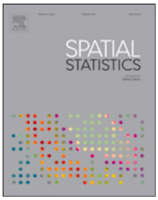
ARTICLE IN PRESS

Spatial Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Spatial Statistics

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



Testing biodiversity using inhomogeneous summary statistics and global envelope tests

M.C. de Jongh[b], M.N.M. van Lieshout[a,b,*]

[a] *CWI, P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands*

[b] *Department of Applied Mathematics, University of Twente, P.O. Box 217, NL-7500 AE Enschede, The Netherlands*

ARTICLE INFO

Article history:
Received 23 August 2021
Received in revised form 3 January 2022
Accepted 11 January 2022
Available online xxxx

Keywords:
Biodiversity
Clustering within species
Independence between species
Inhomogeneous summary statistics
Spatial point pattern
Rank and global envelope tests

ABSTRACT

We discuss recent methodological developments in inhomogeneous summary statistics and envelope testing and study their applications in the field of spatial point pattern analysis. Specifically, we use these methods to test McGill's theory of biodiversity. This theory is based upon three axioms: individuals of the same species cluster together, many rare species co-exist with a few common ones and individuals of different species grow independently of each other. We assess the validity of the first and third axioms for data obtained from a 50 hectare plot on Barro Colorado Island.

1. Introduction

Ecology is concerned with understanding plant and animal communities (Ricklefs, 1990). Of particular interest are the mutual relations that exist between individuals and the way in which the plants or animals interact with their natural environment. The spatial or dynamic patterns that arise from biological mechanisms (e.g. competition, cooperation, reproduction and mortality) or climate and habitat factors such as the amount of sunlight or precipitation, the availability and spatial distribution of nutrients, the type of terrain or properties of the soil, tend to be complex. Therefore, since the human eye is not very good in assessing interaction patterns that go beyond plain trends and second-order aspects (Ripley, 1976), ecologists have long looked towards spatial statistics (Diggle, 1983; Gelfand et al., 2010; Illian et al., 2008; Lieshout, 2019) for the tools that are

---

[*] Corresponding author at: CWI, P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands.
E-mail address: colette@cwi.nl (M.N.M. van Lieshout).

https://doi.org/10.1016/j.spasta.2022.100607

Please cite this article as: M.C. de Jongh and M.N.M. van Lieshout, Testing biodiversity using inhomogeneous summary statistics and global envelope tests. Spatial Statistics (2022) 100607, https://doi.org/10.1016/j.spasta.2022.100607.

needed to quantify interaction patterns as well as to formulate and test hypotheses (Gelfand et al., 2019; Wiegand and Moloney, 2014).

Ecological data come in various formats. A very old technique is quadrat counting (Du Rietz, 1929) in which the number of plants, say, that fall in quadrats placed over the sampling window are recorded. The classic index of dispersion (Fisher et al., 1922) is widely used to test spatial randomness and scaling by the area of the quadrats readily gives an estimator of the intensity. An obvious disadvantage is that the results depend on the sizes, shapes and number of quadrats. Other sampling schemes are based on recording distances, for example from a given reference point or between nearest neighbours. For a survey, we refer the reader to Diggle (1983) or to Szmyt (2014).

The sparse sampling methods mentioned above are quite limited in that they record only a fraction of the available information. Thus, increasingly data come in the form of a mapped point pattern such as shown in Fig. 1. In other words, the locations of all individuals are recorded. This approach is superior to sparse sampling but also more laborious and expensive. Exploratory analysis for mapped patterns relies on functional statistics that capture various aspects of the underlying pattern including tendencies to seek out or avoid other individuals, both within the same species and between species. For example Diggle (1983) considers the distribution function of the distance between nearest-neighbours or the size of the voids, Lieshout and Baddeley (1996) compare them to one another and Ripley (1976, 1977) looks at the expected number of further points within the vicinity of a typical point. Traditionally such functional statistics were developed under strong assumptions of homogeneity. In ecology, for such an assumption to hold one would need to restrict attention to featureless terrains, which is highly restrictive. However, methodological developments in recent decades substantially reduced the need for restrictions (Baddeley et al., 2000; Lieshout, 2011) of this kind. In ecology, where it is typical to have combinations of inhomogeneity and interaction and it is important to prise them apart, the potential is enormous. In this paper, we discuss recently developed methods that circumvent the assumption of homogeneity and use these methods to study the validity of McGill's axioms of biodiversity, which are phrased in terms of interactions within a species and between species. Failing to take into account inhomogeneity can lead to false conclusions in the sense that, for example, preference for different regions may be taken for negative correlation (Volkov et al., 2009) where in fact the species are independent (Lieshout, 2018b). To the best of our knowledge, this is the first work that provides evidence for the correctness of McGill's axioms that does not rely on the assumption of homogeneity.

Functional summary statistics can be used to test hypotheses, but, since their distribution tends to be unknown, Monte Carlo ideas based on simulations need to be applied (Besag and Clifford, 1989; Besag and Diggle, 1977; Ripley, 1977). Envelope tests with their graphical representations are particularly popular in ecology (Kenkel, 1988). However, naive users have the unfortunate habit of falling into a multiple testing trap, even though Ripley (1977, page 181) already warned against this. The issue was forcefully brought to the attention of ecologists in an influential paper by Loosmore and Ford (2006). However, the paper seems to have sparked the new misunderstanding that all envelope tests are invalid (Law et al., 2009). Fortunately, it also led to a new wave of ideas, including tackling uneven envelope widths and asymmetry (Myllymäki et al., 2015) as well as the incorporation of ideas from functional analysis on ordering curves by their functional depth (Myllymäki et al., 2016; Ramsay and Silverman, 2006).

Some other recent trends in point process theory, such as the development of models on non-Euclidean spaces including graphs (Anderes et al., 2020; Lieshout, 2018a), linear networks (Baddeley et al., 2021) or spheres (Møller and Rubak, 2016) so far seem to have had little impact on ecology, although determinantal point processes (Lavancier et al., 2015; Macchi, 2017) offer great potential for design-based sampling as well as numerical integration (Belhadji et al., 2019).

The plan of this paper is as follows. In Section 2, we recall the classic definitions of the inhomogeneous K- and J-function and their cross versions and indicate how inhomogeneity can be taken into account. Kernel estimators of the spatial intensity function form the topic of Section 3. We then move on to discuss Monte Carlo testing in Section 4 and explain the differences between pointwise and global envelopes. To demonstrate how developments in functional summary statistics and envelope testing may impact ecology, we discuss McGill's axioms for biodiversity in Section 5. Section 6 is devoted to intra-specific interactions, whilst Section 7 deals with inter-specific interactions. The results in these sections form compelling evidence for the validity of McGill's axioms and

(a) Locations of *Hybanthus prunifolius* ('hybapr') trees.

(b) Locations of *Coussarea curvigemmia* ('cou2cu') trees.

Fig. 1. Locations of alive trees in the eighth census on Barro Colorado Island for some selected species.

demonstrate the potential of inhomogeneous summary statistics. We will summarize our findings in Section 8.

2. Functional summary statistics

A point process (Diggle, 1983; Gelfand et al., 2010; Illian et al., 2008; Lieshout, 2019) is a random mechanism for generating finite patterns in some bounded observation window $W \subset \mathbb{R}^2$. Let X denote such a point process in $\mathbb{R}^2$. Then X is called *stationary* if its probability distribution is invariant under translations. Intuitively speaking, this means that the output of the random mechanism looks the same in terms of the point density and the inter-point interactions from every viewpoint in $\mathbb{R}^2$.

Structural information about point processes is often expressed by means of summary statistics (Diggle, 1983; Ripley, 1981). The simplest one for stationary X is the *intensity* λ, the average number of points per unit area. When $\lambda > 0$, one measure of interaction is the K-function (Ripley, 1976) defined so that $\lambda K(r)$ is the expected number of further points seen within a distance $r \geq 0$ of a typical point. In the absence of any interaction between the points, $K(r) = \pi r^2$; larger values indicate a clustering of points, whereas smaller values are indicative of repulsion.

The K-function can be estimated by

$$\hat{K}(r) = \frac{1}{|W|} \sum_{x_i \in X \cap W} \sum_{x_j \in X \cap W \setminus \{x_i\}} \frac{w_{x_i,x_j} 1(\|x_i - x_j\| \leq r)}{\lambda^2}.$$

Here w_{x_i,x_j} is a weight factor to compensate for edge effects (Baddeley et al., 2016; Diggle, 1983) and $1(\|x_i - x_j\| \leq r)$ takes the value one if the distance between x_i and x_j is at most r, zero otherwise.

If X is not stationary, the abundance of points varies over space. A version of the K-function that compensates for varying point intensities was developed by Baddeley et al. (2000). Under appropriate technical conditions, this *inhomogeneous K-function* is given by

$$K_{\text{inhom}}(r) = \frac{1}{|W|} \mathbb{E} \left[\sum_{x_i \in X \cap W} \sum_{x_j \in X \cap W \setminus \{x_i\}} \frac{1(\|x_i - x_j\| \leq r)}{\lambda(x_i)\lambda(x_j)} \right], \quad r \geq 0. \tag{1}$$

The term $\lambda(x_i)\lambda(x_j)$ accounts for the habitat associations, so that K_{inhom} is solely a measure of the direct interactions. Again the benchmark curve is that of πr^2 when there is no interaction, larger values suggest clustering, smaller ones repulsion.

An estimator of the inhomogeneous K-function follows directly from the definition (1). Indeed,

$$\hat{K}_{\text{inhom}}(r) = \frac{1}{|W|} \sum_{x_i \in X \cap W} \sum_{x_j \in X \cap W \setminus \{x_i\}} \frac{w_{x_i,x_j} 1(\|x_i - x_j\| \leq r)}{\lambda(x_i)\lambda(x_j)}. \tag{2}$$

Other popular summary statistics used in ecology (Gelfand et al., 2019; Wiegand and Moloney, 2014) include the *empty space function* $F(r)$ defined for stationary point processes X as the probability of finding a point of X within distance $r \geq 0$ of the origin (Ripley, 1977) and, especially, the *nearest-neighbour distance distribution function* $G(r)$ (Diggle, 1978, 1983) defined as the probability that the distance from an arbitrary point of X to the nearest other point is at most r. Both statistics can be expressed in terms of the generating functional Z (Daley and Vere-Jones, 2008). To see this, let us consider the empty space function. Write $B(0, r)$ for the closed ball with radius r centred at the origin. Then

$$1 - F(r) = \mathbb{P}(X \cap B(0, r) = \emptyset) = \mathbb{E}\left[\prod_{x_i \in X} 1(x_i \notin B(0, r))\right]. \tag{3}$$

Eq. (3) states that $1 - F(r)$ is equal to $Z(1 - u_r)$, the generating functional evaluated at $1 - u_r$ for the function u_r that maps a point x to $u_r(x) = 1(x \in B(0, r))$. Similarly, for the nearest-neighbour distance distribution function, we have

$$1 - G(r) = \mathbb{P}(X \cap B(0, r) \setminus \{0\} = \emptyset \mid 0 \in X) = \mathbb{E}\left[\prod_{x_i \in X\setminus\{0\}} 1(x_i \notin B(0, r)) \,\middle|\, 0 \in X\right].$$

The conditioning on X having a point at the origin can be made rigorous by means of the reduced Palm distribution, denoted by $P^{!0}$, and the corresponding generating functional, $Z^{!0}$. With this notation, $1 - G(r) = Z^{!0}(1 - u_r)$. Finally, the J-function is defined as $J(r) = (1 - G(r))/(1 - F(r))$, provided that the denominator is non-zero. To interpret this statistic, recall that a clustered pattern is characterized by small distances between neighbouring points and large empty spaces. Thus, for a clustered pattern J takes values smaller than one. For a regular pattern on the other hand, the nearest-neighbour distances tend to be large and J exceeds 1. For a Poisson point process, $Z = Z^{!0}$ so $J(r) = 1$ for all $r \geq 0$.

Extensions to inhomogeneous point processes were constructed by Lieshout (2011) under appropriate technical conditions. Using the representation in terms of generating functionals, set

$$F_{\text{inhom}}(r) = 1 - Z(1 - \tilde{u}_r) \text{ and } G_{\text{inhom}}(r) = 1 - Z^{!0}(1 - \tilde{u}_r), \tag{4}$$

where $\tilde{u}_r(x) = (\inf_x \lambda(x))\, 1(x \in B(0, r))/\lambda(x)$. As for the K-function, the weighting by $\lambda(x)$ accounts for the fact that the likelihood of observing points of X varies over space. The infimum ensures that $\tilde{u}_r(x)$ takes values in $[0, 1]$ as required in the definition of generating functionals (Daley and Vere-Jones, 2008). The related inhomogeneous J-function J_{inhom} is simply the ratio of $1 - G_{\text{inhom}}$ and $1 - F_{\text{inhom}}$, provided the latter is non-zero. It takes the value 1 in the absence of interaction. Larger values suggest a regular pattern, values smaller than one indicate clustering.

In order to estimate the inhomogeneous F- and G-functions, Lieshout (2011) proposed the following estimators for $Z(1 - \tilde{u}_r)$ and $Z^{!0}(1 - \tilde{u}_r)$:

$$Z\widehat{(1 - \tilde{u}_r)} = \frac{1}{\#L \cap W_{\ominus r}} \sum_{l_i \in L \cap W_{\ominus r}} \prod_{x_j \in X \cap B(l_i, r)} \left[1 - \frac{\inf \lambda}{\lambda(x_j)}\right] \tag{5}$$

and

$$Z^{!0}\widehat{(1 - \tilde{u}_r)} = \frac{1}{\#X \cap W_{\ominus r}} \sum_{x_i \in X \cap W_{\ominus r}} \prod_{x_j \in X\setminus\{x_i\} \cap B(x_i, r)} \left[1 - \frac{\inf \lambda}{\lambda(x_j)}\right] \tag{6}$$

where $L \subset W$ is a finite point grid and $W_{\ominus r}$ is the eroded set $\{x \in W : x + B(0, r) \subset W\}$ containing the points of W whose distance to the boundary of W is at least r.

It is worth emphasizing that the various summary statistics capture different aspects of the data. The inhomogeneous K-function is a second-order statistics: it can be shown that its series expansion relies only on the joint probability densities of finding points of X at location pairs. The inhomogeneous F- and G-function rely on such densities for n-tuples of all orders $n \in \{2, 3, \dots\}$ (Lieshout, 2011). In practice it is wise to consider both a second-order statistic and an infinite-order statistic. In this paper, we use the inhomogeneous K- and J-functions.

In studying inter-species interactions, we make use of the cross versions of summary statistics that describe associations between points of different types. The classic cross K-function, $K_{1,2}(r)$, for example, is proportional to the number of points of type 2 seen within distance $r \geq 0$ from a typical point of type 1 and can be defined in a similar fashion as the univariate K-function upon weighting a point of type $l \in \{1, 2\}$ by λ_l, the average number of points of type l per unit area. To define the inhomogeneous cross K-function, simply allow $\lambda_l(x)$ to be spatially varying (Møller and Waagepetersen, 2004). To interpret the statistic, note that when points of type 2 are mostly found in the r-vicinity of points labelled 1, $K_{\text{inhom; } 1,2}(r)$ takes values larger than πr^2.

An inhomogeneous cross version of the J-function was introduced in Lieshout (2011) and studied in Cronie and Lieshout (2015). The idea again is to exploit the connection to generating functionals. Write X_l for the marginal point process that consists of the points in X having label l. Then

$$1 - G_{\text{inhom; } 1,2}(r) = \mathbb{E}\left[\prod_{x_i \in X_2} (1 - \tilde{u}_{2,r}(x_i)) \,\middle|\, 0 \in X_1\right] \tag{7}$$

for $\tilde{u}_{2,r}(x) = (\inf \lambda_2) 1(x \in B(0, r))/\lambda_2(x)$. With the notation $F_{\text{inhom; } 2}$ for the empty space function of X_2, the inhomogeneous cross J-function is given by $J_{\text{inhom; } 1,2} = (1 - G_{\text{inhom; } 1,2}(r))/(1 - F_{\text{inhom; } 2}(r))$ whenever the denominator is non-zero. When X_1 and X_2 are independent, $J_{\text{inhom};1,2} = 1$. Estimators in the spirit of (2) and (5)–(6) are straightforward. For further technical details and underlying assumptions, we refer to the literature cited above.

3. Estimating the spatially varying intensity function

The estimators (2), (5) and (6) all depend on the spatially varying intensity function λ. Since it is generally unknown, in practice it must be estimated. For this purpose, Diggle (1985) suggested a non-parametric kernel estimator. Recall that a kernel is a non-negative real-valued probability density function κ which satisfies symmetry requirements. Then

$$\hat{\lambda}_\sigma(x) = \sum_{x_i \in X \cap W} \frac{1}{c_{W,\sigma}(x, x_i)} \kappa_\sigma(x - x_i). \tag{8}$$

Here, W denotes the bounded observation window through which the point process X is observed and

$$\kappa_\sigma(x) = \sigma^{-2} \kappa(\sigma^{-1} x), \;\; x \in W. \tag{9}$$

The factor σ is a smoothing parameter called the *bandwidth*. Furthermore, $c_{W,\sigma}(x, x_i)$ is an edge correction factor. In this paper, we shall use the local correction (Lieshout, 2012) given by

$$c_{W,\sigma}(x, x_i) = \int_W \kappa_\sigma(s - x_i) \mathrm{d}s.$$

There is a great variety of possible choices for the kernel. Popular choices include the Gaussian, Beta and Epanechnikov kernels. Here we shall use the smoothest of the aforementioned options, the Gaussian kernel.

As already observed by Diggle (1985) an appropriate value of the bandwidth is essential for a successful estimate. Various techniques have been developed for selecting a proper value of the bandwidth (Diggle, 1985; Loader, 1999; Scott, 1992). We use Cronie and Van Lieshout's criterion (Cronie and Lieshout, 2018) that selects the bandwidth σ that, under the assumption that λ takes strictly positive values, minimizes the expression

$$\text{CvL}(\sigma) = \left(|W| - \sum_{x_i \in X \cap W} \frac{1}{\hat{\lambda}_\sigma(x_i)}\right)^2, \tag{10}$$

where the sum is taken over all data points x_i and $\hat{\lambda}(x_i)$ denotes the kernel estimate of the intensity at x_i with bandwidth σ. Advantages of this approach are the facts that it is completely non-parametric, does not involve any higher-order moments and does not require numerical integration.

4. Rank and global envelope tests

To assess whether an estimated functional summary statistic T, for example an estimator of the classic K-function, deviates significantly from its value under a null hypothesis, say that of *complete spatial randomness* for specificity, ecologists often use Monte Carlo envelope tests (Besag and Clifford, 1989; Besag and Diggle, 1977; Ripley, 1977).

The procedure is as follows. First generate s realizations under the null hypothesis and compute T for each realization to obtain the function cloud $T_i(r)$ for $i = 2, 3, \ldots, s+1$. Here the notation $T_i(r)$ is used for the statistic T evaluated on realization i for r in a given interval I. The band that is bounded by the point-wise minimum and maximum of the T_i is called a *point-wise envelope* and can be displayed graphically. The graph of $T_{\text{obs}}(r)$, that is, the statistic T evaluated on the data pattern, is plotted too for comparison. The interval I ideally should be determined by the range of interaction. In practice, I is usually selected by a combination of knowledge about the size of the region of interest, the statistic at hand and problem-specific knowledge. In testing McGill's axioms, for example, we take into account the diameters of tree trunks.

Consider a fixed $r_0 \in I$. If the null hypothesis were true, then the random variables $T_{\text{obs}}(r_0), T_2(r_0), \ldots, T_{s+1}(r_0)$ would be independent and identically distributed. Therefore the probability that $T_{\text{obs}}(r_0)$ is the largest (or k-largest for any $k \in \{1, \ldots, s+1\}$) is $1/(s+1)$, with suitable modifications in case of ties. Hence a two-sided Monte Carlo test that rejects the null hypothesis when the observed value $T_{\text{obs}}(r_0)$ does not lie within the open interval bounded by

$$T_{\min}(r_0) = \min\{T_{\text{obs}}(r_0), T_2(r_0), \ldots, T_{s+1}(r_0)\}$$

and

$$T_{\max}(r_0) = \max\{T_{\text{obs}}(r_0), T_2(r_0), \ldots, T_{s+1}(r_0)\}$$

has significance level $2/(s+1)$. Being based on a fixed r_0, such a test is called a *pointwise envelope test*. When deviations in one direction are of interest, a one-sided test may be more appropriate. For example, if one would like to test for clustering within a tree species using an estimator of the K-function, a one-sided test would reject complete spatial randomness for large values. It is worth noting that the assumption of independent samples, can be relaxed to exchangeability (Besag and Clifford, 1989).

In practice, ecologists often plot the point-wise envelopes over the entire interval I of r-values. Such plots are useful as diagnostic tools because they indicate at what ranges the data deviate from the assumed model (Besag and Diggle, 1977; Ripley, 1977). However one cannot simply reject the null hypothesis when the data graph does not lie completely within the point-wise envelopes (Loosmore and Ford, 2006). The reason is that considering all r simultaneously leads to a multiple testing problem and therefore requires adjustment. To take into account a range of r-values properly, a *global envelope test* should be used. Such a test rejects the null hypothesis when

$$\phi_{\text{env}}(T_{\text{obs}}) = 1(\exists r \in I : T_{\text{obs}}(r) \notin (T_{\text{low}}(r), T_{\text{upp}}(r))) \tag{11}$$

is equal to one. Below, we shall describe various approaches to determine the bounds T_{low} and T_{upp} in such a way that (11) has a controlled significance level.

Perhaps the oldest example of a global envelope test is the (MAD) test (Diggle, 1979). It is based upon the *maximum absolute deviation* measure given by

$$U = \max_{r \in I}\{|T(r) - T_0(r)|\}, \tag{12}$$

where $T_0(r)$ denotes the expectation of the statistic T under the null hypothesis. Effectively, T is reduced to a scalar. Hence the statistics U_i, $i = 2, \ldots, s+1$ calculated for a sample from the null model can be ordered and, assuming no ties, a p-value can be calculated from the rank of U_{obs}, the value of U for the data pattern. One-sided versions of the test take into account deviations from $T_0(r)$ in one direction only. In terms of (11), the global envelope test at significance level $1/(s+1)$ rejects the null hypothesis when T_{obs} does not lie completely within the global envelope

$$(T_0(r) - U_{\max},\ T_0(r) + U_{\max})$$

where, as before, $U_{\max}$ is the maximum of the set $\{U_{\text{obs}}, U_2, \ldots, U_{s+1}\}$, again under the proviso that there are no ties. Unlike the point-wise envelope, the global envelope has constant width. Empirical evidence suggests that a modest number s of simulations suffices (Diggle, 1979).

Alternative deviation measures may be used, for example the L_2-deviation given by

$$\int_I (T(r) - T_0(r))^2 \mathrm{d}r,$$

which gives rise to the Diggle–Cressie–Loosmore–Ford (DCLF) test (Loosmore and Ford, 2006). As it is more sensitive to the choice of I, Baddeley et al. (2014) recommend to use the DCLF test when there is prior information about the range of interaction and to use the MAD test otherwise.

The MAD measure given in Eq. (12) fails to capture the behaviour of $T(r)$ at different distances r, unless the distribution of $T(r)$ is the same and symmetric for all values of r. The *studentized MAD envelope test* resolves the first of these issues by studentizing T. Specifically, the upper and lower bounds in (11) are set to

$$\left(T_0(r) - U_{S,\max}\sqrt{\mathrm{var}_0(T(r))},\ T_0(r) + U_{S,\max}\sqrt{\mathrm{var}_0(T(r))}\right) \tag{13}$$

where $\mathrm{var}_0(T(r))$ is the variance of $T(r)$ under the null hypothesis and $U_{S,\max}$ is the maximum of the studentized statistics

$$\max_{r\in I} \left| \frac{T(r) - T_0(r)}{\sqrt{\mathrm{var}_0 T(r)}} \right|$$

evaluated at the data pattern and the simulations. The *directional quantile MAD envelope test* scales the upper and lower bound in (13) by $|\underline{T}(r) - T_0(r)|$ and $|\overline{T}(r) - T_0(r)|$, instead of by the standard deviation, in terms of the r-wise lower and upper 2.5% quantiles $\underline{T}(r)$ and $\overline{T}(r)$ under the null hypothesis. Note that the interval in general is asymmetric. Full details can be found in Myllymäki et al. (2016).

So far, we concentrated on deviation measures. A second type of ideas centre around ordering the test statistics based on their 'extremeness' in the cloud of functions $\{T_{\text{obs}}, T_2, \ldots, T_{s+1}\}$ (López-Pintado and Romo, 2009; Ramsay and Silverman, 2006). For ease of notation, we will write $T_{\text{obs}} = T_1$. Doing so, Myllymäki et al. (2016) define the extreme rank R_i of the ith function $T_i(r)$, $i = 1, \ldots, s+1$, as follows. First, for every $r \in I$, write $T_{(1)}(r) \leq T_{(2)}(r) \leq \cdots \leq T_{(s+1)}(r)$ for the point-wise order statistics and set

$$R_i = \max\{k : T_{(k)}(r) \leq T_i(r) \leq T_{(s+2-k)}(r) \text{ for all } r \in [0, r_0]\}$$

with suitable modification in the case of point-wise ties ($T_i(r) = T_j(r)$ for some $i \neq j$). If $R_i < R_j$, the function T_i is closer to the edge of the function cloud than T_j and therefore more extreme. Rejecting the null hypothesis when the fraction

$$\frac{1}{s+1} \sum_{i=1}^{s+1} 1(R_i \leq R_1) \tag{14}$$

is at most α yields a test with significance level at most α. In other words, the test is conservative. Similarly, a liberal test is obtained by considering R_i strictly less than R_1 (Myllymäki et al., 2016). To obtain a global envelope, note that the liberal test rejects precisely when the graph of $T_1(r) = T_{\text{obs}}(r)$ does not lie completely within the open interval bounded by the point-wise k_α-smallest and k_α-largest of the statistics. The rank k_α is the maximal k for which (14) is at most α. For the conservative test, the envelope is closed so that T_{obs} may touch its boundary.

The difference between the liberal and conservative test is due to ties in the ranks. Therefore a Monte Carlo test based on the value of R_1, the rank of the data curve, cannot return an exact p-value, only a range:

$$(p_1, p_2) = \left(\frac{1}{s+1} \sum_{i=1}^{s+1} 1(R_i < R_1), \quad \frac{1}{s+1} \sum_{i=1}^{s+1} 1(R_i \leq R_1) \right).$$

In other words, the significance level α lies in this interval. To minimize the ambiguity, the length of the interval of p-values, $p_2 - p_1$, should be small. As the length depends on the data, the smoothness of the T_i and the number of simulations, some experimentation is needed. Myllymäki et al. (2016) recommend to use at least 2500 simulations. If such a large number of simulations is not feasible, one may use the MAD test or the scaled MAD test.

We close this section with a few words about the null hypothesis. From a mathematical point of view, complete spatial randomness is formalized as a homogeneous Poisson process (Cressie, 1991; Diggle, 1983; Illian et al., 2008; Lieshout, 2019). Note that this null hypothesis is composite, as the intensity parameter λ is left unspecified. A solution is to condition on the number of points and simulate realizations from a binomial point process with the same number of points as the data pattern. For a hypothesis of absence of interaction, formalized by inhomogeneous Poisson processes, no adequate solution seems to have been found. In practice, one would estimate the intensity function and generate simulations from the fitted model. However, doing so violates the Monte Carlo assumption that data and simulations are exchangeable (Besag and Clifford, 1989), and makes the test invalid, typically conservative. Also its power tends to decrease (Baddeley et al., 2014). Some work has been done to deal with this problem. For example Dao and Genton (2014) designed an adjusted test, but their approach is very computationally intensive. In order to mitigate this issue, for the tree census data considered in this paper, we estimate the intensity function using data from the first census which has the least overlap with the observed pattern. Section 6 provides more details on the procedure.

5. Biodiversity on Barro Colorado Island

The set of laws that govern the distribution and co-existence of species is among the most popular topics of research in the field of ecology (Ricklefs, 1990). In 1934, biologist G.F. Gause stated the *competitive exclusion principle*, which asserts that two species cannot stably co-exist if they compete for the same means of living (Gause, 1934). Yet in tropical rain forests or coral reefs a vast number of different species are encountered within small areas (Gentry, 1988; Huisman and Weissing, 1999; Losos and Leigh, 2004). Over the years, various attempts have been made to explain such observed patterns of biodiversity (Chesson, 2000; Clements, 1916; Gleason, 1926; Hubbell, 2001; Tilman, 1982; Wright, 2002). In 2010, biologist McGill set out to unite six different theories into one (McGill, 2010). He established three axioms that would constitute the basis of his unified theory of biodiversity: individuals of the same species cluster together, many rare species co-exist with a few common ones and individuals of different species grow independently of each other.

McGill's axioms have been subjected to careful scrutiny over the past decade. Various studies showed that interactions within a species (as measured by nearest-neighbour distances or the homogeneous K-function) are generally much stronger than inter-specific interactions (Lieberman and Lieberman, 2007; Perry et al., 2009; Wiegand et al., 2007). In Lieberman and Lieberman (2007) and Perry et al. (2009), nearest-neighbour pairs of trees were tested on conformity to a simple random mixing model. The observed frequencies of each species combination were compared to the expected frequencies under a null hypothesis of random mixing. The results indicate that trees from different species tend to grow independently of one another. In Wiegand et al. (2007), spatial patterns in local diversity were analysed from the perspective of individual species. The analyses were carried out using individual species–area relationships (ISAR), a spatial statistic that measures the expected diversity around an arbitrary individual of some target species. The results imply that individual species only affect diversity structures on small spatial scales. In 2012, Wiegand et al. (2012) used the homogeneous cross K- and G-functions to examine the validity of the third of McGill's axioms. The results of applying these statistics to data from three tropical rain forests reveal that the strength of inter-specific interactions fades with species richness. The authors are aware of the limitations of their method 'because they require complete mapping and because of difficulties in teasing apart two major, yet contrasting factors: habitat association and direct species interactions.' The first of these difficulties was addressed in Lieshout (2018b), where summary statistics for random measures were used to eliminate the requirement of a complete mapping. Our goal in this paper is to address the second problem using the recent methodological developments

(a) Kernel smoothed intensity function of 'hybapr'.

(b) Estimated intensity function for 'hybapr'.

(c) Kernel smoothed intensity function of 'cou2cu'.

(d) Estimated intensity function for 'cou2cu'.

Fig. 2. Comparison of estimated intensity functions.

in the analysis of inhomogeneous spatial point patterns discussed in Section 2 and the Monte Carlo testing procedures described in Section 4. We apply these methods to data on the spatial distribution of stems of a large number of woody trees and shrub species measuring at least 1 cm in diameter found in a 50 hectare plot on Barro Colorado Island, Panama (Hubbell and Foster, 1983). This plot was established in 1980 by researchers from the Smithsonian Tropical Research Institute and Princeton University and has been analysed in numerous studies (Law et al., 2009; Volkov et al., 2009; Waagepetersen et al., 2016; Wiegand et al., 2012; Yue and Loh, 2015). The first census was conducted in 1982 after which new censuses were carried out every five years since 1985. We use the data from the 8-th census.[1] For illustration purposes, we pick two species of trees, the locations of which are shown in Fig. 1. Results for a number of other species, ranging from quite rare to most common, can be found in Jongh (2020).

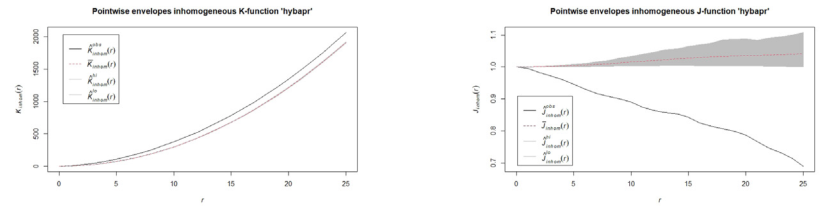
6. Clustering within species

It is evident from the plots in Fig. 1 that trees are not scattered in a homogeneous fashion over the study region. Therefore, in order to check whether trees of the selected species tend to cluster together in accordance to McGill's first axiom, we subject them to the inhomogeneous versions of the summary statistics discussed in Section 2. Since the intensity functions of the point processes are unknown, we estimate them using the non-parametric kernel estimator given by Eq. (8) with local edge correction (Lieshout, 2012). Here, we use a Gaussian kernel (Diggle, 1985), as it provides a kernel estimator with a high degree of smoothness. The quality of the obtained estimates relies heavily on the choice of the smoothing bandwidth. Here we use Cronie and Van Lieshout's criterion (Cronie and Lieshout, 2018) as given by Eq. (10).

For each of the species, we compare the estimates of the inhomogeneous K- and J-functions to the ones that correspond to an inhomogeneous Poisson process, which serves as the null hypothesis of absence of interactions. In order to assess if the deviations of the estimated statistics from those

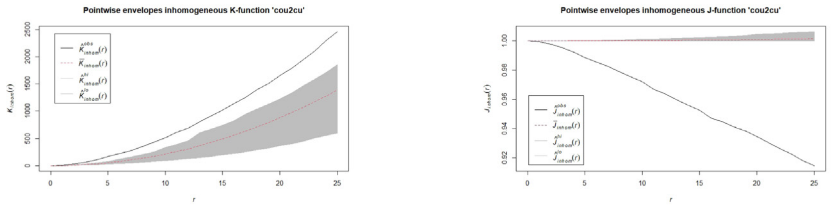
---

[1] https://doi.org/10.15146/5xcp-0d46.



(a) Point-wise envelopes of estimated inhomogeneous K-function for 'hybapr'.

(b) Point-wise envelopes of estimated inhomogeneous J-function for 'hybapr'.



(c) Point-wise envelopes of estimated inhomogeneous K-function for 'cou2cu'.

(d) Point-wise envelopes of estimated inhomogeneous J-function for 'cou2cu'.

Fig. 3. Simulation envelopes for estimated inhomogeneous K- and J-functions for 'hybapr' and 'cou2cu' based on 19 simulations.

of a Poisson process are significant, a range of Monte Carlo tests is conducted. We use the one-sided versions that test against clustered alternatives.

Specifying the intensity function of the null process proves to be difficult. Baddeley et al. suggest to use the observed point pattern to obtain an estimate $\hat{\lambda}(x)$ (Baddeley et al., 2016). However, that would violate the principle of not using the same data for estimation as for testing. Moreover, the simulated patterns are realizations of an inhomogeneous Poisson process with intensity function $\hat{\lambda}(x)$, whereas the observed pattern is an inhomogeneous Poisson process with unknown intensity function $\lambda(x)$, cf. the discussion at the end of Section 4. In order to circumvent this issue, we estimate the intensity function using data from the first census which has the least overlap with the observed pattern. To reduce the dependence of the estimates on the observed data even further, we remove all trees that were still alive at the time of the latest census. The intensity estimate based on the trimmed data is then scaled in such a way that the total number of points it predicts in the plot is equal to the actual number of trees counted during the latest census. Note that this method is not perfect in that it still uses the observed tree counts. To make sure that it is quite a reasonable option, we compare the intensity function thus obtained with a kernel estimate based on the data. Both functions are depicted in Fig. 2 for each of the selected species.

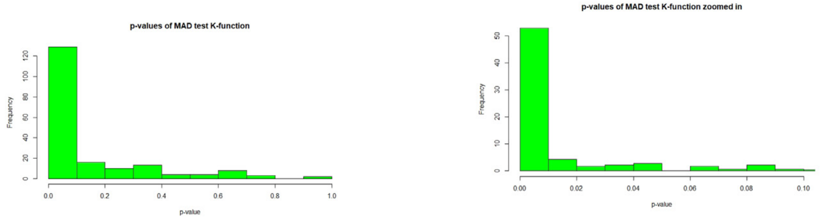
Fig. 2 suggests that the estimates are fairly good. The areas that are rich or barren in vegetation seem to be roughly in harmony. However, the figures do show some differences that may give rise to distortions in the test results. For example, the data of the latest census seem to feature a cluster of 'cou2cu' trees in the bottom right corner of the plot which is not captured by the estimated intensity functions.

Point-wise envelopes for the two species are shown in Fig. 3 based on 19 simulated patterns. We use the inhomogeneous K-function (1) and the inhomogeneous J-function, cf. (4). The functions are estimated using the estimators discussed in Section 2. Both summary statistics indicate that 'cou2cu' trees cluster together. After all, the estimated inhomogeneous K-function lies above the upper envelope, the estimated inhomogeneous J-function lies beneath the lower envelope over the entire range of r.

To assess whether the clustering is statistically significant, Table 1 lists the p-values returned by the MAD test, the DCLF test, the studentized MAD test and the directional quantile MAD test

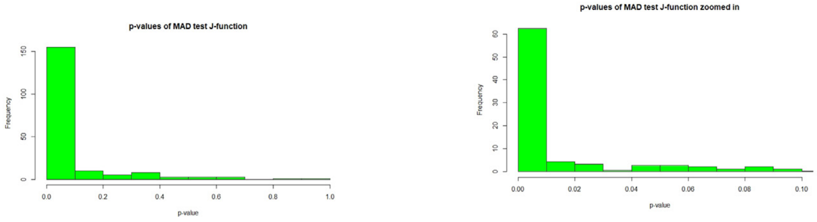
Table 1
P-values of the MAD test, the DCLF test, the studentized MAD test and the directional quantile MAD test for ‘cou2cu’.

Summary statistic	MAD test	DCLF test	Studentized MAD test	Directional quantile MAD test
K	0.01	0.01	0.02	0.01
J	0.01	0.01	0.01	0.01



(a) Histogram of p-values for the MAD test based on the inhomogeneous K-function.

(b) Histogram of zoomed in p-values for the MAD test based on the inhomogeneous K-function.



(c) Histogram of p-values for the MAD test based on the inhomogeneous J-function.

(d) Histogram of zoomed in p-values for the MAD test based on the inhomogeneous J-function.

Fig. 4. Distribution of p-values for the MAD test.

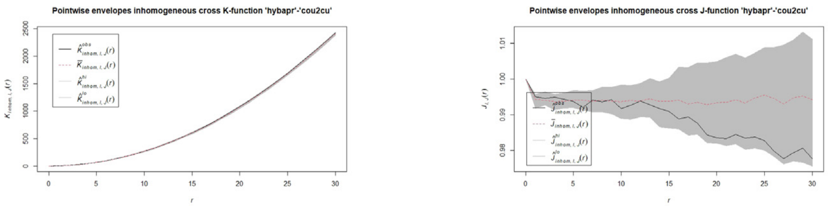
(Baddeley et al., 2016; Myllymäki et al., 2015), as described in Section 4, over the range $[0, 25]$ for ‘cou2cu’. The p-values for ‘hybapr’ are similar. In conducting these tests, we use the envelopes based on 99 simulations. Since one-sided versions of the studentized and the directional quantile MAD test are not available, we use the two-sided versions of these tests. The results of both the inhomogeneous K- and J-functions strongly support clustering.

The rank envelope (Myllymäki et al., 2016) test returns an interval of p-values. For ‘hybapr’ we find $(0, 0.004)$, for ‘cou2cu’ $(0, 0.002)$ when using the inhomogeneous K-function. For the inhomogeneous J-function the intervals are $(0, 0.002)$ and $(0, 0.0004)$ for ‘hybapr’and ‘cou2cu’ respectively. We use 500 simulations. These results are in line with those of the deviation tests.

To examine if the selected species are representative for the vegetation in the plot, we carry out the MAD test for each species featuring more than 50 trees. Due to the high number of species, here we use the kernel smoothed data to generate samples. Fig. 4 shows the distributions of the obtained p-values when using the inhomogeneous K- and J-functions. Both provide convincing evidence for McGill’s first axiom: the histograms show distinct peaks between 0 and 0.01. All tests were carried out in `R` using `spatstat` (Baddeley et al., 2016) and `GET` (Myllymäki et al., 2016).

7. Independence between species

The third of McGill’s axioms states that the patterns of trees of different species are independent of each other. In order to assess the validity of this axiom, we conduct a Lotwick–Silverman test



(a) Point-wise envelopes of estimated inhomogeneous cross K-function.

(b) Point-wise envelopes of estimated inhomogeneous cross J-function.

Fig. 5. Simulation envelopes for estimated inhomogeneous cross K- and J-functions from 'hybapr' to 'cou2cu' based on 99 simulations.

Table 2
P-values of the MAD test, the DCLF test, the studentized MAD test and the directional quantile MAD test for 'hybapr'–'cou2cu'.

Summary statistic	MAD test	DCLF test	Studentized MAD test	Directional quantile MAD test
K	0.28	0.30	0.60	0.53
J	0.06	0.06	0.20	0.08

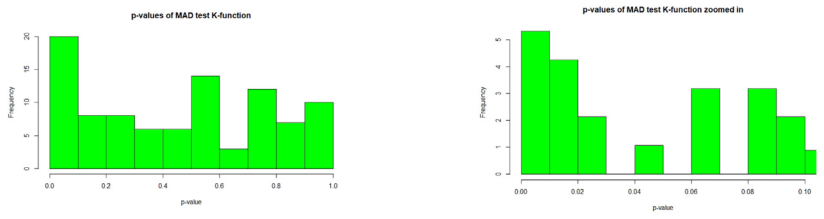
for independence (Lotwick and Silverman, 1982) applied to the inhomogeneous cross K- and J-functions introduced in Section 2. Recall that when species i and j are independent, $K_{\text{inhom; i,j}}(r) = \pi r^2$; larger values suggest positive association, smaller values negative association between the species. Similarly, $J_{\text{inhom; i,j}}(r) = 1$ when the species are independent. Smaller values suggest that trees of species j tend to cluster around those of species i, larger ones that trees of type j tend to avoid those of type i. The idea behind the test is that under the null hypothesis of independence, shifting the pattern and the intensity function of one of the species with respect to those of the other leaves the marginal structures intact and affects only the inter-species interactions. Since such a shift might move some points out of the plot, a torus correction is applied.

Fig. 5 shows the point-wise envelopes of the estimated inhomogeneous cross K- and J-functions for the species pair 'hybapr'-'cou2cu', each based on 99 torus translations of the first species with respect to the other. The graphs do not give us any reason to doubt the hypothesis of independence, as the estimated summary statistics for the data lie mostly within the envelopes.

Table 2 shows the p-values returned by the two-sided MAD test, the DCLF test, the studentized MAD test and the directional quantile MAD test applied to the inhomogeneous cross K- and J-functions for the pair 'hybapr'–'cou2cu' over the range [0, 30]. In conducting these tests, we again use envelopes based on 99 simulations. All p-values are 6 percent or higher and there is no reason to reject the independence hypothesis. Note though, that the p-values for the inhomogeneous cross K-function are higher than those for the inhomogeneous cross J-function so the latter statistic is more sensitive for our data.

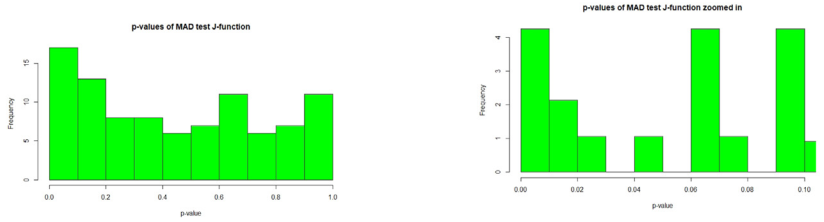
Also the intervals of p-values for the rank envelope test, $(0.637, 0.639)$ for the inhomogeneous cross K-function and $(0.0988, 0.107)$ for the inhomogeneous cross J-function, are in confirmation (2500 toroidal shifts).

In order to gain more insight in the validity of the third axiom, we examine the total pool of species pairs. We restrict ourselves to patterns containing more than 50 trees and pair them randomly. We then estimate the inhomogeneous cross K- and J-functions and apply the MAD test as before. The distributions of the obtained p-values are shown in Fig. 6. Both for the inhomogeneous cross K- and J-function, over 85 percent of the p-values exceed 0.05. Hence, the third axiom holds for the majority of pairs. All tests were carried out in `R` using `spatstat` (Baddeley et al., 2016) and `GET` (Myllymäki et al., 2016).



(a) Histogram of p-values for the MAD test based on the inhomogeneous cross K-function.

(b) Histogram of zoomed in p-values for the MAD test based on the inhomogeneous cross K-function.



(c) Histogram of p-values for the MAD test based on the inhomogeneous cross J-function.

(d) Histogram of zoomed in p-values for the MAD test based on the inhomogeneous cross J-function.

Fig. 6. Distribution of p-values for the MAD test.

8. Conclusions

In this paper, we discussed recent methodological developments in global envelope testing and inhomogeneous summary statistics, which were then used to assess the validity of the first and third of McGill's axioms of biodiversity using data on the spatial distribution of trees found in a 50 hectare plot on Barro Colorado Island (Hubbell and Foster, 1983). Previous studies on McGill's axioms all rely on the strong assumption of homogeneity, which is unlikely to hold for all but highly featureless terrains. Methods that fail to take into account inhomogeneity cannot disentangle species interaction and habitat association and therefore may lead to false conclusions. For example, preference for different regions may be taken for negative correlation, where in fact the species are independent. In this paper, we presented compelling evidence for the correctness of McGill's first and third axioms. To the best of our knowledge, this is the first work that assessed the validity of McGill's axioms without relying on the restrictive assumption of homogeneity.

Although the results of this study are highly indicative of the correctness of McGill's axioms, we should be heedful not to draw too presumptuous conclusions. More and different kinds of ecosystems should be investigated. Furthermore, recall that the results of the Monte Carlo tests depend on various choices, such as the type of edge correction used, the estimated intensity function, the choice of bandwidth and the range of interaction considered. The robustness of the results with respect to these choices should also be investigated. Nevertheless, this paper provides a strong basis for further research on biodiversity and demonstrates the potential of recent developments in the analysis of inhomogeneous spatial point patterns.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks are due to Professor Richard J. Boucherie for helpful discussions.

The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB0640386, DEB0425651, DEB0346488, DEB0129874, DEB00753102, DEB9909347, DEB9615226, DEB9615226, DEB9405933, DEB9221033, DEB9100058, DEB8906869, DEB8605042, DEB8206992, DEB7922197; support from the Center for Tropical Forest Science; the Smithsonian Tropical Research Institute; the John D.and Catherine T. MacArthur Foundation; the Mellon Foundation; the Small World Institute Fund; and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades.

Calculations were carried out using `spatstat` (Baddeley et al., 2016) and `GET` (Myllymäki et al., 2016). The research was mostly conducted during the internship of the first author at CWI.

References

Anderes, E.A., Møller, J., Rasmussen, J.G., 2020. Isotropic covariance functions on graphs and their edges. Ann. Statist. 48, 478–503.

Baddeley, A., Diggle, P.J., Hardegen, A., Lawrence, T., Milne, R.K., Nair, G., 2014. On tests of spatial pattern based on simulation envelopes. Ecol. Monograph 84, 477–489.

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

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., Davies, T.M., 2021. Analysing point patterns on networks - A review. Spatial Stat. 42, 1–35.

Baddeley, A., Rubak, E., Turner, R., 2016. Spatial Point Patterns. Methodology and Applications with R. CRC Press.

Belhadji, A., Bardenet, R., Chainais, P., 2019. Kernel quadrature with DPPs. In: NeurIPS 2019 - Thirty-Third Conference on Neural Information Processing Systems. Vancouver.

Besag, J., Clifford, P., 1989. Generalized Monte Carlo significance tests. Biometrika 76, 633–642.

Besag, J., Diggle, P.J., 1977. Simple Monte Carlo tests for spatial pattern. J. R. Stat. Soc. Ser. C 26, 327–333.

Chesson, P., 2000. General theory of competitive coexistence in spatially-varying environments. Theor. Popul. Biol. 58, 211–237.

Clements, F.E., 1916. Plant Succession: An Analysis of the Development of Vegetation. Carnegie Institute of Washington.

Cressie, N.A.C., 1991. Statistics for Spatial Data. Wiley.

Cronie, O., Lieshout, M.N.M. van, 2015. Summary statistics for inhomogeneous marked point processes. Ann. Inst. Statist. Math. 68, 905–928.

Cronie, O., Lieshout, M.N.M. van, 2018. A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. Biometrika 105, 455–462.

Daley, D., Vere-Jones, D., 2008. An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure. Springer.

Dao, N.A., Genton, M.G., 2014. A Monte Carlo adjusted goodness-of-fit test for parametric models describing spatial point patterns. J. Comput. Graph. Statist. 23, 497–517.

Diggle, P.J., 1978. Statistical methods for spatial point patterns in ecology. In: Proceedings International Statistical Ecology Program.

Diggle, P.J., 1979. On parameter estimation and goodness-of-fit testing for spatial point patterns. Biometrics 35, 87–101.

Diggle, P.J., 1983. Statistical Analysis of Spatial Point Patterns. Academic Press.

Diggle, P., 1985. A kernel method for smoothing point process data. J. R. Stat. Soc. Ser. C 34, 138–147.

Du Rietz, G.E., 1929. The fundamental units of vegetation. In: Proceedings of the International Congress of Plant Science, Vol. 1, pp. 623–627.

Fisher, R.A., Thornton, H.G., Mackenzie, W.A., 1922. The accuracy of the plating method of estimating the density of bacterial populations with particular reference to the use of Thornton's agar medium with soil samples. Ann. Appl. Biol. 9, 325–359.

Gause, G.F., 1934. Struggle for Existence. Williams & Wilkins.

Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P., 2010. Handbook of Spatial Statistics. CRC Press.

Gelfand, A.E., Fuentes, M., Hoeting, J.A., Smith, R.L., 2019. Handbook of Environmental and Ecological Statistics. CRC Press.

Gentry, A.H., 1988. Tree species richness of upper Amazonian forests. Proc. Natl. Acad. Sci. USA 85, 156–159.

Gleason, H.A., 1926. The individualistic concept of the plant association. Bull. Torrey Bot. Club 53, 7–26.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press.

Hubbell, S.P., Foster, R.B., 1983. Diversity of Canopy Trees in a Neotropical Forest and Implications for Conservation. Blackwell, pp. 25–41.

Huisman, J., Weissing, F.J., 1999. Biodiversity of plankton by species oscillations and chaos. Nature 402, 407–410.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. Wiley.

Jongh, M.C. de, 2020. Biodiversity at Barro Colorado Island (MSc Thesis). University of Twente.

Kenkel, N.C., 1988. Pattern of self-thinning in jack pine: testing the random mortality hypothesis. Ecology 69, 1017–1024.

Lavancier, F., Møller, J., Rubak, E., 2015. Determinantal point process models and statistical inference. J. R. Stat. Soc. Ser. B 77, 853–877.
Law, R.J., Illian, J., Burslem, D.F.R.P., Gratzer, G., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., 2009. Ecological information from spatial patterns of plants: insights from point process theory. J. Ecol. 97, 616–628.
Lieberman, M., Lieberman, D., 2007. Nearest-neighbor tree species combinations in tropical forests: the role of chance, and some consequences of high diversity. Oikos 116, 377–386.
Lieshout, M.N.M. van, 2011. A J-function for inhomogeneous point processes. Stat. Neerlandica 65, 183–201.
Lieshout, M.N.M. van, 2012. On estimation of the intensity function of a point process. Methodol. Comput. Appl. Probab. 14, 567–578.
Lieshout, M.N.M. van, 2018a. Nearest-neighbour Markov point processes on graphs with Euclidean edges. Adv. Appl. Probab. 50, 1275–1293.
Lieshout, M.N.M. van, 2018b. Nonparametric indices of dependence between components for inhomogeneous multivariate random measures and marked sets. Scand. J. Stat. 45, 985–1015.
Lieshout, M.N.M. van, 2019. Theory of Spatial Statistics. A Concise Introduction. CRC Press.
Lieshout, M.N.M. van, Baddeley, A.J., 1996. A nonparametric measure of spatial interaction in point patterns. Stat. Neerlandica 50, 344–361.
Loader, C., 1999. Local Regression and Likelihood, first ed. Springer.
Loosmore, N.B., Ford, E.D., 2006. Statistical inference using the G or K point pattern statistics. Ecology 87, 1925–1931.
López-Pintado, S., Romo, J., 2009. On the concept of depth for functional data. J. Amer. Statist. Assoc. 104, 718–734.
Losos, E.C., Leigh, E.G. (Eds.), 2004. Tropical Forest Diversity and Dynamism: Findings from a Large-Scale Plot Network. University of Chicago Press.
Lotwick, H.W., Silverman, B.W., 1982. Methods for analysing spatial processes of several types of points. J. R. Stat. Soc. Ser. B Stat. Methodol. 44, 406–413.
Macchi, O., 2017. Point Processes and Coincidences. Contributions to the Theory. with Applications to Statistical Optics and Optical Communication. Walter Warmuth Verlag.
McGill, B.J., 2010. Towards a unification of unified theories of biodiversity. Ecol. Lett. 13, 627–642.
Møller, J., Rubak, E., 2016. Functional summary statistics on the sphere with an application to determinantal point processes. Spatial Stat. 18, 4–23.
Møller, J., Waagepetersen, R.P., 2004. Statistical Inference and Simulation for Spatial Point Processes. CRC Press.
Myllymäki, M., Grabarnik, P., Seijo, H., Stoyan, D., 2015. Deviation test construction and power comparison for marked spatial point patterns. Spatial Stat. 11, 19–34.
Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., Hahn, U., 2016. Global envelope tests for spatial processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 79, 381–404.
Perry, G.L.W., Enright, N.J., Miller, B.P., Lamont, B.P., 2009. Nearest-neighbour interactions in species-rich shrub lands: the roles of abundance, spatial patterns and resources. Oikos 118, 161–174.
Ramsay, J., Silverman, B., 2006. Functional Data Analysis, second ed. Springer.
Ricklefs, R.E., 1990. Ecology. W.H. Freeman and Company, third ed..
Ripley, B.D., 1976. The second-order analysis of stationary point processes. J. Appl. Probab. 13, 255–266.
Ripley, B.D., 1977. Modelling spatial patterns (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 39, 172–212.
Ripley, B.D., 1981. Spatial Statistics. Wiley.
Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization, third ed. Wiley.
Szmyt, J., 2014. Spatial statistics in ecological analysis: from indices to functions. Silva Fennica 48, 1–31.
Tilman, D., 1982. Resource Competition and Community Structure. Princeton University Press.
Volkov, I., Banavar, J.R., Hubbell, S.P., Maritan, A., 2009. Inferring species interactions in tropical forests. Proc. Natl. Acad. Sci. USA 106, 13854–13859.
Waagepetersen, R., Guan, Y., Jalilian, A., Mateu, J., 2016. Analysis of multispecies point patterns by using multivariate log-Gaussian Cox processes. J. R. Stat. Soc. Ser. C 65, 77–96.
Wiegand, T., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., Huth, A., 2007. How individual species structure diversity in tropical forests. Proc. Natl. Acad. Sci. USA 104, 19029–19033.
Wiegand, T., Huth, A., Getzin, S., Wang, X., Hao, Z., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., 2012. Testing the independent species' arrangement assertion made by theories of stochastic geometry of biodiversity. Proc. R. Soc. B 279, 3312–3320.
Wiegand, T., Moloney, K.A., 2014. Handbook of Spatial Point-Pattern Analysis in Ecology. CRC Press.
Wright, S.J., 2002. Plant diversity in tropical forests: a review of mechanisms of species coexistence. Oecologia 130, 1–14.
Yue, Y., Loh, J.M., 2015. Variable selection for inhomogeneous spatial point process models. Canad. J. Statist. 43, 288–305.