*, PhD Thesis, Department of Mathematics and statistics, Curtin University of Technology.

Chen, J., Baddeley, A.J., Nair, G. (2001). Uncorrected estimators of J-functions in multivariate point patterns, *Paper Presented at 11th International Workshop on Stereology, Stochastic Geometry and Related Fields*, Perth, Australia.

Chiu, S.N., Stoyan, D. (1998). Estimators of distance distributions for spatial patterns. *Statistica Neerlandica*, *52*, 239–246.

Cox, D.R., Lewis, P.A.W. (1972). Multivariate point processes, In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3, 401–448, University of California Press, Berkeley

Cressie, N.A.C. (1991). *Statistics for spatial data*. New York: Wiley.

Daley, D.J., Vere-Jones, D. (1988). *An introduction to the theory of point processes: vol. I. Elementary theory and methods*. New York: Springer.

Diggle, P.J. (1983). *Statistical analysis of spatial point patterns*. London: Academic.

Foxall, R., Baddeley, A.J. (2002). Nonparametric measures of association between a spatial point process and a random set, with geological applications. *Journal of the Royal Statistical Society Series C*, *51*, 165–182.

Hanisch, K.-H. (1984). Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point patterns. *Mathematische Operationsforschung und Statistik, Series Statistics*, *15*, 409–412.

Hansen, M.B., Gill, R.D., Baddeley, A.J. (1996). Kaplan–Meier type estimators for linear contact distributions. *Scandinavian Journal of Statistics*, *23*, 129–155.

Kerscher, M. (1998). Regularity in the distribution of superclusters? *Astronomy and Astrophysics*, *336*, 29–34.

Kerscher, M., Schmalzing, J., Buchert, T., Wagner, H. (1998). Fluctuations in the IRAS 1.2 Jy catalogue, *Astronomy and Astrophysics*, *333*, 1–12.

Kerscher, M., Pons-Bordería, M.J., Schmalzing, J., Trasarti–Battistoni, R., Buchert, T., Martínez, V.J., Valdarnini, R. (1999). A global descriptor of spatial pattern interaction in the galaxy distribution. *Astrophysical Journal*, *513*, 543–548.

Last, G., Holtmann, H. (1999) On the empty space function of some germ-grain models. Pattern Recognition, *32*, 1587–1600.

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

van Lieshout, M.N.M., Baddeley, A.J. (1999). Indices of dependence between types in multivariate point patterns. *Scandinavian Journal of Statistics*, *26*, 511–532.

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

Ogata, Y., Tanemura, M. (1989). Likelihood estimation of soft-core interaction potentials for Gibbsian point patterns. *Annals of the Institute of Statistical Mathematics*, *41*, 583–600.

Paulo, M.J. (2002). *Statistical sampling and modelling for cork oak and eucalyptus stands*, PhD thesis, Department of Mathematics, Wageningen University.

Penttinen, A., Stoyan, D. (1989). Statistical analysis for a class of line segment processes. *Scandinavian Journal of Statistics*, *16*, 153–168.
Ripley, B.D. (1988). *Statistical inference for spatial processes*. Cambridge: Cambridge University Press.
Ripley, B. D. (1989). Gibbsian interaction models. In D.A. Griffiths (Eds), Spatial statistics: past, present and future, (55–57).
Ripley, B.D., and Kelly, F.P. (1977). Markov point processes. *Journal of the London Mathematical Society*, *15*, 188–192.
Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Londan: Chapman and Hall.
Stein, A., Van Lieshout, M.N.M., Booltink, H.W.G. (2001). Spatial interaction of methylene blue stained soil pores. *Geoderma*, *102*, 101–121.
Stoyan, D., Stoyan, H. (1994). *Fractals, random shapes and point fields. Methods of geometrical statistics*. (translated from the 1992 German original), Chichester: Wiley.
Stoyan, D., Stoyan, H. (2000). Improving ratio estimators of second order point process characteristics. *Scandinavian Journal of Statistics*, *27*, 641–656.
Stoyan, D., Kendall, W.S., Mecke, J. (1987). *Stochastic geometry and its applications*. Berlin: Akademie-Verlag.
Thőnnes, E., Van Lieshout, M.N.M. (1999). A comparative study on the power of Van Lieshout and Baddeley's *J*–function, *Biometrical Journal*, *41*, 721–734.
White, S.D.M. (1979). The hierarchy of correlation functions and its relation to other measures of galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, *186*, 145–154.