Computational Statistics and Data Analysis 55 (2011) 627–640

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Estimating inter-group interaction radius for point processes with nested spatial structures

J. Chadœuf [a,*], G. Certain [b], E. Bellier [a], A. Bar-Hen [c,d], P. Couteron [e], P. Monestiez [a], V. Bretagnolle [b]

[a] INRA, Biométrie, Domaine St Paul, Site Agroparc, 84914 Avignon Cedex 9, France
[b] CEBC—CNRS UPR 1934, Villiers-en-Bois, F-79360 Beauvoir Sur Niort, France
[c] Université Paris 5, 45 rue des Saints-Pères, 75270 Paris cedex 06, France
[d] EHESP, Avenue du professeur Léon Bernard, 35043 Rennes cedex, France
[e] IFP, 11 St Louis Street, Pondicherry, 605001, India

ARTICLE INFO

Article history:
Received 15 April 2008
Received in revised form 2 June 2010
Accepted 2 June 2010
Available online 12 June 2010

Keywords:
Interaction radius
Point process
Local statistics

ABSTRACT

A statistical procedure is proposed in order to estimate the interaction radius between points of a non-stationary point process when the process can present local aggregated and regular patterns. The model under consideration is a hierarchical process with two levels, points and clusters of points. Points will represent individuals, clusters will represent groups of individuals. Points or clusters do not interact as soon as they are located beyond a given interaction radius, and are assumed to interact if their distance is less than this interaction radius. Interaction radius estimation is performed in the following way. For a given distance, observations are split into several clusters whose in-between distances are larger than this distance. For each cluster, a neighbourhood and an area in which this cluster is randomly located is defined under the assumption that the distance between the cluster and its neighbourhood is larger than the interaction radius. The *p*-value of a test of this assumption is then computed for each cluster. Modelling the expectation of this *p*-value as a function of the distance leads to an estimate of the interaction radius by a least-square method. This approach is shown to be robust against non-stationarity. Unlike most classical approaches, this method makes no assumption on the point spatial distribution inside the clusters. Two applications are presented in animal and plant ecology.

1. Introduction

The interaction radius, the maximum distance at which two individuals interact, is one crucial parameter of inter-individual interaction processes. When this parameter cannot be estimated experimentally, it is possible to estimate the interaction radius from field data as a parameter of a statistical spatial model that describes the structure of the observed pattern of individuals.

The usual statistical models of point patterns have mainly focused on homogeneous patterns, that is patterns for which the interaction is regular everywhere or aggregated everywhere. The main processes that have been studied include (i) Cox processes, that is Poisson processes with a stationary random intensity in which aggregation comes from variations in the intensity, (ii) Neyman–Scott processes, which are obtained by attaching independently a random set of daughter points to each point of a set of Poisson parents, aggregation being due to the clusters of daughters around each parent, and (iii) Strauss point processes where the presence of a point at a given location depends on the location of points within a given radius (see Stoyan et al., 1995 for such models; Renshaw and Särkkä, 2001 for a Gibbs models in a spatio-temporal case).

* Corresponding author.
E-mail address: joel@avignon.inra.fr (J. Chadœuf).

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.06.004

Thus, statistical estimations of individual patterns have been performed on scales at which the pattern structure is always of the same kind. It can be regular (as for example in Särkkä and Tomppo, 1998) or aggregated (Sterner et al., 1986). Long range variations due to spatially varying factors can be included in these analyzes; see for example Shimatani and Kubota (2004) in the case of an aggregated process, or Stoyan and Stoyan (1999) in the case of a regular process. Estimation methods will depend on the model, mainly maximum of likelihood or pseudo-likelihood (see for example Guyon, 1993 or Möller and Waagepetersen, 2004) and bayesian estimation (Diggle et al., 1998).

However, inter-individual interactions may include several processes that nonetheless all have the same parameter, i.e. the same interaction radius. For example the spatial structure of a fish school may vary depending on the kind of activity in which the school is engaged while the interaction radius between fish remains the same (Couzin et al., 2002). When several inter-individual interaction processes are at work, all depending on the same interaction distances, the resulting spatial patterns may be a mix of clusters presenting locally several kinds of spatial structure. The point pattern of individuals is then not homogeneous and the use of the classical models is no more possible.

These interactions may also depend locally on the value of environmental variables that vary at large scale, e.g. prey distribution, landscape, habitats (Fortin and Dale, 2005). At very large scales, plant species distribution will be governed by global climatic variables, while at intermediate scales external influences by soil or topographic variables often determine variations of plant density (Pélissier and Goreaud, 2001). At a local scale, plants can be aggregated because of limited dispersal of propagules or facilitative influences exerted by other individuals (Callaway, 1995) but plants can also be regularly distributed because of density-dependent factors such as competition or phytophage abundance (Holmgren et al., 1997). In animal ecology, similar variations appear: Veit (1999) for example notes that krill swarms "tended to be located consistently in the same places over the shallowest part of the transects". Then, the patterns of individuals are not homogeneous and methods proposed to deal with it (Shimatani and Kubota, 2004; Stoyan and Stoyan, 1999) cannot be used if the environmental variables are not known.

Estimating the interaction radius as a parameter of a classical point process model cannot be used in the cases we are interested in: the pattern structure may vary locally from aggregation to regularity, and the rules governing these changes are not known. One solution could be to identify and separate the different groups, i.e., separate them in several clusters inside which the structure is constant, and then to analyze separately (i) the process of cluster distribution and (ii) the process of point distribution within each cluster. The main point is then to partition the individuals into clusters. Methods to do this are based on the variations of local point intensity. The most popular of these are the scan statistic developed by Kulldorff (1997) and Kulldorff and Nagarwalla (1995) and its developments as in Duczmal and Assunção (2004) and Duczmal et al. (2007), and the method proposed by Kelsall and Diggle (1995). These methods basically test whether the mean number of points per unit area in a given area is significantly larger than outside this area, using the Poisson assumption, i.e. independence between points, an assumption we cannot make since points interact. Demattei et al. (2006) used a different method based on nearest neighbour distance, but cluster detection relies again on a random distribution of points outside the cluster. Detected clusters may then be very different from the real ones, and could lead to a biased estimation of the interaction radius.

We propose then to estimate directly the interaction radius as the minimum distance beyond which individuals do not interact. In fact, to act as a group, a set of individuals such as birds for example must first detect and monitor other members of a group. Then interaction may occur depending on local conditions, such as aggregation if prey are found as mentioned above. Two individuals at a distance larger than this interaction radius can interact indirectly if they are connected by a chain of individuals such that the distance between two consecutive individuals of the chain is less than the interaction radius, as the behaviour of the first animal can be identified and modify the behaviour of the second animal and so on. From a modelling point of view, the spatial repartition of these individuals can then be seen as a hierarchical model where first, a set of individuals constitutes a group as soon as any two individuals are connected by a chain of individuals, and second the spatial structure of each group may depend on local conditions.

In such a model, clusters of points are not so much defined by a higher intensity with respect to a neighbourhood, as by the fact that points belonging to two different clusters do not interact because they are separated by a larger distance than the interaction radius, i.e. the distance over which two clusters do not interact. In statistics, a caricatural case is the hard-core point process with hard-core distance R_0. In this model, the interaction radius is R_0 and each point is a cluster: the point does not interact with other points located farther than R_0. While this process is stationary, its regularity increases with the process density and the pattern of points can become very regular. Then, the scan statistic and the method of Kelsall & Diggle will fail to detect clusters because they will detect no area with a large local density of points, and then cannot be used to estimate the interaction radius.

We propose in this paper an estimation method with very few assumptions and based on the following property: a given cluster of points is randomly distributed in a given area if its distance to its neighbourhood is larger than the interaction radius. We test this assumption using this property on each cluster and compute their p-values. We estimate the interaction radius with a least-square method by modelling the expectation of the mean of p-values of these tests. Finally, as the local intensity of the point process cannot be assumed constant in many applications we show that our estimator minimizes the effect of a large scale non-stationarity, using the fact that the statistics on which the estimation is based are local since each of them depends only on the position of a cluster conditionally to its neighbours.

In Section 2, we present our model with the proposed statistics and we give some asymptotical results. The method is then tested with simulated examples in Section 3. Its application to the analysis of tree spatial interactions in a semi-arid context is provided in Section 4. In Section 5, we look at the spatial repartition of a seabird in a temperate continental shelf.

2. Model definition, test and estimation

The aim being to test whether the interaction radius R_0 is larger than a specified value R without specifying a parametric model, we will

(i) Formalize the model we propose in a first step. We focus more specifically on the fact that if the distance of a group of points to other points is larger than the interaction radius, then this group is uniformly randomly located in a given area (property 2 in the following). Additional statistical assumptions are then given.
(ii) For a given distance R, we split the observed set of points X into groups such that points in a group are connected by a set of points with distance less than R. Then we define polygons enclosing these groups, and areas into which each group is randomly uniformly distributed if R is larger than the interaction radius. We compute for each group the p-value of the test of this assumption. At the end of this step we have, for a given R, a series of p-values which are randomly uniformly distributed between 0 and 1 if R is larger than the interaction radius.
(iii) We combine these local tests in a global test in the next step.
(iv) We estimate the interaction by modelling the changes of the p-value of the global test with respect to R in step four.
(v) We look at the effect of non-stationarity in the last step.

2.1. Interaction radius definition and model assumptions

We consider in the following a point process X in $\mathbf{R}^2$. For B a compact set of $\mathbf{R}^2$, the number of points of X inside B is a random variable $X(B)$. We note $\lambda(.)$ its intensity. It verifies $E(X(B)) = \int_B \lambda(x)\mathrm{d}x$.

Definition. We define the interaction radius of X as the minimum distance R_0 such that the following Markovian property exists: let B a compact set, $B \oplus B(0, R_0) = \{x + y; x \in B, y \in B(0, R_0)\}$ the dilation of B by $B(0, R_0)$, the disc centered on 0 of radius R_0, X_u a finite subset of X, then

$$P(X_u = x_u \mid X_u \in B, X \setminus \{X_u\}) = P(X_u \mid X_u \in B, (X \setminus \{X_u\}) \cap B \oplus B(0, R_0)). \tag{1}$$

That is, the position of a set of points inside a subset conditionally to the position of the other points of the process depends only on the positions of the points of the process inside a neighbourhood of the subset. This Markovian property is verified for pair interaction with finite range processes, as for example the Strauss process.

Consequences: (a) if B_1 and B_2 are two compact sets of $\mathbf{R}^2$, for which a closed curve without intersection C exists such that B_1 belongs to the interior of $C_{R_0} = C \oplus B(0, R_0)$ and B_2 belongs to its exterior, then Eq. (2) holds:

$$P(X(B_1)X(B_2) \mid X(C_{R_0}) = 0) = P(X(B_1) \mid X(C_{R_0}) = 0)P(X(B_2) \mid X(C_{R_0}) = 0). \tag{2}$$

This property is classically used by coding methods to build independent sub-samples from one realization (Guyon, 1993).
(b) If $X(B_1) = X \cap C_{R_0}$, i.e. if the points inside B_1 are the only points of X inside C, then:

$$P(X(B_1) \mid X(C_{R_0}) = 0) = P(X(t(B_1)) \mid X(C_{R_0}) = 0). \tag{3}$$

if t is a translation such that $t(B_1)$, the translation of B_1 by t, belongs to the interior of C.

Statistical assumption: We assume classically that the process is α-mixing. More precisely, we assume that if V_x is a bounded random vector measured at a point $x \in \mathbf{R}^2$, depending on $B(x, D) \cap X$, the points of X inside a disc of radius D centered on x, then for W a compact convex set:

- there exits a vector V such that $\frac{1}{\|W\|}\sum_{x\in X\cap W} V_x \to V$ in probability when $W \to \mathbf{R}^2$
- there exists a matrix Σ_V such that $\sqrt{\|W\|}\left(\frac{1}{\|W\|}\sum_{x\in X\cap W} V_x - V\right) \to \mathcal{N}(0, \Sigma_V)$ in law when $W \to \mathbf{R}^2$.

This property, which means that no long-distance interaction exists, is necessary to ensure that estimators based on coding methods are statistically convergent and asymptotically Gaussian once normalized.

2.2. Testing locally if $R > R_0$

Defining groups

Let $R > 0$ be a proposed interaction radius. R defines a partition $C_R(X)$ of X, the set of R-connected groups defined as follows. If A and B are two points of X, A and B belong to the same R-connected group (Ling, 1973) if there exists a finite set of points $x_1, \ldots, x_n$ belonging to X such that:

- $\|A - x_1\| \le R$,
- $\|x_i - x_{i+1}\| \le R$ for $i < n$,
- $\|x_n - B\| \le R$.

For the set of points in Fig. 1(a), $R = 0.05$. Fig. 1(c) presents the groups of R-connected points.

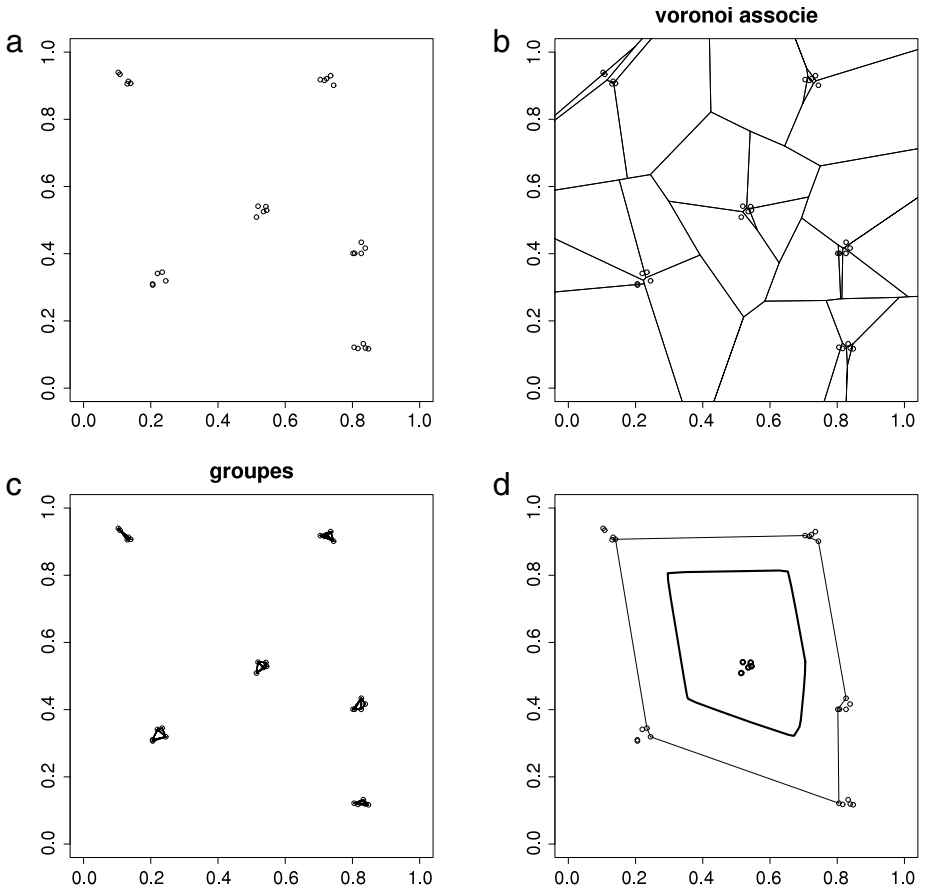

Fig. 1. Definition of a group and its neighbourhood. (a) Original set of points. (b) Voronoï tessellation based on the points in (a). (c) Points of (a) whose distances are less than R are connected together and define groups. (d) For a given group, here in the center of the picture, points of (a) not belonging to the group and neighbours of the point of the group define a polygon around the group (in thin lines), the bound eroded set of this polygon by the union of the discs of radius R centered on the points of the group at the center of the picture appears in thick lines.

Defining neighbours

Let $V(X)$ be the Voronoï tessellation (Stoyan et al., 1995) associated with the point process X and denote $v(x)$ the cell with centroid x for $x \in X$. This tessellation is defined by associating to each $x \in X$ the set of points $\{z \in \mathbf{R};\ \|z - x\| \leq \min_{\{y \in X\}} \|z - y\|\}$, that is all points of $\mathbf{R}$ which are nearer to x than to other points of X.

To each point x is associated a set of neighbours, two points being neighbours if their Voronoï cells have a common edge. Let $N(x)$ denote the neighbours of x. Fig. 1(b) shows the Voronoï tessellation obtained with points of Fig. 1(a).

Let s be an R-connected group. The neighbours of s are defined as $N(s) = \cup_{y \in s} N(y) \setminus \{s\}$. $N(s)$ is composed of the points of X having a common edge with the set $v(s) = \cup_{y \in s} v(y)$. $v(s)$ being a finite convex set a.s., the set $N(s)$ can be ordered, for example with respect to the trigonometric orientation, and defines a polygon $P(N(s))$. Points connected by a thin line in Fig. 1(d) correspond to neighbours of the group in the center of Fig. 1(c).

Testing on one group if $R_0 > R$

Suppose that $\lambda(x) = \lambda$, i.e. that the point process X is stationary. If $R > R_0$, then s is randomly uniformly distributed in $P(N(s)) \ominus B(0, R) = P_R(s)$, the eroded set of $P(N(s))$ by the disc centered on 0 and of radius R, conditionally to the fact that s is included in $P_R(s)$. The eroded set associated to the group of points in Fig. 1(c) is the set limited by thick lines in Fig. 1(d).

Let $d(s) = \min\{\|x - y\|; x \in s, y \in N(s)\} = P_{R_0}^{-1}(U)$ the minimum distance between the points of s and the points of its neighbourhood, where U is the p-value of $d(s)$ under $H(R_0)$, which represents the uniform repartition inside $P_{R_0}(s)$. If $R > R_0$,then $U_R(s) = P_R(d(s))$, the p-value of $d(s)$ under the uniform spatial repartition of s in $P_R(s)$, is uniformly distributed in $[0 - 1]$. It presents departures from this uniform distribution if $R < R_0$, the departure depending on the local properties of the model. If the model is aggregative around s, one expects to find more frequently s near the nodes of the polygon $P(N(s))$, and the distribution of $\widehat{U_R}(s)$ should peak around 0, whereas it should present a peak around 1 if the process is locally regular around s.

2.3. Testing if the interaction radius is larger than a given value R

To build a global test, we proceeded as in coding methods (Besag, 1974; Guyon, 1993). Let S be the set of all R-connected sets of X; define a coding $\mathcal{A}_R$ as a subset of S such that if s and s' belong to $\mathcal{A}_R$, then $s' \notin N(s)$. Then, conditionally to $S \setminus \mathcal{A}_R$, any two surfaces $P_R(s)$ and $P_R(s')$ are distant from R or more by construction for s and s' in $\mathcal{A}$, such that, if $R > R_0$, the spatial repartitions of s and s' in these areas are independent. The p-values $(U_R(s))_{s\in\mathcal{A}_R}$ are then an i.i.d. sample of the uniform distribution and any statistics $T(R) = T(U_R(s), s \in \mathcal{A}_R)$ classically used to test adequacy to the uniform distribution can be used.

2.4. Estimating the interaction radius

$\widehat{R_0}$ definition

Let R be a given distance, $n(R)$ the number of R-connected sets s defined for this distance. For $R > R_0$, $U_R(s)$ is uniform between 0 and 1 and $E(U_R(s)) = m(R_0, R) = 1/2$. For $R < R_0$, $U_R(s) = Q_R(P_R^{-1}(V_s))$ where:

- V_s is a uniform random variable on $[0\ -\ 1]$,
- Q is the cumulative distribution function, under the uniform distribution of the group s, of the statistics used to measure the distance of the group from the neighbouring group.
- P_R is the cumulative distribution function, under the actual distribution of the group s, of the statistics used to measure the distance of the group to the neighbouring groups.

Its expectation, $E(U_R(s)) = E\left(Q_R(P_R^{-1}(V))\right)$ with respect to Q_R depends on the configurations in the neighbourhood of width R_0 of $P(s)$ and the ergodicity assumption leads to $\frac{1}{n(R)}\sum_s E(U_R(s)) \to m(R_0, R)$.

Let (R_j) J ordered distances, $R_1 < R_0$, $R_J > R_0$, and denote $S(R)$ the function

$$S(R) = \sum_{j\le J}\left(\frac{1}{n(R_j)}\sum_s U_{R_j}(s) - m(R, R_j)\right)^2. \tag{4}$$

We propose to estimate R_0 as a mean squared error estimator, by $\hat{R_0} = \text{argmin}\{S(r)\}$, the value for which $S(R)$ is minimum.

$\widehat{R_0}$ asymptotic property

When the sampling window W tends to infinity, $\sqrt{\|W\|}(\hat{R_0} - R_0)$ is asymptotically Gaussian with mean 0 and variance

$$\frac{\text{grad}(m(R, .))'_{|R=R_0}\,\Gamma\,\text{grad}(m(R, .))_{|R=R_0}}{\left(\text{grad}(m(R, .))'_{|R=R_0}\,\text{grad}(m(R, .))_{|R=R_0}\right)^2}.$$

This result is obtained by first noting that

- the vector $\sqrt{\|W\|}\left(\frac{1}{n(R_j)}\sum_s U_R(s) - m(R, R_j)\right)$ is asymptotically unbiased and Gaussian if W denotes the sampling window,
- $S(R)$ tends to the function $\sum_{j\le J}\left(m(R_O, R_j) - m(R, R_j)\right)^2$ which is minimum at $R = R_0$,
- if Γ denotes its asymptotic variance matrix, $\lim_{|W|\to\infty}\text{var}(\text{grad}(S(R))) = 4\,\text{grad}(m(R, .))'_{|R=R_0}\,\Gamma\,\text{grad}(m(R, .))_{|R=R_0}$,
- $\lim_{|W|\to\infty}\frac{d^2S(R)}{dR^2} = -2\,\text{grad}(m(R, .))'_{|R=R_0}\,\text{grad}(m(R, .))_{|R=R_0}$.

$S(R)$ is thus a contrast (Dacunha-Castelle and Duflo, 1993), such that:

(1) $\hat{R_0} = \text{argmin}S(r)$ is asymptotically consistent
(2) $\sqrt{|W|}(\hat{R_0} - R_0)$ is asymptotically Gaussian with mean 0 and variance

$$\frac{\text{grad}(m(R, .))'_{|R=R_0}\,\Gamma\,\text{grad}(m(R, .))_{|R=R_0}}{\left(\text{grad}(m(R, .))'_{|R=R_0}\,\text{grad}(m(R, .))_{|R=R_0}\right)^2}.$$

In practice the functions $m(R_0, R)$ are unknown for $R < R_0$. Suppose that for a given group s P_R is K times left-side derivable at R_0, then $Q_R o P_R^{-1}$ is K times left-side derivable at R_0. We propose to approach $m(R_0, R)$ by using polynomial functions:

$$m(R_0, R) = 1/2 + 1_{\{R<R_0\}}\sum_{1<k<K} a_k(R - R_0)^k. \tag{5}$$

The values a_k correspond to the ergodic limits of the mean of the k terms of the development of $Q_R o P_R^{-1}$. The asymptotic limits cannot be explicited, all the more when $R < R_0$. Confidence intervals of the curve $R \to S(R)$ and of $\hat{R_0}$ can be approached by block bootstrap (Lahiri, 1999).

Remark. Estimating R_0 using the p-value of the Kolmogorov–Smirnov statistics on the U_s is not so straightforward. The main reason is that for $R > R_0$ it is uniformly distributed consequently the preceding method will not work. One will

then rely on other procedures, as for example estimating R_0 as the first value under which the test remains constantly outside its confidence interval under uniform assumption. However, this means dealing with multiple dependent tests. One classical procedure is the traditional Bonferroni procedure (Manly, 1991), but whose quality depends heavily on the number of comparisons. On the other hand, Benjamini and Yekutieli (2001) proposed to change criteria, by controlling the false discovery rate instead of the type I error.

Choosing the test statistic T. As mentioned before, any test of uniformity of the position of a group can theoretically be used, with any statistic T. The choice of T can also be directed by the structure of the point process if prior information exists. (i) Suppose that the point pattern structure in each group is constant, either aggregative or regular. A natural choice is then to look at an extremum, either the min or the max of the empirical distribution, as this statistic will be very sensitive to departures from uniform distribution in this direction. (ii) If the two kinds of structures can be present, a more useful approach will be to use the maximum distance between the empirical distribution and the first diagonal, then use a Kolmogorov–Smirnov test. (iii) Another approach can be to use the mean value, which will be sensitive to the ratio of aggregative structures against regular structures. This last approach is used in the following.

2.5. Dealing with non-stationarity

When scanning large windows, individual and group spatial repartitions cannot generally be assumed stationary. Spatial trends can be due for example to environmental long range smooth variations. In the stationary case, the probability density of the typical point X_0 of a group $B \oplus X$ inside the polygon S_R, $p(X_0 \mid X_0 \in S_R) = p(B \oplus X_0, W_B)$, depends on the local interactions between $B \oplus X_0$ and the points W_B in the neighbourhood of $B \oplus X_0$. It then depends only on the relative position of $B \oplus X$ with respect to W_B. In the non-stationary case, suppose that these interactions, depending on local behaviour between individuals, do not change, but that the probability of finding a group at a given location depends on the location. In such a case, the probability of finding the typical point at X_0 conditionally to the neighbourhood W_B may also depend on the location X_0:

$$p(X_0 \mid X_0 \in S_R) = p(a(X_0); B \oplus X_0, W_B). \tag{6}$$

Suppose that $|a(x)| \to 0$, then $p(a(X_0); B \oplus X_0, W_B) = p(B \oplus X_0, W_B) + a(X_0)\alpha(B \oplus X_0, W_B) + o(a)$ and the p-value of the statistic T is $Q_a(t) = Q(t) + \int_{S_t} a(x)\alpha(B \oplus \mathrm{d}x, W_B)\mathrm{d}x + o(a)$ if S_t denotes the subset of S_R on which $T(x) \geq t$.

Therefore the estimated p-value of $Q(U) = T(X_0)$ verifies

$$P(T(X_0)) = P(Q_a^{-1}(U)) = P(Q^{-1}(U)) - \frac{\int_{\partial S_{T(X_0)}} \mathrm{d}x}{\int_{S_R} \mathrm{d}x} \frac{\int_{\partial S_{T(X_0)}} a(x)\alpha(B \oplus x, W_B)\mathrm{d}x}{\int_{S_R} p(B \oplus x, W_B)\mathrm{d}x} + o(a) \tag{7}$$

and the proposed global statistics are consequently biased. In particular, the bias of the mean value tends to 0 when the relative non-stationarity $a(x)$ tends to 0. Therefore, the proposed estimation is asymptotically unbiased when $a(x) \to 0$.

The proposed statistic is local, in the sense that it depends on the position of each typical point conditionally to the fact that it belongs to a given bounded subset. Therefore, it does not depend on the absolute value of the non-stationarity, but only on its relative variations inside the subset.

To illustrate this fact, consider the case of an inhomogeneous Poisson point process X on the horizontal line with intensity $\lambda(x)$. Classical tests of Poisson distribution in the homogeneous case are based on the distribution of the distance between two consecutive segments.

Then, for a given tested distance R, a group of connected points is composed of points whose consecutive distances are shorter than R. Let the typical point of such a group be the left point of the group, l the length of the group, Y_0 the left neighbour point of X_0 and Y_1 the right neighbour point of the group. Under the model, the probability density of observing the length $X_0 - Y_0$ conditionally to the fact that $X_0 - Y_0 > R$ is $p(X_0 - Y_0 \mid R) = \lambda(X_0) \exp\{-\int_{Y_0+R}^{X_0} \lambda(x)\mathrm{d}x\}$ whereas the probability of observing the typical point at X_0 knowing Y_0, Y_1 and l is $q(X_0 \mid Y_0, Y_1, l, R) = \frac{\lambda(X_0)}{\int_{Y_0+R}^{Y_1-R-l} \lambda(x)\mathrm{d}x}$.

If $M = (Y_0 + Y_1 - 2R - l)/2$ is the middle of the segment $[Y_0 + R, Y_1 - R - l]$, let us denote $\lambda(x) = \lambda(M)(1 + a(x))$. $a(x)$ is the relative variation of $\lambda(x)$ around M. The two density probabilities become $p(Y_0 - X_0 \mid R) = \lambda(M) \exp\{-\lambda(M) \int_{Y_0+R}^{X_0} (1 + a(x))\mathrm{d}x\}$ and $q(X_0 \mid Y_0, Y_1, l, R) = \frac{1+a(X_0)}{\int_{Y_0+R}^{Y_1-R-l} (1+a(x))\mathrm{d}x}$.

For small values of $a(x)$, $q(X_0)$ is the density of the uniform distribution, up to a bounded term in $a(x)$ whereas the value of $p(X_0)$ can vary greatly even if $a(x)$ is small. This comes from the fact that $p(X_0)$ depends on $\lambda(M)$, i.e. the value of the intensity itself, which be subject to large variations, even if its relative variations are small.

3. Tests on simulated examples

3.1. Neyman–Scott process

The proposed procedure was first applied on a realization of a Neyman–Scott model. Parent points followed a Poisson point process with constant intensity 100. Five daughter points were spread independently from each other around each parent point, uniformly on the square of side length 0.03 centered on the parent point.

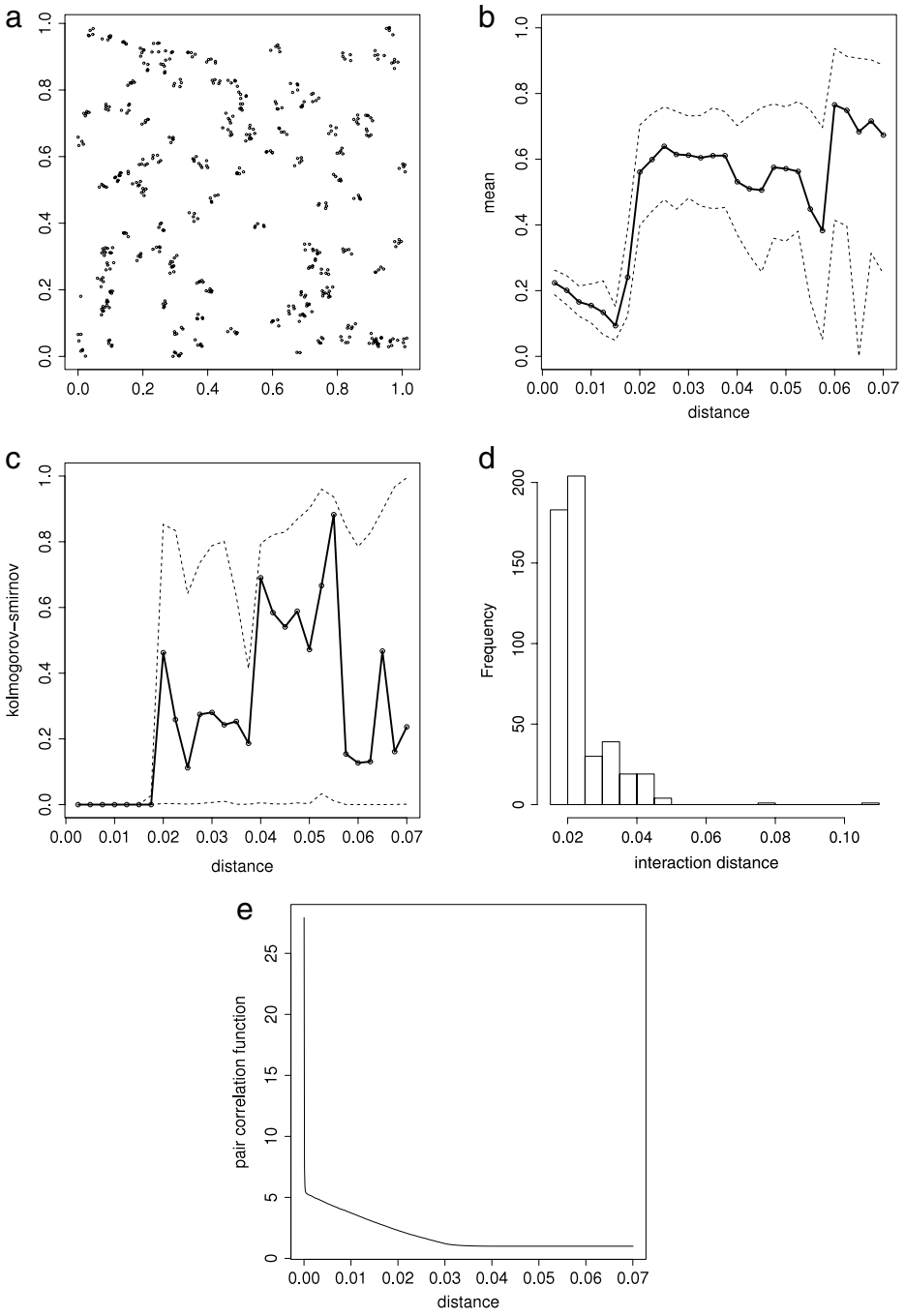

Fig. 2. Test of the estimation procedure based on a Neyman–Scott model. (a) Simulated data set, parent point process is Poisson with intensity λ, 5 daughter points are spread independently around each parent point in a square of side length 0.03. (b) Evolution of the global statistics based on the mean value of p-values of individual statistics with respect to tested distances, together with its individual confidence band. (c) Evolution of Kolmogorov–Smirnov statistics from p-values of individual statistics with respect to tested distances, together with its individual confidence band. (d) Distribution of the block-bootstrap estimates of the interaction radius. (e) Pair correlation function for the Neyman–Scott model defined in (a).

The process of daughter points was observed in a square sampling window of side length 1. Fig. 2(a) presents a realization of the process. A group is then defined as the set of daughter points around a given parent point, or the union of such sets if their distances (defined as the minimum distance) are less than 0.03.

Fig. 2(b) and (c) present the results of the proposed global tests. Plain lines correspond to the changes in each global test with respect to R, dotted lines are the individual confidence bounds at 95% obtained using a random block-bootstrapping method (Lahiri, 1999), the side length of a square block being 0.2. 500 bootstrap pseudo-repetitions were used. The two

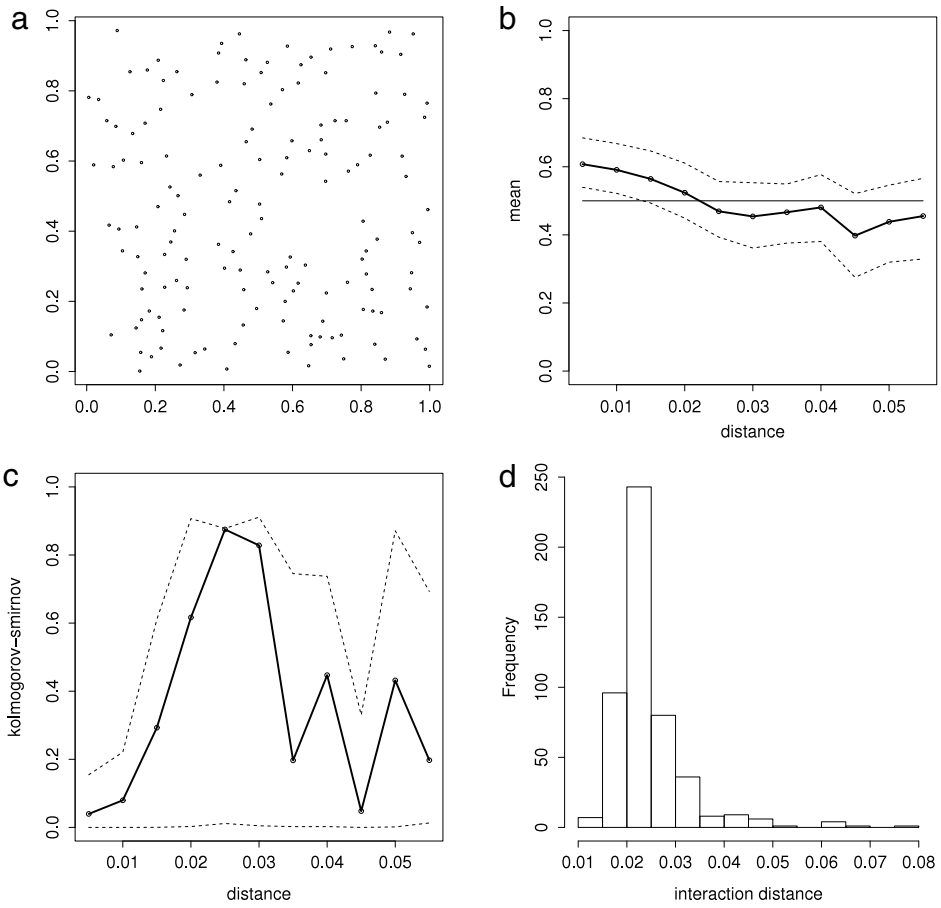

Fig. 3. Test of the estimation procedure based on a hard-core model. (a) Simulated data set of a Strauss model with hard-core radius 0.025 and intensity 200. (b) Evolution of the global statistics based on the mean value of p-values of individual statistics with respect to tested distances, together with its individual confidence band. (c) Evolution of Kolmogorov–Smirnov statistics from p-values of individual statistics with respect to tested distances, together with its individual confidence band. (d) Distribution of the block-bootstrap estimates of the interaction radius.

statistics show globally the same results. Tests based on the Kolmogorov–Smirnov test (Fig. 2(c)) show however a much sharper change, allowing a much easier graphical estimation of R_0. The confidence bands are on the other side smaller for the mean p-value statistics.

Estimation based on the curve modelling of the mean p-value leads to $\hat{R}_0 = 0.0203$, with a confidence band at 95% equal to [0.0199; 0.0227]. The estimator is biased, the confidence bound does not contain the theoretical value of 0.03. This is due to the low dependence between points when $r > 0.02$, which is not detected by the method. This dependence, measured through the pair correlation density (Stoyan et al., 1995) is equal to

$$\rho(r) = 1 + \frac{4}{20\pi}\frac{f(r)}{r} \tag{8}$$

if $f(r)$ denotes the probability density of the distance of two random points in the square of side length 0.03. This pair correlation density, shown on Fig. 2(e), is close to 1 for $r > 0.02$.

3.2. Hard-core process

The preceding test was based on a realization where clusters of points were randomly spread in space. We focus now on a drastically different example where a cluster is composed of only one point and where the cluster spatial distribution is very regular: a hard-core process. Fig. 3(a) presents a realization of such a point process with intensity 200 on the unit square. Interaction radius is equal to 0.025. It was obtained by exact simulation, using the procedure "rStrauss" from the package "spatstat" in R (Baddeley and Turner, 2005).

Fig. 3(b) and (c) present the changes in the two proposed statistics together with their individual confidence intervals at the 5% level using a block-bootstrap procedure with 500 pseudo-repetitions and a square block of size length 0.2.

Estimated interaction radius estimation was equal to $\hat{R}_0 = 0.0224$ and a confidence interval equal to [0.015, 0.061]. The distribution of the bootstrap pseudo-repetitions are given in Fig. 3(d). The procedure accurately estimates the interaction radius, though with a large upper confidence value.

4. Estimation of the interaction radius between trees

4.1. Local dependence analysis in a patch of vegetation

Fig. 4(a) presents the spatial distribution of trees in a patch of vegetation in a tiger bush in South-West Niger (see Couteron et al., 2000, for details on the study site). Trees, observed in a 50 m × 50 m area, are mainly *C. micranthum* G. Don (about three-fourths of the population). If trees present an aggregated pattern at a global scale, as can be seen on Fig. 4(a), they present at a local scale a regular pattern, as shown by the position of the nearest neighbour distance function (Diggle, 2003) which lies outside the bounds of the individual confidence band built under Complete Spatial Randomness (CSR) at the 5% level (Fig. 4(e)).

The proposed statistic, the changes with R in the mean and the p-value of the Kolmogorov–Smirnov statistics computed from the p-values of the individual statistics respectively, are presented in Fig. 4(b) and (c), together with their individual confidence bounds computed by block bootstrapping with 10 m × 10 m blocks. Estimated R_0 is $\hat{R}_0 = 0.8$ m with an estimated confidence interval [0.61, 1.3] at 5% level.

The distribution of $\hat{R}_0$, obtained by block bootstrap, is shown in Fig. 4(e). It peaks around the estimated value, with a relatively small standard deviation equal to 0.22 m, but presents large values with low probability densities.

4.2. Local dependence analysis for trees in a periodic vegetation pattern

Fig. 5(a) presents the spatial distribution of a small tree, *pterocarpus lucens Lepr.* in a 320 m × 320 m area in NW Burkina-Faso. The overall tree pattern is marked by periodic bands devoid of trees. The wavelength of the pattern of bands is approximatively 60–70 m (see Couteron, 2001, for more details). Trees present here an aggregated pattern, as shown with the nearest neighbour distance function. The observed nearest neighbour distance function lies outside or on the upper bound of the individual confidence band built under CSR at the 5% level as long as the distance is less than 1.5 m (Fig. 5(d)).

The proposed statistics, the changes with R of the mean and the p-value of the Kolmogorov–Smirnov statistics computed from the p-values of the individual statistics respectively, are presented in Fig. 5(b) and (c), together with their individual confidence bounds computed by block bootstrapping with 10 m × 10 m blocks. Unlike the preceding case, no consistent variation of the mean statistic with the distance appears. It mainly oscillates around 0.5, this value being outside the confidence band for R around 3 m only. The same pattern appears with the Kolmogorov statistic. Consequently, no local dependence is detected and no R_0 value can be estimated.

4.3. Conclusion

These two data sets were already studied in order to test whether they were locally independent, i.e could be considered as realizations of inhomogeneous Poisson point processes (Couteron et al., 2003). The results obtained with our method lead to the same conclusions but the present method provides additional information, an estimation of the interaction radius R_0.

5. Estimation of the interaction radius between northern gannets

5.1. Data and objective

Our method was also applied to seabird count data collected during aerial surveys in the Bay of Biscay, (100 000 km^2 off the French Atlantic coast, see Certain et al. (2007) for details on the study site and data). Sightings of northern gannets were collected during 30 000 km of monthly aerial strip transect carried out in winter 2001–2002 from October to March. Each sighting corresponds to one or several individuals: in the latter, individuals are very close to each other (<40 m). Data recording was achieved using a standard strip transect method (Briggs et al., 1985a,b). Fig. 6(a) presents the observation collected in January 2002 as an example. Data collected at each date were considered as independent observations, and we supposed that the interaction radius did not vary between dates. One can notice the presence of first-order non-stationarity in sightings distribution, fewer animals being observed in the center of the study area. Northern gannets are known to feed in groups, and it is commonly accepted that they share information on the location of prey patches, at least passively (Nelson, 2002). Our aim was to define at which interaction radius R_0 a gannet could share information and interact with a neighbour, in order to define clusters of foraging gannets. Theoretically, this interaction radius R_0 is bounded by the perception (visual) capacities of gannets.

5.2. Estimation of the interaction radius between gannets

We based the R_0 estimation on the coordinates of the projected groups on the transect axis, such that the analysis is based on the relative position of coordinate X_{2i} with respect to coordinates X_{2i-1}, X_{2i+1} once clusters along a transect are

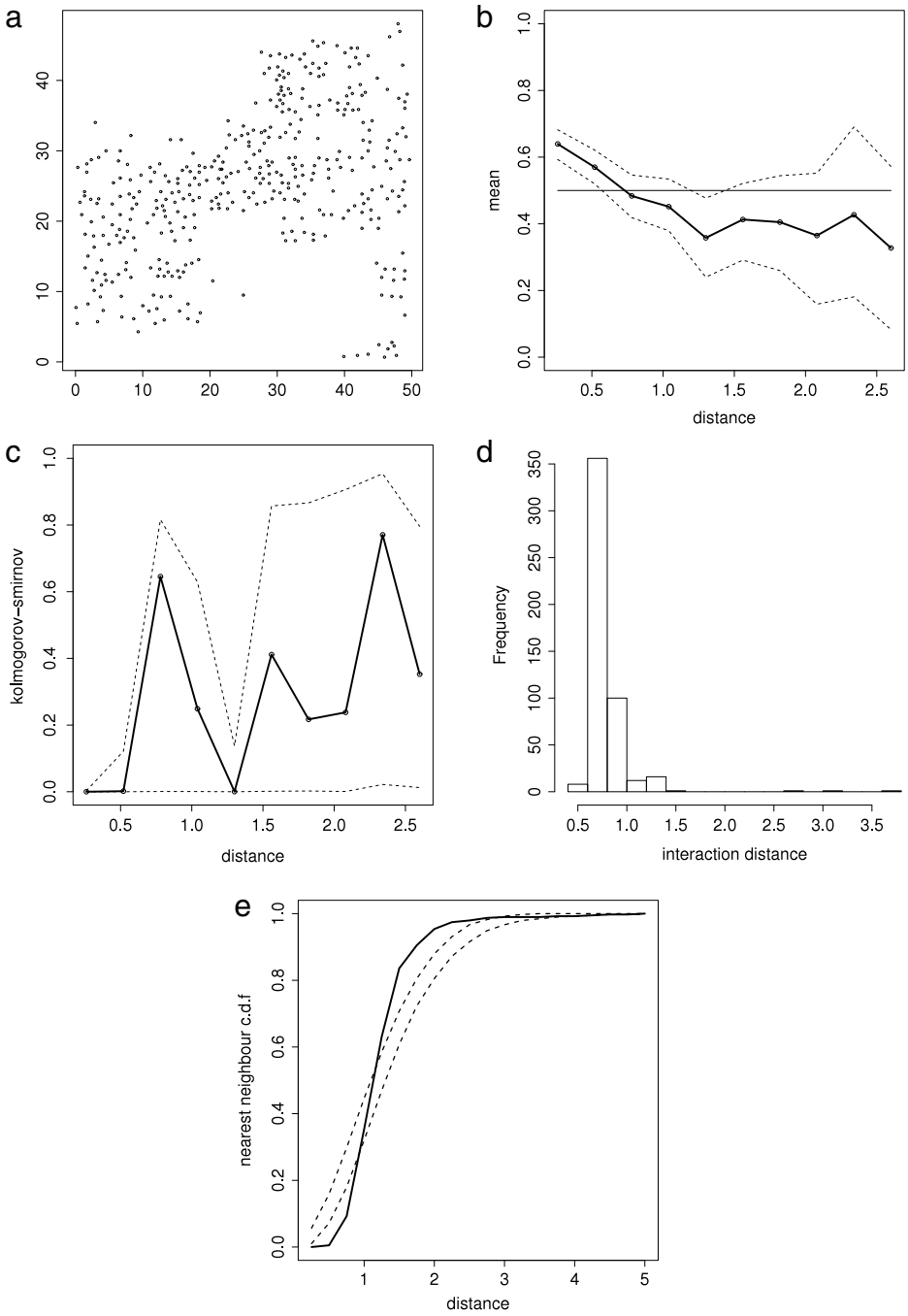

Fig. 4. Estimation of the interaction radius between trees in a tiger bush forest. (a) Tree distribution in a 50 m × 50 m area. (b) Evolution of the global statistics based on the mean value of p-values of individual statistics with respect to tested distances, together with its individual confidence band. (c) Evolution of the Kolmogorov–Smirnov statistics of p-values individual statistics with respect to tested distances, together with its individual confidence band. (d) Distribution of the block-bootstrap estimates of the interaction radius. (e) Nearest neighbour distance distribution computed on map (a), together with its individual confidence band under independence assumption.

ordered with respect to their positions along the transect axis. Samples verifying $|X_{2i+1} - X_{2i-1}| > 10$ km were excluded from the analysis to ensure the asymptotic estimator convergence.

Block bootstrap was performed conditionally to the observation date. Block length was equal to 40 km. Fig. 6(b) and (c) show the variations with respect to R of the mean statistic and of the Kolmogorov–Smirnov statistic together with their confidence bands at 5%. One may notice a regular increase in the Kolmogorov statistic, with a sharp increase in its confidence bound (Fig. 6(c)). The confidence bound of the mean value statistic changes more regularly, but remains large

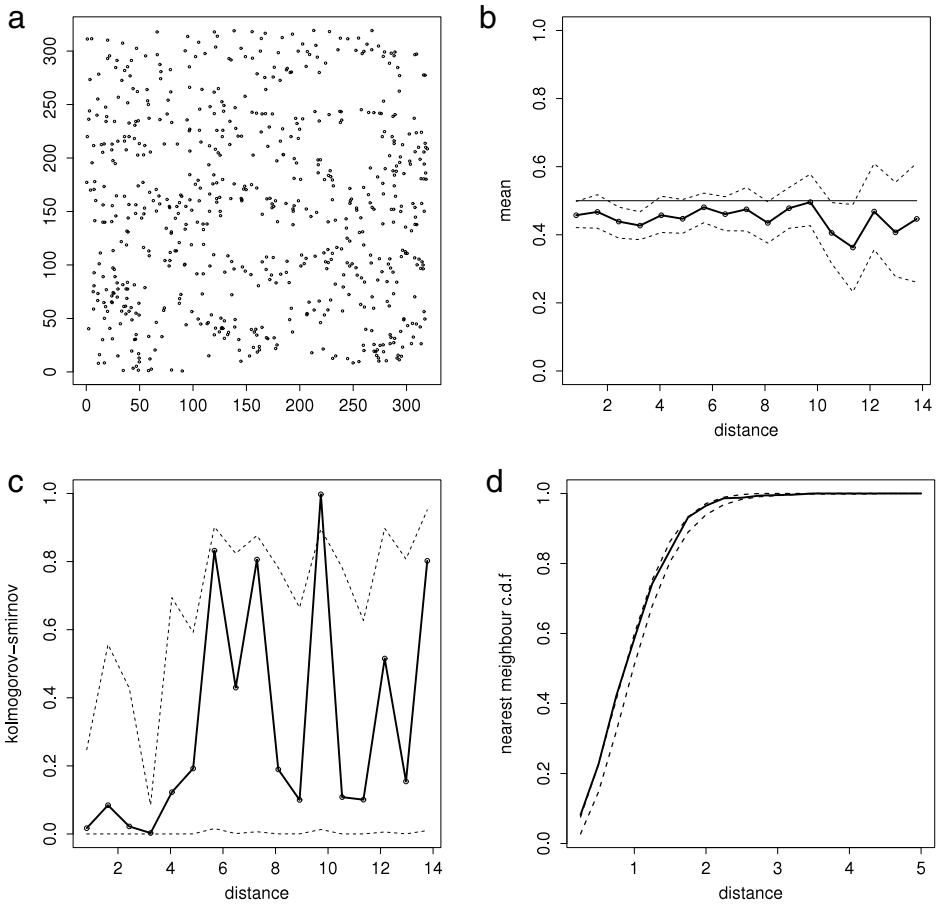

Fig. 5. Estimation of the interaction radius between pterocarpus lucens in a Burkina-Faso forest. (a) Tree distribution in a 320 m × 320 m area. (b) Evolution of the global statistics based on the mean value of p-values of individual statistics with respect to tested distances, together with its individual confidence band. (c) Evolution of Kolmogorov–Smirnov statistics from p-values of individual statistics with respect to tested distances, together with its individual confidence band. (d) Nearest neighbour distance distribution computed on map (a), together with its individual confidence band under independence assumption.

enough to allow very different mean value shapes, and so a large confidence bound of the estimated R_0. Estimated value was $\hat{R}_0 = 2512$ m with a confidence interval of [1810, 6400] at 95% level. Fig. 6(d) shows the distribution of $\hat{R}_0$ obtained by block bootstrap. The distribution peaks at around 2507 m, but a second mode for large R values is responsible for a large confidence interval.

5.3. Effect of non-stationarity

To check whether the non-stationarity affected the estimation, we looked for a spatial dependence between the estimated p-values of the tests of random distribution of each cluster. First, to check whether or not the relative position of a group with respect to its neighbours was influenced by the non-stationarity, for example if it was more often near the neighbour having the largest local intensity, we computed the variogram of the p-values and tested by randomization whether any spatial dependence existed. Second, to test whether the process intensity could affect the interaction radius, we computed the variogram of the absolute value of the local p-values minus 0.5, $|\hat{U}(s) - 0.5|$. Fig. 6(e) and (f) show these variograms together with their confidence bounds. No spatial dependence was detected. To check whether the value of the interaction radius depended on the observation date, we plotted the mean and variance of the p-values for each observation date (not shown). No date effect was detected.

5.4. Discussion on the estimated interaction radius

First we should mention that the interaction radius estimated in this case study is a projected distance in one dimension of a biological distance in two dimensions. So we must keep in mind that the real, biological distance is greater than the estimated one. Nonetheless, our results are coherent with the biology of the animal under study: gannets, like all birds, have an acute sense of sight (Nelson, 2002). In procellariiforms, Haney et al. (1992) calculated that the mean horizontal

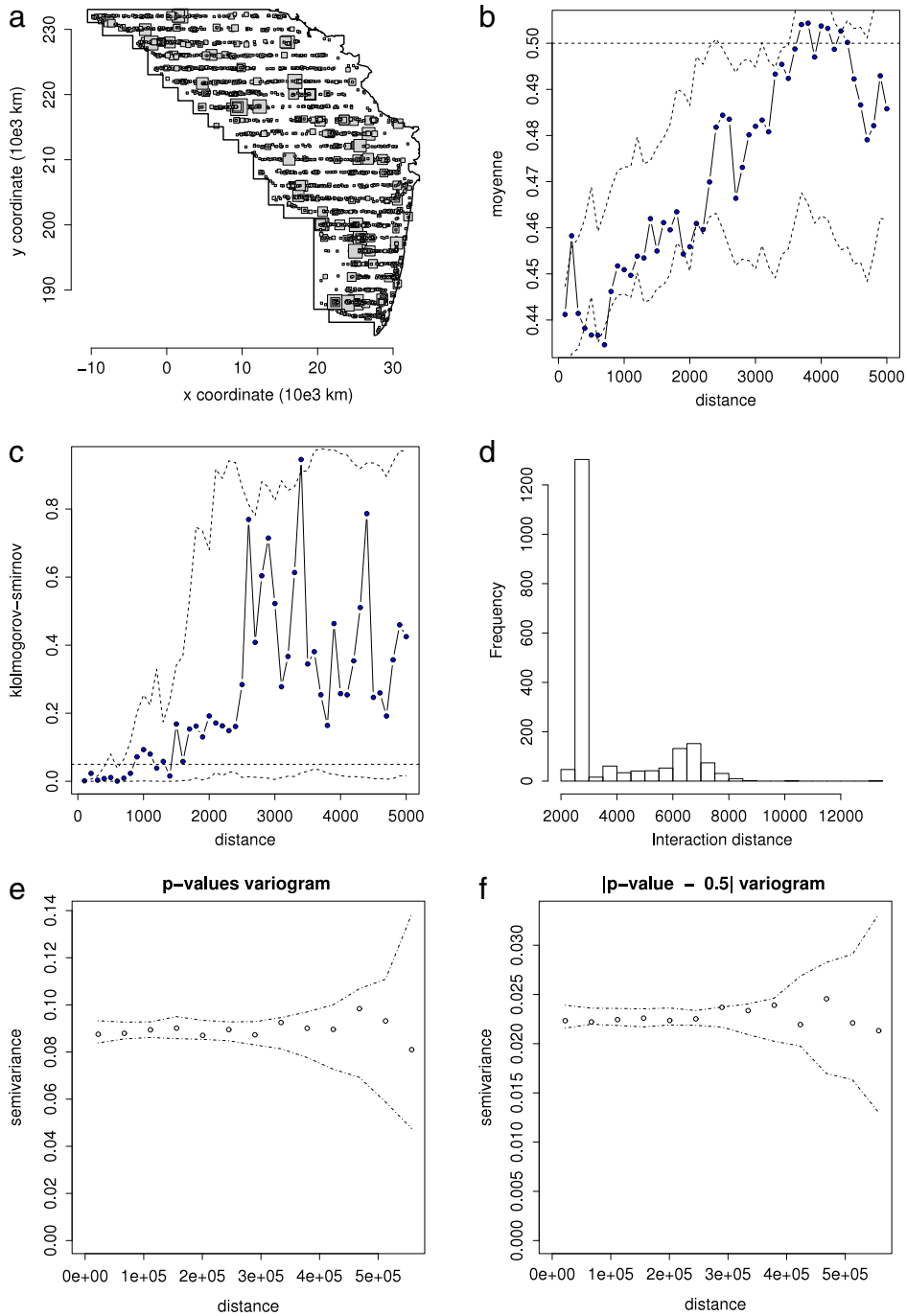

Fig. 6. Estimation of the interaction radius between gannet sightings in the Bay of Biscaye. (a) Spatial distribution of sightings observed in January 2002. (b) Evolution of the global statistics based on the mean value of p-values of individual statistics with respect to tested distances, together with its individual confidence band. (c) Evolution of Kolmogorov–Smirnov statistics from p-values of individual statistics with respect to tested distances, together with its individual confidence band. (d) Distribution of the block-bootstrap estimates of the interaction radius. (e) Variogram of the individual p-values p_i together with individual confidence band under independence. (f) Variogram of $|p_i - 0.5|$ with individual confidence band under independence.

distance over which procellariiforms seabirds were recruited visually to a feeding flock was around 4–5 km, which is highly consistent with our projected distance estimated on gannets. Moreover, these birds live in open seas where objects can be detected at great distances without obstacles. If the interaction radius R_0 is bounded by the perception abilities of gannets, it is also a function of the reward one individual can expect by joining foraging neighbours. This reward depends on the distance separating both individuals. Therefore, two limits exist, the perception distance and the distance up to which a bird is ready to join another bird. These two limits vary according to environmental conditions, physics and resources and

can lead to estimation uncertainty. By comparison, studies on other, less dynamic systems such as plant distributions shown in Section 4 resulted in much smaller confidence intervals.

6. Conclusion

Our method shows that the interaction radius can be estimated without assuming an explicit interaction model. The proposed method does not rely on the structure of the points inside the groups, consequently the interaction radius can be estimated without having to explicit a model for the point pattern. This can be very useful in behavioural studies, where the first question is to determine which individuals are interacting and how they are interacting before analyzing the spatial structure inside each group.

Interaction radius estimates were close to the theoretical ones for the simulated examples, in the expected range for the data sets. If the distribution mode is clear, the tail of the distribution slowly decreases and therefore the confidence intervals of the interaction radius remain very large. Because the number of individual statistics drops dramatically as the value of the tested interaction radius increases, the power of the global test decreases accordingly which contributes to an inflation of the upper limit of the confidence interval.

Scan statistics and our method are complementary methods. (i) The two methods differ mainly by the way the departure from the Poisson distribution is used. Scan statistics tests whether the number of points inside an area of predetermined shape, generally a circle, is in accordance with the Poisson assumption. We propose to measure whether the position of a given group of points is random in an area defined by its neighbours. (ii) The objectives of the two methods are different, scan statistics being primarily used to detect groups whereas we focus only on interaction radius estimation. However, the objective of either of the two methods can be seen as a sub-objective of the other one, since (a) one can study the distances between groups to estimate an interaction radius when scan statistics are used, (b) one can define groups by splitting the pattern with respect to the interaction distance when applying our method.

The two methods will detect the same groups as soon as the groups are sufficiently far apart such that (i) the scan statistics will be able to detect that the number of points is higher than expected in circles englobing the groups, (ii) our method will find a distance beyond which group positions are independent conditionally to their neighbours. This should be the case with the example based on the Neyman–Scott process above. Our method will fail if the interaction radius is not constant but spatially varying. The scan statistics will fail if no density variation is present, as with hard-core processes where each group is composed of only one point.

Estimating the interaction radius can be performed in the two cases. Our method has been developed in order to get an estimator and a measure of its variability. One also can estimate an interaction radius using the groups found by scan statistics. For that purpose, one must first use scan statistics so as to split all the set of points into separate groups, then analyze distances between groups to estimate an interaction radius. The problem will appear when estimating its variability, either its variance or a confidence interval, because the estimator is given as a function of the distance between the estimated groups. Adapted statistical procedures have then to be developed.

Considering future studies of animal or plant distribution, ecologists can easily assume that the interaction radius depends primarily on abilities of the individuals to detect other fellows in given environmental conditions. This parameter is relatively stable whereas groups can be quite variable, either in size or in structure. It is then interesting, in an exploratory stage with a minimum of assumptions, to estimate the interaction radius and its uncertainty, before going into a more specific modelling of the group structure itself.

References

Baddeley, A., Turner, R., 2005. Spatstat: an **R** package for analyzing spatial point patterns. Journal of Statistical Software 12 (6), 1–42.
Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B 36, 192–236.
Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics 29 (4), 1165–1188.
Briggs, K.T., Tyler, W.B., Lewis, D.B., 1985a. Aerial survey for seabirds: methodological experiment. Journal of Wildlife Management 49, 412–417.
Briggs, K.T., Tyler, W.B., Lewis, D.B., 1985b. Comparison of ship and aerial surveys of bird at sea. Journal of Wildlife Management 49, 405–411.
Callaway, R.M., 1995. Positive interactions among plants. Botanical Review 61, 306–349.
Certain, G., Bellier, E., Planque, B., Bretagnolle, V., 2007. Characterising the temporal variability of the spatial distribution of animals: an application to seabirds at sea. Ecography 30 (5), 695–708.
Couteron, P., Mahamane, A., Ouedraogo, P., Seghieri, J., 2000. Differences between banded thickets (tiger bush) at two sites in West Africa. Journal of Vegetation Science 11, 321–328.
Couteron, P., 2001. Using spectral analysis to confront distributions of individual species with an overall periodic pattern in semi-arid vegetation. Plant Ecology 156, 229–243.
Couteron, P., Seghieri, J., Chadœuf, J., 2003. A test for spatial relationships between neighbouring plants at various densities. Journal of Vegetation Science 14, 163–172.
Couzin, I.D., Krause, J., James, R., Ruxton, G.D., Franks, N.R., 2002. Journal of Theoretical Biology 218 (1), 11.
Dacunha-Castelle, D., Duflo, M., 1993. Probabilités et Statistiques; 2 problèmes à temps mobile. Masson, Paris, France.
Demattei, C., Molinari, N., Daurès, J.P., 2006. Arbitrarily shaped multiple spatial cluster detection for case event data. Computational Statistics and Data Analysis 51 (8), 3931–3945.
Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. Journal of the Royal Statistical Society. Series C 47, 299–350.
Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns. Academic Press, London, UK, p. 159.
Duczmal, L., Assunção, R., 2004. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. Computational Statistics and Data Analysis 45 (2), 269–286.
Duczmal, L., Cançado, A.L.F., Takahashi, R.H.C., Bessegato, L.F., 2007. A genetic algorithm for irregularly shaped spatial scan statistics. Computational Statistics and Data Analysis 52 (1), 43–52.

Fortin, M.J., Dale, M., 2005. Spatial Analysis. A Guide for Ecologists. Cambridge University Press, Cambridge, p. 365.
Guyon, X., 1993. Champs aléatoires sur un réseau. Masson, Paris, p. 226.
Haney, J.C., Fristrup, K.M., Lee, D.S., 1992. Geometry of visual recruitment by seabirds to ephemeral foraging flocks. Ornis Scandinavica 23, 49–62.
Holmgren, M., Scheffer, M., Huston, M.A., 1997. The interplay of facilitation and competition in plant communities. Ecology 78, 1966–1975.
Kelsall, J.E., Diggle, P.J., 1995. Kernel estimation of relative risk. Bernoulli 1, 3–16.
Kulldorff, M., 1997. A spatial scan statistic. Communications in Statistics. Theory and Methods 6, 1481–1496.
Kulldorff, M., Nagarwalla, N., 1995. Spatial disease clusters: detection and inference. Statistics in Medecine 14, 799–810.
Lahiri, S.N., 1999. Theoretical comparisons of block bootstrap methods. Annals of Statistics 27 (1), 386–404.
Ling, R.F., 1973. A probability theory of cluster analysis. Journal of the American Statistical Association 68 (341), 159–164.
Manly, F.J., 1991. Randomization, Bootstrap and Monte-Carlo Methods in Biology. Chapman and Hall, London, UK.
Möller, J., Waagepetersen, R.P., 2004. Statistical Inference and Simulation for Spatial Point Processes. Chapman and Hall, Boca Raton, p. 315.
Nelson, J.B., 2002. The Atlantic Gannet, 2nd ed. Fenix Books, Norfolk, p. 396.
Pélissier, R., Goreaud, F., 2001. A practical approach to the study of spatial structure in the cases of heterogeneous vegetation. Journal of Vegetation Science 12, 99–108.
Renshaw, E., Särkkä, A., 2001. Gibbs point processes for studying the development of spatial–temporal stochastic processes. Computational Statistics and Data Analysis 36 (1), 85–105.
Särkkä, A., Tomppo, E., 1998. Modelling interactions between trees by means of field observations. Forest Ecology and Management 108 (1–2), 57–62.
Shimatani, K., Kubota, Y., 2004. Spatial analysis for continuously changing point patterns along a gradient and its application to an *Abies sachalinensis* population. Ecological Modelling 180 (2–3), 359–369.
Sterner, R.W., Ribic, C.A., Schatz, G.E., 1986. Testing for life historical changes in spatial patterns of four tropical tree species. Journal of Ecology 74, 621–633.
Stoyan, D., Kendall, W., Mecke, J., 1995. Stochastic Geometry and its Applications, 2nd ed.. John and Sons, Chichester.
Stoyan, D., Stoyan, H., 1999. Non-homogeneous Gibbs process models for forestry—a case study. Biometrical Journal 40 (5), 521–531.
Veit, R.R., 1999. Behavioural responses by foraging petrels to swarms of krill Euphosia superba. Ardea 87, 41–50.