Poles of pair correlation functions: When they are real?

Ka Yiu Wong[1] · Dietrich Stoyan[1]

Abstract

The most common standard estimator of the pair correlation function (PCF) of a point process has a pole at zero, which is in most cases a statistical artifact. However, sometimes it makes sense to assume that a pole really exists. We propose two independent approaches for the proof of existence of a PCF's pole and for the determination of its order. In the first, we use a summary characteristic F that transforms the PCF's pole order to the location of F's pole, while the other one uses a natural estimation method based on Ripley's K-function. These methods are applied to simulated samples of two classical point process models and two cluster point process models with special geometries. Finally, we use the approach in the statistical analysis of a classical point pattern of pine trees and a highly clustered pattern of nonmetallic inclusions in steel.

Keywords Pair correlation function · High degree of clustering · Pole

1 Introduction

Today the pair correlation function is a strong tool of point process statistics (see Baddeley et al. 2015; Illian et al. 2008). (To make the paper self-contained, we explain the basic concepts in the beginning of Sect. 2.) Empirical pair correlation functions help to understand the structure of many point patterns. This is easy if the patterns are completely random or more or less regular, but problems appear when the patterns are clustered. Then, the pair correlation function $g(r)$ has large values for small r and the corresponding estimates are not precise for these r.

Often, even in textbooks, figures in Fig. 1 show empirical pair correlation functions, where it is not clear whether there is a pole of $g(r)$ at $r = 0$; see also

✉ Ka Yiu Wong
  wky313@gmail.com

  Dietrich Stoyan
  stoyan@math.tu-freiberg.de

[1] Institut für Stochastik, Technische Universität Bergakademie Freiberg, Prüferstraße 9,
  09599 Freiberg, Germany

Fig. 1 An empirical pair
correlation function from the
book Chiu et al. (2013), there
Figure 4.4(b). It is not clear
whether there is a pole at $r = 0$.
See the discussion in Sect. 6.1
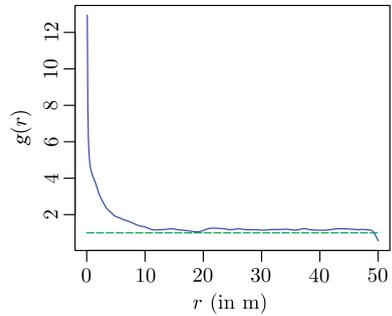
Figure 16.5 in Gelfand et al. (2010), Figure 7.24 in Baddeley et al. (2015), and Figure 5.8 in Illian et al. (2008).

Surely, the question of existence of a pole of $g(r)$ is a difficult one. A pole is a mathematical idea, and therefore it is difficult to detect a pole from statistical data. There are mathematical point process models the pair correlation functions of which have poles, i.e., are proportional to $r^{-\gamma}$ for small r. This order γ is limited, and it is always between 0 and the space dimension d; the present paper always considers the case $d = 2$. (By the way, a pair correlation function may have poles also at other values of r, not only at $r = 0$. Section 4 gives examples.) In empirical point patterns, there is always a positive minimum inter-point distance, which means that at very short distances, there is not the theoretical behavior of a cluster process with a pole at $r = 0$. Nevertheless, there are natural point patterns where it makes sense to assume the existence of a pole, at least for an approximative elegant model. A prominent example is the three-dimensional pattern of galaxies, where the assumption is that $\gamma = 1.8$, see Totsuji and Kihara (1969), Peebles (1974), Davis and Peebles (1983). Also one of the patterns discussed in the present paper may belong to a structure where a pole at $r = 0$ is credible. It was just this pattern of inclusions in samples of steel that has led to the present paper. If these inclusions would have a pair correlation function with a pole, the choice of a suitable statistical model is directed to special models not (yet) popular in recent point process statistics.

The problem with a pole at $r = 0$ is statistically complicated since one of the standard estimators quite often leads to estimates having such a pole. Thus, in many cases, poles of empirical pair correlation functions are just statistical artifacts. This even happens in the case of Poisson point process samples; only for hard-core processes, this effect does not play a role. Therefore, a statistical proof of the existence of a pole is a difficult task. For small r, which to consider is necessary in pole detection, the statistics of pair correlation functions is difficult, since then there is a complicated interplay of bandwidth h and inter-point distance r.

In the present paper, we suggest two independent approaches for the determination of poles at $r = 0$ and their order. One uses the natural way to employ Ripley's K-function, while the other works with a new summary characteristic that transforms the order of a possible pole at $r = 0$ to the location of a pole.

The paper is organized as follows. A short review of pair correlation function estimators is given in Sect. 2. Then, a mathematical definition of a pole of a pair

correlation function and its order and the two approaches for detection and determination of poles are provided in Sect. 3. Section 4 describes point process models without and with poles. The two approaches are then applied to simulated data in Sect. 5 and two real data sets in Sect. 6.

2 Pair correlation function estimators

Suppose N is a stationary and isotropic point process i.e., a random subset of $\mathbb{R}^2$ with the property that the intersection of N with any bounded subset of $\mathbb{R}^2$ is of finite cardinality. 'Stationarity' and 'isotropy' mean invariance of the distribution under translations and rotations. The mean number of points per area unit of N is called 'intensity' and denoted by λ. The pair correlation function $g(r)$ of N is the normalized second-order product density $\varrho(r)$, normalized by division by λ^2. Intuitively, $\varrho(r)\mathrm{d}x_1\mathrm{d}x_2$ is the probability that N has each a point in infinitesimally small regions of area $\mathrm{d}x_1$ and $\mathrm{d}x_2$ around two points x_1 and x_2 of distance r. Large values of $g(r)$ for some r indicate that there is a tendency that N has more point pairs of an inter-point distance r than a completely random point process. See also Illian et al. (2008) and Daley and Vere-Jones (2008).

For statistical analysis, N is observed in a compact window $W \subset \mathbb{R}^2$ of positive Lebesgue measure $v(W)$. An edge-corrected kernel estimator of its pair correlation function is

$$\hat{g}_k(r) = \frac{1}{2\pi r \hat{\lambda}^2} \sum_{x_1, x_2 \in N \cap W}^{\neq} \frac{k_h\big(\|x_1 - x_2\| - r\big)}{v\big(W \cap (W - x_1 + x_2)\big)} \quad \text{for } r \geq 0,$$

where $\hat{\lambda}$ is an intensity estimator of N, $v(\cdot)$ is the Lebesgue measure, $W \cap (W - x_1 + x_2) = \{x - x_1 + x_2 \in W : \text{for all } x \in W\}$ and $k_h(t)$ is a kernel function in the form $k(t/h)/h$ for some probability density $k(\cdot)$ and bandwidth $h > 0$, see Illian et al. (2008, p. 230). Because of the inverse term $1/r$, the estimator $\hat{g}_k(r)$ tends to infinity as $r \downarrow 0$.

An alternative estimator (see Guan 2007) of $g(r)$ is obtained by replacing the r at the denominator of $\hat{g}_k(r)$ by pairwise distances $\|x_i - x_j\|$. It is in the form

$$\hat{g}_d(r) = \frac{1}{2\pi \hat{\lambda}^2} \sum_{x_1, x_2 \in N \cap W}^{\neq} \frac{k_h\big(\|x_1 - x_2\| - r\big)}{\|x_1 - x_2\| v\big(W \cap (W - x_1 + x_2)\big)} \quad \text{for } r \geq 0.$$

Notice that this estimator $\hat{g}_d(r)$ has a large bias at small lags r. To reduce this bias, Guan (2007) proposed a modified estimator which is expressed as

$$\hat{g}_c(r) = \frac{\hat{g}_d(r)}{\int_{-h}^{\min\{r,h\}} k_h(t)\,\mathrm{d}t} \quad \text{for } r \geq 0.$$

The estimators $\hat{g}_d(r)$ and $\hat{g}_c(r)$ are always finite at $r = 0$. Thus, they are reasonable choices for estimating the pair correlation functions of most of the classical point process models such as Poisson point process, Matérn cluster process and modified

Thomas process, which have no pole at $r = 0$. On the other hand, in some special cases, the pair correlation function of a point process may approach infinity as r tends to 0. In such cases, the estimator $\hat{g}_k(r)$ may be preferable for estimating $g(r)$ at small r.

3 Detection of PCF's pole

In this section, we give a formal definition of poles of pair correlation functions and propose two methods to detect, determine and characterize the poles. The first approach is a summary characteristic whose expectation has a pole at the PCF's pole order minus two (the dimension of the space). The second approach is least-square estimation of the pole order based on empirical K-functions. Estimation of pole orders requires precise information of point pairs with small inter-point distances and thus, a large number n of points is necessary in order to obtain accurate results in pole statistics.

3.1 Definition of PCF's pole

Given a pair correlation function $g(r)$, we say that the function $g(r)$ has a pole at $r = 0$ of order $\gamma > 0$ if

$$g(r) = \Theta(r^{-\gamma}) \quad \text{for } 0 \le r < r_0,$$

for some small r_0. This is equivalent to saying that there exist constants C_L and C_U such that

$$C_L r^{-\gamma} \le g(r) \le C_U r^{-\gamma} \quad \text{for } 0 \le r < r_0.$$

When a point process has finite $g(0)$, its $g(r)$ has not a pole at $r = 0$ and therefore $\gamma = 0$. To simplify wording, we will say in the following that a point process has a pole order γ if its pair correlation function has a pole of order γ at $r = 0$.

3.2 The *F*-approach for detecting the pole

Suppose that the pair correlation function of a point process N has a pole of order $\gamma \ge 0$ at $r = 0$. Consider the following function

$$F(s; r_f) := \sum_{x_1, x_2 \in N}^{\ne} \frac{\mathbf{1}_W(x_1)\mathbf{1}_W(x_2)\mathbf{1}_{[0, r_f]}(\|x_1 - x_2\|)\|x_1 - x_2\|^s}{\lambda^2 \nu\big(W \cap (W - x_1 + x_2)\big)} \quad \text{for } r_f \le r_0,$$

where $\mathbf{1}_A(\cdot)$ is the indicator function of the set A and r_f is a pre-chosen distance. The mean of the right-hand side is determined by using equation (4.3.19) in Illian et al. (2008), i.e.,

$$\mathbb{E}\left[\sum_{x_1,x_2\in N}^{\neq} f(x_1,x_2)\right] = \int\int f(x,x+h)\varrho(h)\,\mathrm{d}h\,\mathrm{d}x,$$

where $\varrho(h)$ denotes the second-order product density of N, which is a consequence of the Campbell-Mecke theorem (see Chiu et al. 2013, p. 130) or the refined Campbell theorem (see Daley and Vere-Jones 2008, p. 288).

With

$$f(x_1,x_2) = \frac{\mathbf{1}_W(x_1)\mathbf{1}_W(x_2)\mathbf{1}_{[0,r_f]}(\|x_1-x_2\|)\|x_1-x_2\|^s}{\lambda^2\nu\big(W\cap(W-x_1+x_2)\big)},$$

the lower bound E_L and upper bound E_U of the expectation $\mathbb{E}\big[F(s;r_f)\big]$ can be expressed as

$$E_X = \frac{2\pi C_X}{s-\gamma+2}r_f^{s-\gamma+2} \quad \text{for } s > \gamma-2,$$

where $X = L$ or U. Thus, $\mathbb{E}\big[F(s;r_f)\big]$ can be represented by using the Big-Theta notation as

$$\mathbb{E}\big[F(s;r_f)\big] = \Theta\left(\frac{2\pi}{s-\gamma+2}r_f^{s-\gamma+2}\right) \quad \text{for } s > \gamma-2,$$

and has a pole at $s = \gamma - 2$. That means that the order γ of the pole of $g(r)$ is transformed to the location of the pole of $\mathbb{E}\big[F(s;r_f)\big]$. The function $F(s;r_f)$ may give information about the pole order γ as its expectation tends to infinity when $s \downarrow \gamma - 2$. With a simple plot of $F(s;r_f)$ against s under a proper ordinate scaling, there is hope to prove the existence of a PCF's pole. However, this is a difficult issue as we have to consider values of $F(s;r_f)$ that are close to infinity. Notice that depending on r_f, the test function $F(s;r_f)$ may exist even when its expectation is undefined, i.e., if $s \leq \gamma - 2$.

The pre-chosen distance r_f plays a crucial role in detecting the pole. The bounds E_L and E_U are decreasing functions with respect to s if and only if r_f is small enough. With a suitable choice of r_f, such as $r_f = 1$, which always results in an $F(s;r_f)$ decreasing in s, one can have a clearer view on where and how the test function approaches infinity. On the other hand, the choice of r_f is restricted by the minimum pairwise distance of the points. A simple strategy to relax the restriction is rescaling the point pattern since the pole order is invariant to rescaling. To see the rescaling effect on the test function, denote $\alpha N = \{\alpha x : x \in N\}$ as the rescaled point pattern of N with intensity $\lambda_{\alpha N} = \alpha^{-2}\lambda_N$. The window W of N is then rescaled as $\alpha W = \{\alpha x : x \in W\}$ such that $\nu(\alpha W) = \alpha^2\nu(W)$. Simple calculation reveals that the test function $F_N(s;r_f)$ of N can be re-expressed in terms of the test function $F_{\alpha N}(s;\alpha r_f)$ of αN as $F_N(s;r_f) = \alpha^{-s-2}F_{\alpha N}(s;\alpha r_f)$. Thus, one may apply the detection method without actually rescaling the point pattern but the test function.

3.3 The *K*-approach for estimating pole order

Another method to estimate the pole order is the *K*-approach. If a point process has a pole order γ, its *K*-function satisfies

$$K(r) = \Theta\left(r^{2-\gamma}\right) \approx Cr^{2-\gamma} + O\left(r^\zeta\right) \quad \text{for } 0 \le r < r_0,$$

for some constant C and $\zeta > 2 - \gamma$. Hence, a natural approach for estimating the pole order is fitting a power law to the *K*-function. Unlike the case of pair correlation function, estimators of $K(r)$ do not require selection of bandwidths, which is a complicated issue in practice, and they have no statistical artifact at $r = 0$. Therefore, empirical *K*-functions are relatively less sensitive to choices of parameters and estimators, and hence their use is preferable for direct estimation of pole orders.

A popular translational-edge-corrected estimator (see Chiu et al. 2013, Section 4.7.4) of the *K*-function has the form

$$\hat{K}(r) = \frac{1}{\hat{\lambda}^2} \sum_{x_1, x_2 \in N}^{\neq} \frac{\mathbf{1}_{[0,r]}(\|x_1 - x_2\|)}{v\left(W \cap (W - x_1 + x_2)\right)} \quad \text{for } r \ge 0.$$

To estimate the pole order, the *K*-function is estimated at $r = d_i, i = 1, \ldots, k_n$, where d_i are the pairwise distances of the point pattern which are smaller than or equal to a pre-chosen distance r_k. Nonlinear least square methods can then applied on $\left(d_i, \hat{K}(d_i)\right)$ to estimate the order $\hat{\gamma}$. Throughout this paper, we use the trust-region-reflective least squares algorithm, implemented by the nonlinear least-squares solver lsqnonlin in the MATLAB optimization toolbox, with the restriction $\hat{\gamma} \in [0, 2]$.

A pole of pair correlation function is a local property. Thus, the pre-chosen distance r_k should be small enough to avoid considering the global structure of the point pattern. Otherwise, the estimates $\hat{\gamma}$ will tend to zero.

4 Examples of point process models without and with poles

Below, we provide some examples of point process models and state their corresponding pole orders at $r = 0$. The classical point process models, such as the Poisson point process and modified Thomas process, have pair correlation functions without poles. On the other hand, models with a special geometry of clusters, such as processes with chain-like clusters, can have a pole at $r = 0$.

Our examples show that the geometry of clusters plays a crucial role to the existence of a pole but not the number of daughters. Thus, a pole cannot be created by just simply making denser clusters. We mention here that one can obtain a point process with pole by superposition of point processes with and without poles.

4.1 Poisson point process

A homogeneous Poisson point process has a pair correlation function which is identical to 1, which means it has no pole at $r = 0$ or it is $\gamma = 0$. The expected value of the test function is

$$\mathbb{E}\big[F(s;r_f)\big] = \frac{2\pi}{s+2} r_f^{s+2} \quad \text{for } s > -2.$$

It has a pole at $s = -2$, which is equivalent to the fact that the pair correlation function has no pole at $r = 0$.

4.2 Modified Thomas process

A modified Thomas process (see Illian et al. 2008, p. 377) is a special Neyman-Scott cluster point process. Denote λ_p as the intensity of the Poisson parent points. The representative cluster is an isotropic centered Poisson process with mean total number $\bar{c}$. The distribution of the daughter points around the parent points is the symmetric normal distribution with parameter σ. Its pair correlation function is

$$g(r) = 1 + \frac{1}{4\pi\lambda_p\sigma^2} \exp\left(-\frac{r^2}{4\sigma^2}\right) \quad \text{for } r \geq 0.$$

It is obvious that $g(r)$ has no pole at $r = 0$ as $g(0) = 1 + (4\pi\lambda_p\sigma^2)^{-1} < \infty$. The function $g(r)$ is independent of the mean daughter numbers $\bar{c}$. Thus, even a very large value of $\bar{c}$ would still result in a pattern without a pole. The expectation $\mathbb{E}\big[F(s;r_f)\big]$ is in the form

$$\mathbb{E}\big[F(s;r_f)\big] = \frac{2\pi}{s+2} r_f^{s+2} + \frac{1}{\lambda_p(2\sigma)^s} \Gamma_L\left(\frac{s}{2}+1, \frac{r_f^2}{4\sigma^2}\right) \quad \text{for if } s > -2,$$

where $\Gamma_L(a,x) = \int_0^x y^{a-1}e^{-y}\,dy$. Similar to the Poisson process case, $\mathbb{E}\big[F(s;r_f)\big]$ tends to infinity when $s \downarrow -2$.

4.3 Pair-cluster process

The pair-cluster process was introduced by Stoyan (1994). Its parent points form a Poisson process with intensity η. Each parent point is the center of a randomly orientated segment of random length, and the i.i.d. lengths of the segments follow a probability density function $f(r)$. The cluster process is then formed by the two endpoints of the random segments. The pair correlation function of the process is given by

$$g(r) = 1 + \frac{f(r)}{4\eta\pi r} \quad \text{for } r \geq 0.$$

With a suitable choice of $f(r)$, the pattern may have a pole at $r = 0$. An example is the Weibull distribution with

$$f(r) = \frac{k}{\sigma}\left(\frac{r}{\sigma}\right)^{k-1} \exp\left(-\left(\frac{r}{\sigma}\right)^k\right) \quad \text{for } r \geq 0,$$

where $k > 0$ and $\sigma > 0$ are the shape and scale parameters, respectively. When the segment lengths follow a Weibull distribution, its pair correlation function has the form

$$g(r) = 1 + \frac{k}{4\eta\pi\sigma^k} r^{k-2} \exp\left(-\left(\frac{r}{\sigma}\right)^k\right) \quad \text{for } r \geq 0.$$

Thus, $g(r)$ can be shown to have a pole of order $\gamma = 2 - k$ at $r = 0$. The expectation of the test function is

$$\mathbb{E}\left[F(s;r_f)\right] = \frac{\pi}{s+1} r^{s+2} + \frac{1}{4\eta\pi\sigma^s} \Gamma_L\left(\frac{s}{k}+1, \left(\frac{r_f}{\sigma}\right)^k\right) \quad \text{for } s > \max\{-2, -k\}.$$

For any $k \in (0, 2]$, the expectation of $F(s;r_f)$ tends to infinity when $s \downarrow -k$. Here, we consider the case of $k \leq 2$ only as we consider the two-dimensional space in the present paper.

If one considers a three-parameter Weibull distribution, i.e., a Weibull distribution with an additional shift parameter $a > 0$ which shifts the distribution and controls the lower bound of the distribution support, the resultant pair-cluster process is then a process with nonzero pole order at $r = a$ as mentioned in Sect. 1.

4.4 Segment Cox process

Instead of using the two endpoints of the random segments as the daughter points, a segment Cox process (Martinez and Saar 2002) uses 1D Poisson point processes with intensity ξ as the daughter points on each segment, and the union of all these Poisson point process forms the resultant cluster process. Suppose the segment length probability density function $f(r)$ has support $[a, b]$. By using a formula for the 'Poisson segment process' of Stoyan (1983), the pair correlation function of the cluster process can be expressed as

$$g(r) = 1 + \frac{\mathbf{1}_{[0,b)}(r)}{\eta\mu\pi r} \int_{\max\{a,r\}}^{b} (x-r)f(x)\, dx \quad \text{for } r \geq 0,$$

where μ is the mean of $f(r)$ and η, as for the previous model, is the intensity of parent points. If the length follows uniform distribution with support $[0, b]$, the pair correlation function is

$$g(r) = 1 + \mathbf{1}_{[0,b)}(r)\left(\frac{r^2 - 2rb + b^2}{\eta\pi r b^2}\right).$$

The function $g(r)$ has a pole of order $\gamma = 1$ at $r = 0$. The expectation $\mathbb{E}\left[F(s;r_f)\right]$ for any $s > -1$ is given by

$$\mathbb{E}\big[F(s;r_f)\big] = \frac{\pi}{s+2}r_f^{s+2} + \frac{2}{\eta b^2}\Big(\frac{1}{s+3}\min\{b,r_f\}^{s+3}$$

$$- \frac{2b}{s+2}\min\{b,r_f\}^{s+2} + \frac{b}{s+1}\min\{b,r_f\}^{s+1}\Big).$$

It has a pole at $s = -1$.

We remark here that we suppose that a pole of order 1 appears also if we replace the uniform distribution by another length distribution or the segments by smooth curves. (We could show this for circular lines.) Replacing the 1D Poisson point process by renewal processes can yield poles at other values of r. (For example, at $r = a$ if the points on the segments have constant inter-point distance a.)

5 Simulations

To study the behavior of the test function $F(s;r_f)$ and the performance of the pole-order estimation method for point process data, six different point process models are considered in simulations.

- Model 1 is the Poisson point process.
- Models 2 and 3 are modified Thomas processes with parameters $(\lambda_p, \mu, \sigma) = (20, 0.05, 0.5)$ and $(20, 0.05, 1)$, respectively.
- Models 4 and 5 are pair-cluster processes with Weibull length distribution. The parameters are $(\eta, k, \sigma) = (0.5, 1.5, 0.15)$ and $(0.5, 0.5, 0.15)$, respectively.
- Model 6 is a segment Cox process with uniform length distribution. The parameters are $(\eta, \xi, b) = (0.02, 5, 2.5)$.

The six models were generated under four combinations of parameters (W, λ): $([0, 10]^2, 10)$, $([0, 10]^2, 5)$, $([0, 20]^2, 1)$ and $([0, 30]^2, 1)$. For each combination, 100 point patterns of each model were simulated. When computing the test function $F(s;r_f)$, the patterns are rescaled to $[0, 1]^2$ to ensure that $F(s;r_f)$ is a decreasing function.

The test function and the K-approach depend on pre-chosen distances r_f and r_k, respectively. In the simulations, we considered six choices of distances for r_f and r_k. The first three choices $r^{(1)}$, $r^{(2)}$ and $r^{(3)}$ are defined as

$$r^{(i)} = \max_{j \in \{1,\ldots,100\}} \big\{x_i\text{th smallest pairwise distances of the }j\text{th pattern}\big\},$$

where $x_1 = 25$, $x_2 = 50$ and $x_3 = 100$. The other three choices $r^{(4)}$, $r^{(5)}$ and $r^{(6)}$ are 0.1 0.5 and 1 over the square root of the intensity, respectively. The values of $r^{(k)}$ for each model and (W, λ) are shown in Table 1. The distances $r^{(5)}$ and $r^{(6)}$ are largest in all cases. For model 1, 2 and 3, $r^{(4)}$ is smaller than $r^{(1)}$, while for model 5 and 6, $r^{(4)}$ is larger than or close to $r^{(3)}$. For model 4, $r^{(4)}$ is between $r^{(1)}$ and $r^{(3)}$.

Table 1 Values of $r^{(k)}$ for the simulated models under different parameters (W, λ)

Model	Window	λ	$r^{(1)}$	$r^{(2)}$	$r^{(3)}$	$r^{(4)}$	$r^{(5)}$	$r^{(6)}$
1	$[0, 10]^2$	10	0.0510	0.0695	0.0945	0.0316	0.1581	0.3162
		5	0.1018	0.1418	0.1895	0.0447	0.2236	0.4472
	$[0, 20]^2$	1	0.2531	0.3442	0.4792	0.1000	0.5000	1.0000
	$[0, 30]^2$		0.1641	0.2145	0.3060	0.1000	0.5000	1.0000
2	$[0, 10]^2$	10	0.0497	0.0684	0.0927	0.0316	0.1581	0.3162
		5	0.0995	0.1409	0.1882	0.0447	0.2236	0.4472
	$[0, 20]^2$	1	0.2417	0.3432	0.4464	0.1000	0.5000	1.0000
	$[0, 30]^2$		0.1575	0.2197	0.2820	0.1000	0.5000	1.0000
3	$[0, 10]^2$	10	0.0495	0.0675	0.0928	0.0316	0.1581	0.3162
		5	0.0988	0.1345	0.1935	0.0447	0.2236	0.4472
	$[0, 20]^2$	1	0.2806	0.3776	0.5175	0.1000	0.5000	1.0000
	$[0, 30]^2$		0.1799	0.2341	0.3111	0.1000	0.5000	1.0000
4	$[0, 10]^2$	10	0.0230	0.0335	0.0507	0.0316	0.1581	0.3162
		5	0.0434	0.0618	0.0926	0.0447	0.2236	0.4472
	$[0, 20]^2$	1	0.0502	0.0749	0.1366	0.1000	0.5000	1.0000
	$[0, 30]^2$		0.0294	0.0449	0.0698	0.1000	0.5000	1.0000
5	$[0, 10]^2$	10	0.0009	0.0039	0.0147	0.0316	0.1581	0.3162
		5	0.0053	0.0157	0.0581	0.0447	0.2236	0.4472
	$[0, 20]^2$	1	0.0068	0.0217	0.1132	0.1000	0.5000	1.0000
	$[0, 30]^2$		0.0015	0.0045	0.0167	0.1000	0.5000	1.0000
6	$[0, 10]^2$	10	0.0052	0.0103	0.0202	0.0316	0.1581	0.3162
		5	0.0114	0.0225	0.0452	0.0447	0.2236	0.4472
	$[0, 20]^2$	1	0.0153	0.0296	0.0834	0.1000	0.5000	1.0000
	$[0, 30]^2$		0.0048	0.0100	0.0194	0.1000	0.5000	1.0000

5.1 Simulation results

5.1.1 *F*-approach

The *F*-approach tries to detect the existence of a pole by plotting $F(s; r_f)$ against s with a pre-chosen r_f. If $F(s; r_f)$ has a pole at some $s > -2$, it is expected that the corresponding pattern has a pole of nonzero order. Since $F(s; r_f)$ is a decreasing function for all s, one may observe the occurrence of pole at different values of s by using different ordinate scalings. Nevertheless, our simulation results reveal that plots of $F(s; r_f)$ with the y-axis limit $[0, 25]$ usually can show where the steep increase of $F(s; r_f)$ starts. The simulation results also suggest that the choices of r_f are not important since the values of $F(s; r^{(i)})$ are similar under different $r^{(i)}$. Thus, we report only the results of $F(s; r^{(6)})$.

Figure 2 shows the plots of the envelopes of $F(s; r^{(6)})$ against s for the six models. The width of the envelopes, i.e., the variation of F at fixed s, increases as the
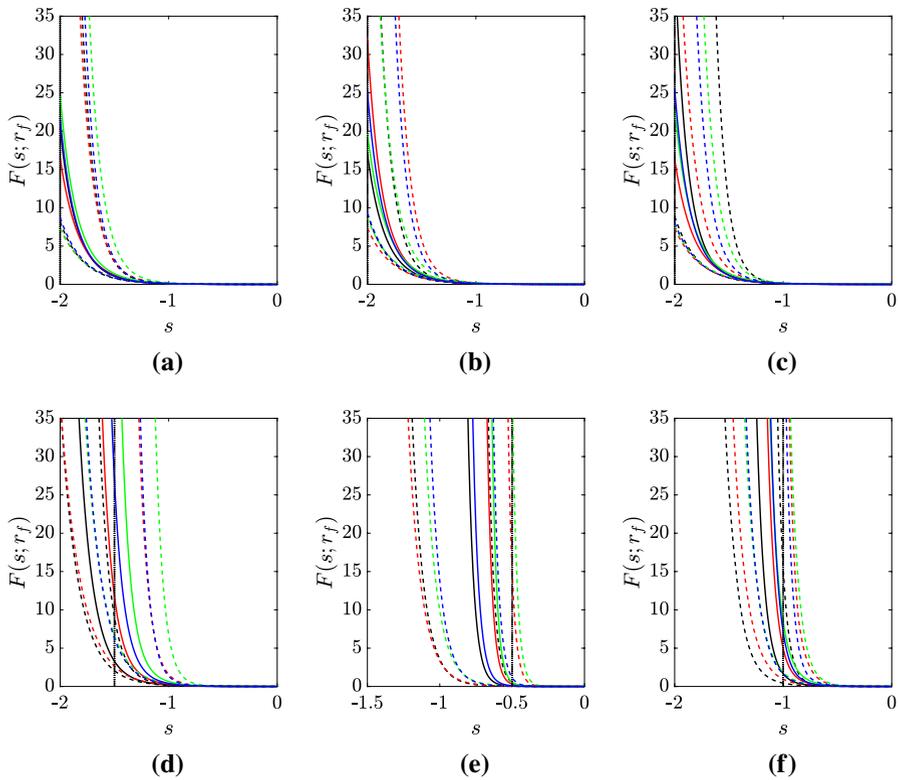
Fig. 2 Envelopes of the test function $F(s;r_f)$ for **a** model 1, **b** model 2, **c** model 3, **d** model 4, **e** model 5 and **f** model 6 under different parameters (W, λ) (black line $([0, 10]^2, 10)$, red line $([0, 10]^2, 5)$, green line $([0, 20]^2, 1)$ and blue line $([0, 30]^2, 1)$). The solid lines indicate the means of $F(s;r_f)$. The dashed lines show the lower and upper bounds of the envelopes. The dotted black lines show the corresponding pole orders. The orders of the three models are 1.5, 0.5 and 1, respectively

window size or the intensity decreases, i.e., as the number of points decreases. For the cases of simulated patterns without pole, all the $F(s;r^{(6)})$ have no pole at $s > -2$, i.e., the F-approach indicates correctly that the pair correlation functions of models 1, 2 and 3 have no pole at $r = 0$. For model 4, the steep increases in mean of $F(s, r_f)$ start near $s = 2 - \gamma$ as expected. For model 5 and 6, the steep increases start at some s greater than $2 - \gamma$.

5.1.2 *K*-approach

The *K*-approach estimates the pole orders by applying the least square method to the empirical *K*-functions. Table 2 shows the means and standard deviations (in brackets) of the estimated pole orders $\hat{\gamma}$ for all simulated patterns under different values of r_k. Both the means and standard deviations decrease when $r^{(k)}$ is large because the interaction between points vanishes at large r.

Table 2 Means and standard deviations (in brackets) of the estimated $\hat{\gamma}$ by the K-function approach

Model	γ	Window	λ	r_k					
				$r^{(1)}$	$r^{(2)}$	$r^{(3)}$	$r^{(4)}$	$r^{(5)}$	$r^{(6)}$
1	0	$[0, 10]^2$	10	0.31 (.17)	0.11 (.14)	0.07 (.09)	0.48 (.07)	0.04 (.06)	0.02 (.03)
			5	0.13 (.18)	0.09 (.14)	0.09 (.11)	0.53 (.12)	0.07 (.09)	0.03 (.04)
		$[0, 20]^2$	1	0.15 (.20)	0.12 (.15)	0.08 (.11)	0.40 (.39)	0.08 (.10)	0.03 (.05)
		$[0, 30]^2$		0.13 (.18)	0.09 (.14)	0.07 (.10)	0.24 (.26)	0.06 (.08)	0.02 (.04)
2	0	$[0, 10]^2$	10	0.31 (.17)	0.12 (.15)	0.07 (.10)	0.49 (.06)	0.04 (.06)	0.02 (.03)
			5	0.11 (.16)	0.10 (.13)	0.08 (.11)	0.51 (.12)	0.06 (.08)	0.03 (.05)
		$[0, 20]^2$	1	0.14 (.17)	0.10 (.13)	0.07 (.10)	0.45 (.40)	0.07 (.09)	0.07 (.07)
		$[0, 30]^2$		0.16 (.17)	0.11 (.15)	0.09 (.10)	0.29 (.31)	0.05 (.06)	0.06 (.05)
3	0	$[0, 10]^2$	10	0.31 (.18)	0.14 (.14)	0.11 (.12)	0.50 (.06)	0.05 (.07)	0.03 (.04)
			5	0.13 (.17)	0.09 (.14)	0.08 (.10)	0.50 (.10)	0.07 (.09)	0.04 (.05)
		$[0, 20]^2$	1	0.15 (.20)	0.10 (.13)	0.10 (.11)	0.35 (.37)	0.10 (.11)	0.03 (.05)
		$[0, 30]^2$		0.14 (.19)	0.10 (.13)	0.08 (.10)	0.33 (.32)	0.05 (.07)	0.03 (.03)
4	0.5	$[0, 10]^2$	10	0.63 (.15)	0.46 (.21)	0.41 (.17)	0.46 (.22)	0.50 (.07)	0.50 (.04)
			5	0.48 (.25)	0.48 (.21)	0.52 (.15)	0.48 (.25)	0.70 (.07)	0.68 (.07)
		$[0, 20]^2$	1	0.56 (.25)	0.60 (.19)	0.75 (.11)	0.66 (.17)	1.18 (.05)	0.95 (.08)
		$[0, 30]^2$		0.52 (.28)	0.57 (.19)	0.61 (.13)	0.68 (.10)	1.18 (.03)	0.95 (.06)
5	1.5	$[0, 10]^2$	10	1.45 (.15)	1.52 (.06)	1.53 (.05)	1.53 (.04)	1.18 (.05)	0.64 (.06)
			5	1.52 (.10)	1.54 (.07)	1.54 (.04)	1.55 (.05)	1.26 (.08)	0.70 (.08)
		$[0, 20]^2$	1	1.53 (.08)	1.55 (.07)	1.59 (.04)	1.59 (.04)	1.42 (.08)	0.81 (.10)
		$[0, 30]^2$		1.51 (.09)	1.52 (.06)	1.53 (.05)	1.58 (.03)	1.41 (.04)	0.82 (.06)
6	1	$[0, 10]^2$	10	1.08 (.12)	1.10 (.13)	1.12 (.11)	1.13 (.11)	1.07 (.08)	0.87 (.08)
			5	1.11 (.17)	1.15 (.14)	1.16 (.13)	1.16 (.13)	1.12 (.11)	0.99 (.11)
		$[0, 20]^2$	1	1.13 (.19)	1.15 (.18)	1.20 (.13)	1.21 (.12)	1.28 (.08)	1.22 (.08)
		$[0, 30]^2$		1.09 (.17)	1.12 (.15)	1.15 (.13)	1.23 (.07)	1.28 (.05)	1.23 (.06)

When the models have zero pole order, using the small pre-chosen distances $r^{(1)}$ and $r^{(4)}$ results in relatively large estimated $\hat{\gamma}$. The estimated orders $\hat{\gamma}$ become closer to zero when $r^{(k)}$ increases. In the case of models with real poles, small pre-chosen distances, $r^{(1)}$, $r^{(2)}$, $r^{(3)}$ and $r^{(4)}$, produce estimated pole orders that are close to the true orders. The estimated orders deviate from the true orders at large pre-chosen distances, $r^{(5)}$ and $r^{(6)}$. The distance $r^{(3)}$ gives relatively good estimates to the pole order for all models. Thus, it is better to choose $r^{(k)}$ based on pairwise distances and we recommend to use $r^{(k)} = r^{(3)}$.

6 Real data examples

6.1 Longleaf pine trees

The longleaf data set, which is accessible from R under the package spatstat, was frequently studied in the statistical literature. The locations of 584 longleaf pine (*Pinus palustris*) trees were recorded from a 200 m × 200 m region in the Wade Tract, an old-growth forest in Thomas County, Georgia. The data were collected by W. J. Platt and S. L. Rathbun with the support of Tall Timbers Research Station, Tallahassee, Florida (see Rathbun 1990; Cressie 1991). The corresponding point pattern shows a high degree of clustering, which is caused by frequent ground fires and recruitment primarily in open spaces created by the decline of large trees.

Despite the non-stationarity of the pattern (see Cressie 1991, p. 600; Guan 2008; Chiu and Liu 2013), some stationary cluster processes offer good fits. Stoyan and Stoyan (1996), Mecke and Stoyan (2005) and Tanaka et al. (2008) fitted generalized Thomas process models, which contain two types of clusters, to the data and found that it yields a better fit than the classical Thomas process. Ghorbani (2013) modeled the data set by a Cauchy cluster process and tested the goodness-of-fit by the Cramér-von Mises test and the empty space function.

Figure 3a shows the empirical pair correlation functions $\hat{g}_k(r)$, $\hat{g}_d(r)$ and $\hat{g}_c(r)$ for the longleaf pine data; the curve for $\hat{g}_k(r)$ is the same as that in Fig. 1. It looks as having a pole at $r = 0$ m. Is there perhaps really a pole? Are the models fitted not suitable and overlook the pole?

To answer these questions, we applied the methodology of Sect. 3. First, we determined the corresponding test function $F(s;r_f)$. The two used values $r_f = 0.92$ and $r_f = 1.28$ were the 50th and 100th smallest pairwise distances, respectively. The
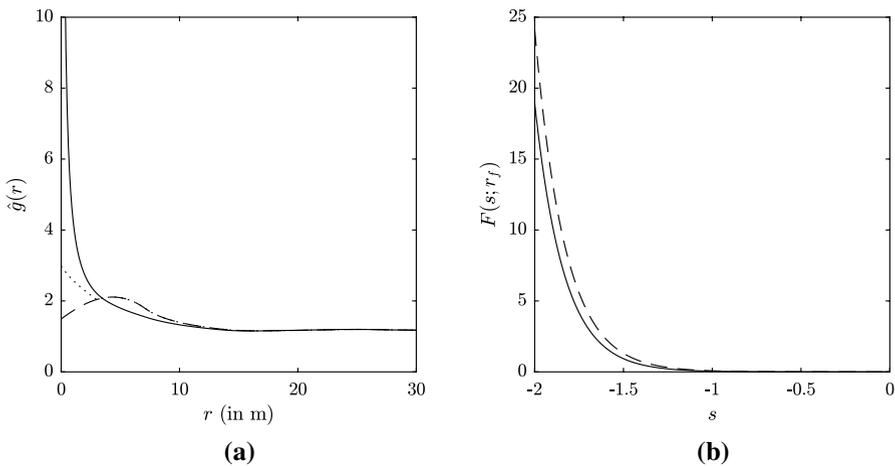
Fig. 3 Plots of the **a** empirical pair correlation functions (line $\hat{g}_k(r)$, dashed line $\hat{g}_d(r)$ and dotted line $\hat{g}_c(r)$) and **b** $F(s;r_f)$ for the 584 longleaf pine trees for two values of r_f (line 0.92 and dashed line 0.28). The bandwidth of the empirical pair correlation functions used in the estimation is 3 m

plot in Fig. 3b shows that the pair correlation function has no pole at $r = 0$: $F(s; r_f)$ has a pole at $s = -2$, which means that the order γ of the pole of the PCF at $r = 0$ is zero. The least-squares fit based on the K-function supports this statement: With $r_k = 0.92$ and $r_k = 1.28$, we obtained pole order estimates equal to zero because of the restriction in the least square method. When we changed the r_k to 1.92 we obtained $\gamma = 0.079$.

In the interpretation of the estimated pole orders $\hat{\gamma}$, we have to consider that the sample size of 584 trees is perhaps too small for a reliable estimation of the K-function. Thus, we conclude that also the estimated order $\hat{\gamma} = 0$ indicates that the pair correlation function of a possible model behind the longleaf pines should not have a pole. Thus in this point, we have full agreement with the fitted models considered in the literature. And because we consider trees in a natural forest, there is no reason for clustering with very small inter-tree distances or points on curves.

6.2 Nonmetallic inclusions in steel

The damage properties of two-phase materials depend heavily on their microstructure. Often the second phase consists of many small objects, inclusions or pores, which can be described as points. The corresponding point patterns are random, often clustered. As already noted by Brechet (1994), the geometrical form of these clusters plays a crucial role, see also Buryachenko (2007, p.3 and 151).

In a study reported in Seleznev et al. (2018), clusters of inclusions in 42CrMo4 steel were statistically analyzed, and their number and spatial distribution were considered. Figure 4a shows one of the 90 analyzed point patterns, a planar section through a steel sample, called 'Section C' in Seleznev et al. (2018). Section C contains 1564 inclusions. Because of the large scaling, sometimes a chain-like cluster can only be seen as a point in the plot of the whole pattern. Figure 4b shows the corresponding empirical pair correlation functions $\hat{g}_k(r)$, $\hat{g}_d(r)$ and $\hat{g}_c(r)$, which are here, in contrast to the case of the pines, almost identical. The functions start from very high values for small r and eventually come close to the value one after $r = 50\,\mu m$.

Visual inspection of Fig. 4a shows chain-like clusters and supports the assumption that the corresponding pair correlation function may have a pole at $r = 0\,\mu m$. We applied the methodology of Sect. 3 to look for a pole.

The plot of the curves of $F(s; r_f)$ with $r_f = 400\,\mu m$ for the 90 specimens in Fig. 4c shows the great variability of the patterns, here expressed by the $F(s; r_f)$ curves. It suggests that the pole of $F(s; r_f)$ is between $s = -1.6$ and $s = -1.3$, which leads to an estimate of the order γ of the pair correlation function pole between 0.4 and 0.7.

The means of the pole orders estimated by the $K(r)$ approach with $r_f = 400\,\mu m$ and $r_f = 800\,\mu m$ are 0.90 and 0.54, respectively.

Thus, we conclude that a possible mathematical model for the inclusion patterns should have a pair correlation function with a pole at $r = 0\,\mu m$. To find a suitable model is of course not the topic of this paper. Possible models are: a generalization of the Cox segment process where the 1D Poisson processes are replaced by renewal
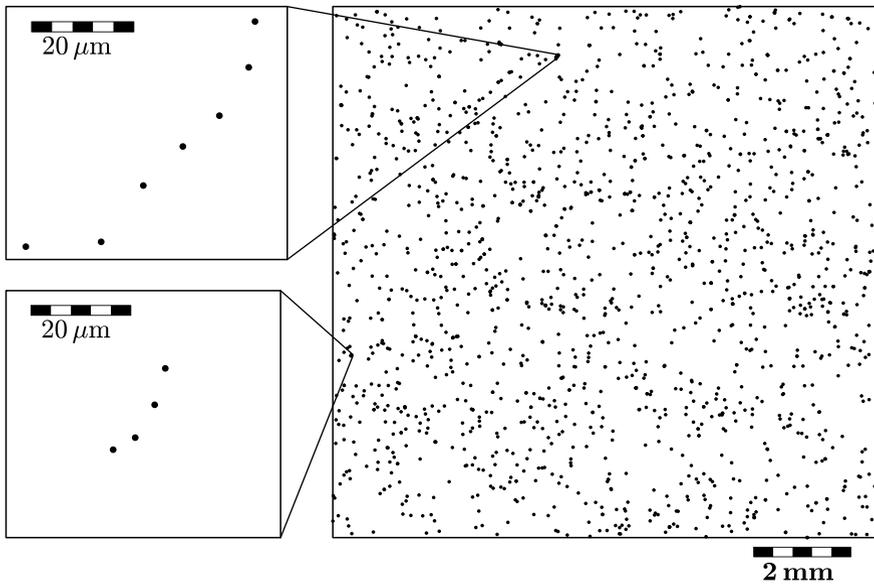
(a)
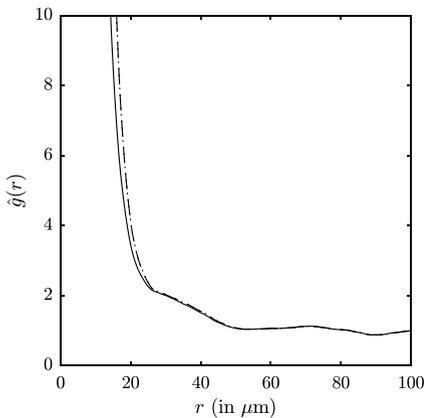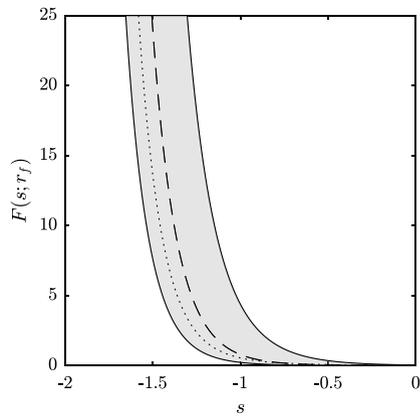
(b)   **(c)**

Fig. 4 Plots of the **a** locations of nonmetallic inclusions in section C of Seleznev et al. (2018), **b** empirical pair correlation functions (line $\hat{g}_k(r)$, dashed line $\hat{g}_d(r)$ and dotted line $\hat{g}_c(r)$)) of Section C and **c** envelopes of the $F(s;r_f)$ curves of the 90 specimens with $r_f = 400\,\mu$m. The bandwidth of the empirical pair correlation functions used in the estimation is 5 μm. The dashed line in (**c**) represents the mean, the dotted line represents $F(s;r_f)$ of section C

processes or Cox segment processes with a length probability density function with a pole at $r = 0$. A detailed discussion of these models is deferred to future work.

We mention that we discussed only planar sections, while of course the steel samples are three-dimensional. Therefore, if no 3D samples are available, stereological methods should be used in order to estimate the pair correlation function of

the three-dimensional pattern. Such methods are discussed in Chiu et al. (2013, section 10.7.2). Perhaps the approximation (10.86) given there is true also for the inclusions; if so this would lead to a pole order of the spatial pair correlation function similar to the values given above for the planar case. But this is only speculation and needs further research.

References

Baddeley, A., Rubak, E., Turner, R. (2015). *Spatial point patterns*: *Methodology and applications with R*. Boca Raton: Chapman and Hall/CRC.

Brechet, Y. J. M. (1994). Clusters, plasticity and damage: A missing link? *Materials Science and Engineering: A*, *175*(1–2), 63–69.

Buryachenko, V. (2007). *Micromechanics of heterogeneous materials*. New York: Springer.

Chiu, S. N., Liu, K. I. (2013). Stationarity tests for spatial point processes using discrepancies. *Biometrics*, *69*(2), 497–507.

Chiu, S. N., Stoyan, D., Kendall, W. S., Mecke, J. (2013). *Stochastic geometry and its applications* 3rd ed. Chichester: Wiley.

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

Daley, D. J., Vere-Jones, D. (2008). *An introduction to the theory of point processes*: *Volume II*: *General theory and structure*. New York: Springer.

Davis, M., Peebles, P. J. E. (1983). A survey of galaxy redshifts. V-The two-point position and velocity correlations. *The Astrophysical Journal*, *267*, 465–482.

Gelfand, A. E., Diggle, P., Fuentes, M., Guttorp, P. (2010). *Handbook of spatial statistics*. Boca Raton: Chapman & Hall/CRC.

Ghorbani, M. (2013). Cauchy cluster process. *Metrika*, *76*(5), 697–706.

Guan, Y. (2007). A least-squares cross-validation bandwidth selection approach in pair correlation function estimations. *Statistics & Probability Letters*, *77*(18), 1722–1729.

Guan, Y. (2008). A kpss test for stationarity for spatial point processes. *Biometrics*, *64*(3), 800–806.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*, Vol. 70. Chichester: Wiley.

Martinez, V. J., Saar, E. (2002). *Statistics of the galaxy distribution*. Boca Raton: Chapman & Hall/CRC.

Mecke, K. R., Stoyan, D. (2005). Morphological characterization of point patterns. *Biometrical Journal*, *47*(4), 473–488.

Peebles, P. J. E. (1974). The nature of the distribution of galaxies. *Astronomy & Astrophysics*, *32*, 197–202.

Rathbun, S. L. (1990). Estimation and statistical inference for space-time point processes. Ph.D. thesis, Iowa State University, Ames.

Seleznev, M., Wong, K. Y., Stoyan, D., Weidner, A., Biermann, H. (2018). Cluster detection of non-metallic inclusions in 42CrMo4 steel. *Steel Research International*, *89*(11), 1800216.

Stoyan, D. (1983). Inequalities and bounds for variances of point processes and fibre processes. *Series Statistics*, *14*(3), 409–419.

Stoyan, D. (1994). Caution with fractal point patterns!. *Statistics*, *25*(3), 267–270.

Stoyan, D., Stoyan, H. (1996). Estimating pair correlation functions of planar cluster processes. *Biometrical Journal*, *38*(3), 259–271.

Tanaka, U., Ogata, Y., Stoyan, D. (2008). Parameter estimation and model selection for Neyman–Scott point processes. *Biometrical Journal*, *50*(1), 43–57.

Totsuji, H., Kihara, T. (1969). The correlation function for the distribution of galaxies. *Publications of the Astronomical Society of Japan*, *21*, 221–229.