TEST (2022) 31:416–441
https://doi.org/10.1007/s11749-021-00783-x

ORIGINAL PAPER

Spatial distribution of invasive species: an extent of occurrence approach

Alberto Rodríguez-Casal[1] · **Paula Saavedra-Nieves**[1]

Received: 1 February 2021 / Accepted: 26 July 2021 / Published online: 16 August 2021

Abstract
Ecological Risk Assessment faces the challenge of determining the impact of invasive species on biodiversity conservation. Although many statistical methods have emerged in recent years in order to model the evolution of the spatio-temporal distribution of invasive species, the notion of extent of occurrence, formally defined by the International Union for the Conservation of Nature, has not been properly handled. In this work, a novel and flexible reconstruction of the extent of occurrence from occurrence data will be established from nonparametric support estimation theory. Mathematically, given a random sample of points from some unknown distribution, we establish a new data-driven method for estimating its probability support S in general dimension. Under the mild geometric assumption that S is $r-$convex, the smallest $r-$convex set which contains the sample points is the natural estimator. A stochastic algorithm is proposed for determining an optimal estimate of r from the data under regularity conditions on the density function. The performance of this estimator is studied by reconstructing the extent of occurrence of an assemblage of invasive plant species in the Azores archipelago.

Keywords Ecological risk · Extent of occurrence (EOO) · Invasive species · Spacing · Testing $r-$convexity

Mathematics Subject Classification 62G05 · 62G07 · 62G10 · 62G20

A. Rodríguez-Casal and P. Saavedra-Nieves acknowledge the financial support of Ministerio de Economía y Competitividad and Ministerio de Ciencia e Innovación of the Spanish government under grants MTM2016-76969P, MTM2017-089422-P, PID2020-118101GB-I00 and PID2020-116587GB-I00 and ERDF. The authors are grateful to Rosa M. Crujeiras, Antonio Cuevas and Ignacio Munilla Rumbao for their help, and two anonymos referees for their useful and constructive comments. They also thank the computational resources of the CESGA Supercomputing Center.

✉ Paula Saavedra-Nieves
paula.saavedra@usc.es

Alberto Rodríguez-Casal
alberto.rodriguez.casal@usc.es

[1] Universidade de Santiago de Compostela, Santiago de Compostela, Spain

1 Introduction

Ecological Risk Assessments (ERA) are performed to evaluate the likelihood of negative ecological effects as a result of exposure to a biological, physical or chemical factor that provokes adverse responses in the environment. Then, a remarkable challenge of ERA is to analyze the impact of invasive species on biodiversity conservation or habitat protection. According to Martínez-Minaya et al. (2018), statistical (and specially Bayesian) modeling of the distribution of (invasive) species has increased substantially in the last years in order to understand their spatio-temporal dynamics. A key issue to determine the species evolution is to characterize its extent of occurrence (EOO). Although EOO is one of the most widely handled concepts in natural reserve network designs involving occurrence data, it has not been considered in the literature under this perspective yet. The International Union for the Conservation of Nature[1] (IUCN) establishes the EOO as a key measure of extinction risk. Roughly speaking, the IUCN defines the EOO as the area contained within the shortest continuous imaginary boundary which can be drawn to encompass all the known, inferred or projected sites of present occurrence of a taxon, excluding cases of vagrancy. For a complete review on this subject, see IUCN (2012) and Rondinini et al. (2006).

The problem of EOO reconstruction will be illustrated via the analysis of a real dataset containing 740 geographical coordinates (or occurrences) for 28 species of terrestrial invasive plants distributed in two of the Azorean islands (Terceira and São Miguel) from 2010 until 2018. In Fig. 1, a satellite image of major Azorean islands (top, left) and five of the invasive species are shown (bottom). The 740 geographical locations (slightly jittered) are represented on the map of Terceira and São Miguel islands in Fig. 1 (top, right). This dataset is available from the Global Biodiversity Information Facility (GBIF) website (see GBIF.org 2019).

An initial estimation of the EOO for this assemblage of invasive plants was obtained from GeoCAT.[2] It is an open source, browser-based tool endorsed by IUCN that allows us to reconstruct the EOO from the geographical locations of species or taxon. Users can quickly combine data from multiple sources including GBIF datasets which can be easily imported. The GeoCAT reconstruction of the EOO for the assemblage of plant species is given by the convex hull of the sample of the 740 coordinates, $H(\mathcal{X}_{740})$. Mathematically, $H(\mathcal{X}_{740})$ is the smallest convex set that contains the original random sample $\mathcal{X}_{740}$. In fact, it is computed as the intersection of all half spaces containing $\mathcal{X}_{740}$. For more details, compare Fig. 5 (first row, left) and Fig. 5 (second row, left). Note that this EOO estimation presents some practical limitations because a marine area is inside the $H(\mathcal{X}_{740})$. Obviously, none of the plant species considered here can occur in open sea which should remain outside the EOO. Therefore, convexity can be a too restrictive shape condition to be assumed in some situations. Of course, an expert can disconnect the original dataset in different spatially homogeneous groups and work separately. But, sometimes, it is not obvious how to split the data and/or finding groups in the dataset is one of the objectives of the study.

[1] IUCN website: www.iucn.org

[2] GeoCAT website: http://geocat.kew.org/

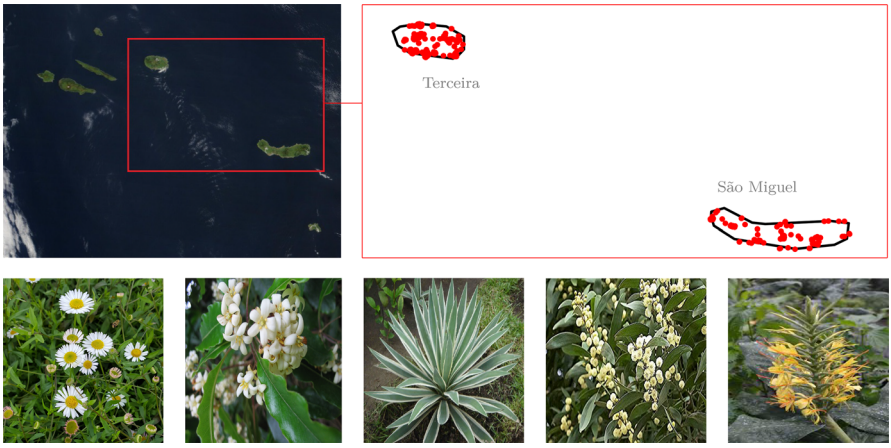

Fig. 1 Location of Terceira and São Miguel islands in the Azores Archipelago, NASA satellite image (top, left). The enlarged area (top, right) shows the 740 geographical locations used to reconstruct the EOO of an assemblage of 28 invasive plant species including: Erigeron karvinskianus, Pittosporum undulatum, Agave americana, Acacia melanoxylon, Hedychium gardnerianum (bottom, from left to right)

Our goal is to propose a novelty, more realistic, flexible and automatic EOO reconstruction using occurrence data from nonparametric support estimation perspective. This methodological approach has proved to be useful in different disciplines such as image analysis (see Rodríguez-Casal and Saavedra-Nieves 2016), quality control (see Devroye and Wise 1980 or Chevalier 1976) or animals home range estimation (see De Haan and Resnick 1994 or Baíllo and Chacón 2018). However, the problem of studying the spatio-temporal distribution of invasive species from the EOO estimation has not been proposed yet.

In general, support estimation deals with the problem of reconstructing the compact and nonempty support $S \subset \mathbb{R}^d$ of an absolutely continuous random vector X from a random sample $\mathcal{X}_n = \{X_1, ..., X_n\}$ (see Cuevas and Fraiman 2010 for a complete survey on the subject). Of course, when the support S is assumed to be convex, then the convex hull of the sample points, $H(\mathcal{X}_n)$, provides a natural support estimator. See Schneider (1988, 2014), Dümbgen and Walther (1996) or Reitzner (2003), for thorough analysis of this estimator. This estimator is indeed simple and fully data driven, but it may not be suitable for some practical situations, failing to provide a satisfactory support estimator when S has holes or it is disconnected as in the example of invasive plants in Azores archipelago where the occurrences are distributed within two different islands.

In this work, we will propose a new data-driven support estimator for general dimension and, as a consequence, an original, realistic and easy to use EOO reconstruction that will overcome the limitations derived from convexity restriction. Concretely, we assume that the support S satisfies the $r-$convexity shape condition for $r > 0$, a much more flexible geometrical property than convexity as it will be shown. Our proposal considers the smallest r-convex set containing $\mathcal{X}_n$ ($r-$convex hull of $\mathcal{X}_n$, namely $C_r(\mathcal{X}_n)$) as the natural estimator for the unknown support. This estimator is well known

in the computational geometry literature for providing reasonable global reconstructions if the sample points are (approximately) uniformly distributed on the set S (see Edelsbrunner 2014). In fact, despite being $r-$convexity a more general condition than convexity, $C_r(\mathcal{X}_n)$ can achieve the same convergence rates than $H(\mathcal{X}_n)$ as proved by Rodríguez-Casal (2007). However, this estimator presents an important disadvantage in practice: it depends on the commonly unknown parameter r. Although the influence of r can be considerable, it must be specified by the practitioner (see Joppa et al. 2016) or selected through practical procedures without theoretical guarantees (see Burgman and Fox 2003). For the example of invasive species in Azorean islands, Fig. 5 shows $C_r(\mathcal{X}_{740})$ for different values of r. Small values of r provide fragmented estimators (many isolated points and connected components) leading to an EOO reconstruction which resembles $\mathcal{X}_n$ (Fig. 5: second row, right). For an intermediate value of r, a realistic reconstruction of the EOO is obtained since sea areas are not inside the estimator (Fig. 5: third row, left). However, if large values of r are considered, then $C_r(\mathcal{X}_n)$ basically coincides with $H(\mathcal{X}_n)$ (Fig. 5: third row, right). Therefore, arbitrary choices of r may provide incongruous EOO estimations.

Most of the available results in the literature about support estimation make special emphasis on asymptotic properties, especially consistency and convergence rates, but they do not usually give any criterion for selecting the unknown parameter r in $C_r(\mathcal{X}_n)$ from the sample. The aim of this paper is to overcome this drawback and present a method for selecting the parameter r for the $r-$convex hull estimator from the available data. This problem has scarcely been studied in the statistical literature with just a couple of references available on the topic. First, Mandal and Murthy (1997) proposed a selector for r based on the concept of minimum spanning tree, but only consistency of the method was provided without considering optimality issues. Later, Rodríguez-Casal and Saavedra-Nieves (2016) proposed an automatic selection criterion based on a very intuitive idea for the selection of r but under the important restriction that the sample distribution is uniform. The idea for selecting r is as follows. According to Fig. 5 (bottom, right), sea areas are contained in $C_r(\mathcal{X}_n)$ if the selected r is too large. So, the estimator contains a large ball empty of sample points, see gray balls in Fig. 5 (top, left) and (bottom, right). Janson (1987) calibrated the size of this maximal ball (or spacing) when the sample distribution is uniform on S. Berrendero et al. (2012) used this result to test uniformity when the support is unknown. However, Rodríguez-Casal and Saavedra-Nieves (2016) followed an opposite approach. They assume that $\mathcal{X}_n$ comes from a uniform distribution on S and if a big enough spacing is found in $C_r(\mathcal{X}_n)$, then it is incompatible with the assumption that data are uniform. As a consequence, it is concluded that r is too large. Therefore, it seems natural to select the largest value of r compatible with the uniformity assumption on $C_r(\mathcal{X}_n)$.

Recently, Aaron et al. (2017) extended the results by Janson (1987) to the case where the data are generated from a density f that is bounded from below and Lipschitz continuous restricted to its bounded support. Here, we will use this extension in order to derive a test to decide, given a fixed $r > 0$, whether the unknown support S is $r-$convex with no more information apart from $\mathcal{X}_n$. In this case, if a large enough spacing is found in $C_r(\mathcal{X}_n)$, then the null hypothesis of $r-$convexity will be rejected. A new data-driven selector for the shape index r will be established from this test. Following the scheme in Rodríguez-Casal and Saavedra-Nieves (2016), it is proposed

to choose the largest value of r compatible with the $r-$convexity assumption. Once the parameter r is estimated from $\mathcal{X}_n$, a new data-driven support reconstruction, based on the estimator of r, will be proposed. As a consequence, a flexible reconstruction for the EOO, based on available data, will be obtained. Furthermore, when the support is convex, our EOO estimator will be similar to $H(\mathcal{X}_n)$. Therefore, the EOO definition given by IUCN is generalized.

This paper is organized as follows. Mathematical tools are introduced in Sect. 2. First, the geometric assumptions on S and the optimal value of the parameter r to be estimated are introduced. Then, the maximal spacing and its estimator are formally defined. Some regularity assumptions on f are also established. In Sect. 3, we propose a procedure for testing the null hypothesis that S is $r-$convex for a given $r > 0$. This test will play a key role in the definition of the consistent estimator of r. Then, a new data-driven estimator for the support S is proposed in Sect. 4 and it will be seen that it achieves the same convergence rates as the convex hull for estimating convex sets. The main numerical features involving the practical application of the algorithm are exposed in Sect. 5. Section 6 contains a simulation study in order to analyze the performances of the $r-$convexity test and the proposed estimator of the parameter r. In Sect. 7, the behavior of the new support reconstruction will be analyzed estimating the EOO of an assemblage of terrestrial plant species in two Azorean islands. Conclusions are exposed in Sect. 8. Finally, proofs of theoretical results are deferred to Sect. 9.

2 Mathemathical tools

Regularity conditions, namely shape assumptions on S, will be introduced next. In addition, we will discuss which is the optimal value of the shape index r to be estimated. Next, required conditions on the density function f will be also presented. Finally, basic notions on maximal spacings are established.

2.1 About geometric assumptions on S and the optimal r

In this work, S is assumed to be $r-$convex for some $r > 0$. Definition 1 establishes the formal definition of this geometric property.

Definition 1 A closed set $A \subset \mathbb{R}^d$ is said to be $r-$convex, for some $r > 0$, if $A = C_r(A)$, where

$$C_r(A) = \bigcap_{\{B_r(x):B_r(x)\cap A=\emptyset\}} (B_r(x))^c$$

denotes the $r-$convex hull of A and $B_r(x)$, the open ball with center x and radius r, whereas D^c denotes the complementary of D .

In practice, $C_r(\mathcal{X}_n)$ can be computed as the intersection of the complements of all open balls of radius larger than or equal to r that do not intersect $\mathcal{X}_n$. To illustrate the importance of a good choice of r, Fig. 2 shows the computation of $C_r(\mathcal{X}_{740})$ for

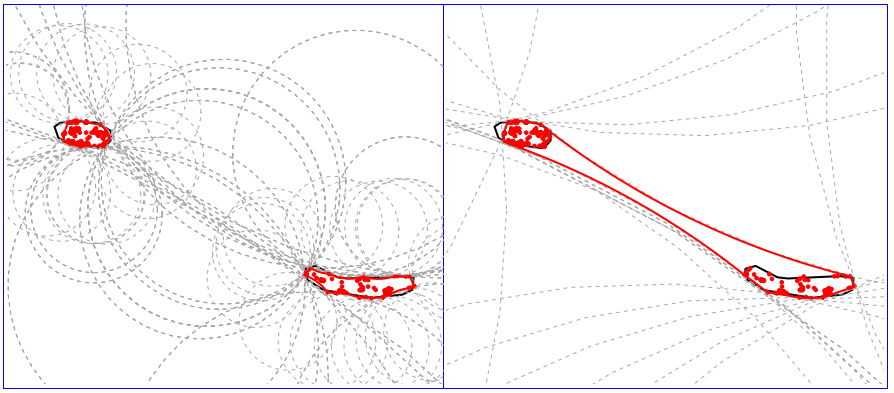

Fig. 2 $C_r(\mathcal{X}_{740})$ (red color) and $B_{r^*}(x)$ for $r^* \geq r$ (gray color) such that $B_{r^*}(x) \cap \mathcal{X}_{740} = \emptyset$ taking $r = 0.3$ (left) and $r = 5$ (right)

$r = 0.3$ (left) and $r = 5$ (right) for the example in Azorean islands. Computations have been done using the alphahull package in R, see Pateiro-López and Rodríguez-Casal (2010). Note that $C_{0.3}(\mathcal{X}_{740})$ is an acceptable EOO reconstruction equal to the intersection of the complements of all gray open balls represented. However, if we select $r = 5$, marine areas are clearly inside the $C_5(\mathcal{X}_{740})$.

It is well-known that the concept of $r-$convex hull is closely related to the closing of A by $B_r(0)$ from the mathematical morphology, see Serra (1983). It can be shown that

$$C_r(A) = (A \oplus rB) \ominus rB,$$

where $B = B_1(0)$, $\lambda C = \{\lambda c : c \in C\}$, $C \oplus D = \{c + d : \ c \in C, d \in D\}$ and $C \ominus D = \{x \in \mathbb{R}^d : \ \{x\} \oplus D \subset C\}$, for $\lambda \in \mathbb{R}$ and sets C and D.

The problem of reconstructing a $r-$convex support S using a data-driven procedure could be easily solved if the parameter r is selected from the data set. The first step is to determine precisely the optimal value of r to be estimated, which is established in Definition 2: we propose to estimate the largest value of r which verifies that S is $r-$convex.

Definition 2 Let $S \subset \mathbb{R}^d$ a compact, nonconvex and $r-$convex set for some $r > 0$. It is defined

$$r_0 = \sup\{\gamma > 0 : C_\gamma(S) = S\}. \tag{1}$$

Remark 1 For simplicity in the exposition, it is assumed that S is not convex; otherwise, r_0 would be infinity, and the convex hull of the sample provides a good reconstruction.

Remark 2 If the supreme in (1) is a maximum, then S is r convex for $r \leq r_0$. In this case, if $r < r_0$, $C_r(\mathcal{X}_n)$ is a non-admissible estimator since it is always outperformed by $C_{r_0}(\mathcal{X}_n)$. This happens because, with probability one, $C_r(\mathcal{X}_n) \subset C_{r_0}(\mathcal{X}_n) \subset S$.

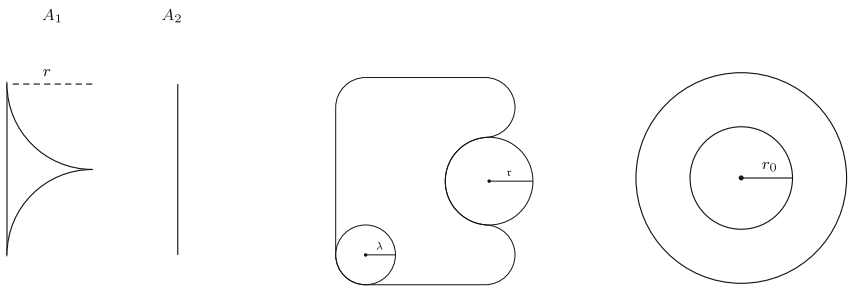

Fig. 3 $A_1 \cup A_2$ fulfills the $r-$rolling condition $\not\Rightarrow$ $A_1 \cup A_2$ is $r-$convex (left). (R) is a more general condition (center). Circular ring with inner circle of radius r_0 (right)

Proposition 2.4 in Rodríguez-Casal and Saavedra-Nieves (2016) ensures that under the shape restriction detailed below, the supreme in (1) is a maximum. The mild regularity condition we need is the following:

(R) S an S^c satisfy the rolling property with rolling positive constants τ and λ, respectively.

Following Cuevas et al. (2012), it is said A satisfies the (outside) $r-$rolling condition if each boundary point $a \in \partial A$ is contained in a closed ball with radius r whose interior does not meet A. There exist interesting relationships between this property and $r-$convexity. In particular, Cuevas et al. (2012) proved that if A is compact and $r-$convex, then A fulfills the $r-$rolling condition. According to Fig. 3 (left), the reciprocal is not always true. Proposition 2.2 in Rodríguez-Casal and Saavedra-Nieves (2016) shows that (R) is a (mild) sufficient condition to ensure the $\tau-$rolling condition implies $\tau-$convexity. Condition (R) was essentially analyzed by Walther (1997, 1999) but just the case $\tau = \lambda$ was taken into account. In this work, the radius λ can be different from τ, see Fig. 3 (center).

2.2 About maximal spacings

The optimal value of the shape index r to be estimated is just established in Definition 2. Some concepts on maximal spacings theory must be handled to propose a consistent estimate of r_0.

The notion of maximal-spacing in several dimensions was introduced and studied by Deheuvels (1983) for uniformly distributed data on the unit cube. Later on, Janson (1987) extended these results to uniformly distributed data on any bounded set and derived the asymptotic distribution of different maximal-spacings notions without conditions on the shape of the support S. Aaron et al. (2017) generalized the results by Janson (1987) to the non-uniform case.

The shape of the considered spacings will be defined by a given set $A \subset \mathbb{R}^d$. For the validity of the theoretical results, it is sufficient to assume that A is a compact and

convex set. For practical purposes, the usual choices are $A = [0, 1]^d$ or $A = B_1[0]$, being $B_r[x]$ the closed ball with center x and radius r. For a general dimension d, the first definition of maximal spacing is that used by Janson (1987) under the restriction of data are uniformly distributed:

$$\Delta_n^*(\mathcal{X}_n) = \sup\{\gamma : \exists x \text{ such that } \{x\} \oplus \gamma A \subset S \backslash \mathcal{X}_n\}.$$

If the Lebesgue measure of the set A is one, $\Delta_n^*(\mathcal{X}_n)^d$ represents the Lebesgue measure of the largest set $\{x\} \oplus \gamma A \subset S \backslash \mathcal{X}_n$. The concept of maximal spacing can be related easily to the maximal inner radius when $A = B_1[0]$. If $\text{Int}(S) \neq \emptyset$, the maximal inner radius of S is defined as

$$\mathcal{R}(S) = \sup\{\gamma > 0 : \exists x \in S \text{ such that } B_\gamma[x] \subset S\}.$$

Note that the value of the maximal spacing depends on S and also on $\mathcal{X}_n$. However, the definition of the maximal inner radius relies only on S.

Aaron et al. (2017) extended the definition of maximal-spacing assuming that $\mathcal{X}_n$ is drawn according to a density f with bounded support S, the Lebesgue measure of the set A is one and its barycenter is the origin of $\mathbb{R}^d$. In this more general setting, the maximal spacing is defined as

$$\Delta_n(\mathcal{X}_n) = \sup\left\{\gamma : \exists x \text{ such that } \{x\} \oplus \frac{\gamma}{f(x)^{1/d}} A \subset S \backslash \mathcal{X}_n\right\}$$

and

$$V_n(\mathcal{X}_n) = \Delta_n(\mathcal{X}_n)^d.$$

The previous definition of maximal spacing relies also on density f. In this way, it distinguishes between low and high density regions. Throughout this paper, $A = w_d^{-1/d} B_1[0]$ where w_d denotes the Lebesgue measure of $B_1[0]$.

Janson (1987) calibrated the volume of the maximal spacing under uniformity assumptions without conditions on the shape of the support S. The corresponding extension established in Theorem 2 in Aaron et al. (2017) is shown in Theorem 1 modifying slightly the original hypotheses on f and on the shape of S. The result remains true if it is assumed that S is under (R) and the density function f satisfies $(f_{0,1}^L)$:

$(f_{0,1}^L)$ The restriction of the density f to S is Lipschitz continuous (there exists k_f such that $\forall x, y \in S$, $|f(x) - f(y)| \leq k_f \|x - y\|$), and there exists $f_0 > 0$ such that $f(x) \geq f_0$ for all $x \in S$. Furthermore, denote $f_1 = \max_{x \in S} f(x)$.

All through this paper, we assume that the random sample of points, $\mathcal{X}_n$, is generated from a density f that satisfies the regularity condition $(f_{0,1}^L)$. Note that it includes the uniform distribution and also, more realistic scenarios with non-uniform sampling allowing to deal with observer biased data.

Theorem 1 (Aaron et al. (2017)) *Let $\mathcal{X}_n$ be a random and i.i.d sample drawn according to a density f that satisfies ($f_{0,1}^L$) with compact and nonempty support S under (R).*

Let U be a random variable with distribution

$$\mathbb{P}(U \leq u) = \exp(-\exp(-u)) \quad \text{for } u \in \mathbb{R}$$

and let β be a constant specified in Janson (1987). Then, we have that

$$U(\mathcal{X}_n) \xrightarrow{d} U \text{ when } n \to \infty,$$
$$\liminf_{n\to\infty} \frac{nV_n(\mathcal{X}_n) - log(n)}{log(log(n))} \geq d-1 \text{ a.s.}, \quad \limsup_{n\to\infty} \frac{nV_n(\mathcal{X}_n) - log(n)}{log(log(n))} \leq d+1 \text{ a.s.}$$

where

$$U(\mathcal{X}_n) = nV_n(\mathcal{X}_n) - log(n) - (d-1)log(log(n)) - log(\beta).$$

Remark 3 The value of constant β does not depend on S. It is explicitly given in Janson (1987). Specifically,

$$\beta = \frac{1}{d!}\left(\frac{\sqrt{\pi}\,\Gamma\left(\frac{d}{2}+1\right)}{\Gamma\left(\frac{d+1}{2}\right)}\right)^{d-1}.$$

In particular, for the bidimensional case, $\beta = 1$.

2.2.1 About nonparametric estimation of maximal spacings

A plug-in estimator of the maximal spacing $\Delta_n(\mathcal{X}_n)$ will be introduced next. Note that the definition of $\Delta_n(\mathcal{X}_n)$ relies on the support S and also on the density function f (both are usually unknown). Under the assumption of $r-$convexity, S will be estimated as $C_r(\mathcal{X}_n)$. As for the density function f, following the ideas in Aaron et al. (2017), a non-conventional density estimator will be introduced in Definition 3.

Definition 3 Let $r > 0$ and let $Vor(X_i)$ be the Voronoi cell of the point X_i (i.e., $Vor(X_i) = \{x : \|x - X_i\| = \min_{y\in\mathcal{X}_n} \|x - y\|\}$). If K is a kernel function (i.e., $K \geq 0$, $\int K = 1$ and $\int uK(u)du = 0$) and $f_n(x) = \frac{1}{nh_n^d}\sum K((x - X_i)/h_n)$ denotes the usual kernel density estimator, we define

$$\hat{f}_n(x) = \max_{i:x\in Vor(X_i)} f_n(X_i).$$

This nonparametric estimator only takes n different values: the evaluation of the usual kernel estimator on the sample points. In fact, for each point $x \in S$, $\hat{f}_n(x)$ is equal to $f_n(X_i)$ where X_i is the closest sample point to x. It can be checked that, with probability one and for n large enough, there exists a point in $\mathcal{X}_n$ as close to x as desired. Therefore, this density estimator is only a simplification of the usual one

with clear computational advantages for estimating Δ_n and V_n. This estimator is just a slight modification of the one proposed in Aaron et al. (2017), avoiding zero values.

Finally, some technical hypotheses on the kernel function must be established. Observe that this condition is satisfied, for instance, by the Gaussian kernel.

($\mathcal{K}_\phi^p$) The kernel function K belongs to the set of kernels $\mathcal{K}$ such that $K(u) = \phi(p(u))$ where p is a polynomial and ϕ is a bounded real function of bounded variation, verifying that $c_K = \int \|u\| K(u)du < \infty$, $K \geq 0$ and there exists r_K and $c_K^{'} > 0$ such that $K(x) \geq c_K^{'}$ for all $x \in B_{r_K}[0]$.

Then, we define the following plug-in estimator of $\Delta_n(\mathcal{X}_n)$

$$\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n) = \sup\left\{\gamma : \exists x \text{ such that } \{x\} \oplus \frac{\gamma}{\hat{f}_n(x)^{1/d}} A \subset C_r(\mathcal{X}_n)\backslash\mathcal{X}_n\right\} \quad (2)$$

and $\hat{V}_{n,r} = \hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n)^d$. Given the definition of $\hat{f}_n$ and the assumption $(f_{0,1}^L)$, it is expected that $\hat{f}_n$ does not go to zero on $C_r(\mathcal{X}_n)$, see Lemma 1. This is important in formulae (2). For instance, if S is $r-$convex, $\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n)$ should converge to zero as the sample size increases. However, if $S \subsetneq C_r(S)$, the plug-in estimator of $\Delta_n(\mathcal{X}_n)$ is expected to converge to a positive constant.

3 A new test for $r-$convexity

We will introduce a consistent hypothesis test based on $\mathcal{X}_n$ drawn according to an unknown density f on the unknown support S, to assess $r-$convexity for a certain $r > 0$. This test is crucial for defining an estimator of r_0 that would allow the data-driven estimation of the support S.

Given $r > 0$, the null hypothesis that S is $r-$convex will be tested taking $\hat{V}_{n,r}$ as statistic. The idea that supports this procedure is simple: Under $(f_{0,1}^L)$ and (R), Theorem 1 allows us to detect which values of $V_n(\mathcal{X}_n)$ are large enough to be incompatible with these two assumptions. A similar reasoning can be also applied if we consider $\hat{V}_{n,r}$, the test is based on the opposite approach: Under $(f_{0,1}^L)$ and (R), if the test statistic takes large enough values, it will mean that the selected r is not appropriate and a smaller value of r should be considered.

Theorem 2 *Let $r > 0$ and let $\mathcal{X}_n$ be a random and i.i.d sample drawn according to a density f that satisfies $(f_{0,1}^L)$ with compact and nonempty support S under (R). Let f_n be the modified density estimator introduced in Definition 3 whose kernel function is supposed to satisfy condition $(\mathcal{K}_\phi^p)$ and the sequence h_n of smoothing parameters fulfills $h_n = O(n^{-\zeta})$ for some $0 < \zeta < 1/d$. Let $\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n)$ be the maximal spacing estimator established in equation (2). Given the statistical testing problem,*

$$H_0 : S \text{ is } r - \text{convex versus } H_1 : S \text{ is not } r - \text{convex}.$$

(a) The test based on the statistic $\hat{V}_{n,r} = \hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n)^d$ *with critical region* $RC = \{\hat{V}_{n,r} > c_{n,\alpha}\}$, *where*

$$c_{n,\alpha} = \frac{1}{n}(-log(-log(1-\alpha)) + log(n) + (d-1)log(log(n)) + log(\beta))$$

has an asymptotic level not larger than α.
(b) Moreover, if S is not r−convex, it is verified that $\mathbb{P}(\hat{V}_{n,r} > c_{n,\alpha},\ eventually) = 1$.

Remark 4 Note that the optimal kernel sequence size for estimating f, $h_n = h_0 n^{-1/(d+4)}$, satisfies the hypotheses under which Theorem 2 holds. Therefore, any reasonable bandwidth selector should be suitable for testing $r-$convexity.

Remark 5 Under (R) r_0 is the maximum of the set $\{\gamma > 0 : C_\gamma(S) = S\}$. Hence, the hypotheses of the test introduced in Theorem 2 can be rewritten as follows:

$$H_0 : r \leq r_0 \text{ versus } H_1 : r > r_0.$$

Observe that, under H_1, $S = C_{r_0}(S) \subsetneq C_r(S)$.

The performance of this test can be illustrated using the real database of invasive plants in Azorean islands. Given the sample $\mathcal{X}_{740}$, the practitioner could be interested in testing the null hypothesis that the EOO is $r-$convex, for instance, for $r = 5$. According to Fig. 5 (third row, right), it is clear that large Atlantic Ocean areas are inside $C_5(\mathcal{X}_{740})$ and the EOO is overestimated. Moreover, $\hat{V}_{740,5}$ will be too large. In fact, although larger samples sizes were considered, its volume would take a constant value (see gray ball inside the EOO reconstruction). Therefore, the null hypothesis of $5-$convexity should be rejected. Note that the situation is the opposite if testing $r-$convexity for $r = 0.3$ is the goal. In this case, $\hat{V}_{740,0.3}$ should be clearly smaller. Furthermore, when the sample size increases, this volume tends to zero.

3.1 Selection and consistency results of the optimal smoothing parameter

The optimal estimation of the smoothing parameter r_0 from $\mathcal{X}_n$ is based on the test previously proposed. Specifically, according to Definition 2, r_0 will be estimated by

$$\hat{r}_0 = \sup\{r > 0 :\ \text{The null hypothesis } H_0 \text{ that } S \text{ is } r-\text{convex is accepted}\}. \quad (3)$$

That is, it is proposed to select the largest value of r compatible with the $r-$convexity assumption. Note that this choice depends on the significance level of the test, but its dependence is not explicitly given in the notation for the sake of clarity. The theoretical properties for the estimator of r_0 are considered next. First, the existence of the supreme defined in (3) must be guaranteed, a result which is proved in Theorem 3. In addition, it is also proved that $\hat{r}_0$ consistently estimates r_0.

Theorem 3 *Let* f *be a density function that satisfies* $(f^L_{0,1})$ *with compact, nonconvex and nonempty support S under* (R). *Let* $\hat{f}_n$ *be the density estimator introduced in*

Definition 3 whose kernel function is supposed to satisfy condition $(\mathcal{K}_\phi^p)$ *and the sequence* h_n *of smoothing parameters fulfills* $h_n = O(n^{-\zeta})$ *for some* $0 < \zeta < 1/d$. *Let* r_0 *be the parameter defined in (1) and* $\hat{r}_0$ *defined in (3). Let* $\{\alpha_n\} \subset (0, 1)$ *be a sequence of significance levels converging to zero such that* $\log(\alpha_n)/n \to 0$. *Then,* $\hat{r}_0$ *converges to* r_0 *in probability.*

Remark 6 For the sake of clarity, S is assumed non-convex throughout the paper. However, if S is convex, it can be shown that $\hat{r}_0$ goes to infinity (which is the value of r_0 in this case) because, with high probability, the test is not rejected for all values of r.

We use again the example of invasive plants in Azorean islands in order to illustrate the behavior of this estimator. Under $(f_{0,1}^L)$ and (R), if $\hat{V}_{n,r}$ is large enough, then the null hypothesis of $r-$convexity will be rejected. Therefore, a smaller value of r should be selected. This case corresponds to Fig. 5 (third row, right) taking $r = 5$. Observe that the null hypothesis of $r'-$convexity would be also rejected for all $r' \geq r$ because $C_r(\mathcal{X}_n) \subset C_{r'}(\mathcal{X}_n)$ and, consequently, $\hat{V}_{n,r'} \geq \hat{V}_{n,r}$. However, the situation is completely opposite in Fig. 5 (second row, right) when $r = 0.03$. Here, the size of the maximal spacing found in $C_{0.03}(\mathcal{X}_{740})\backslash\mathcal{X}_{740}$ does not allow to reject that the support is $0.03-$convex. As a consequence, a bigger r than 0.03 should be considered.

4 Consistency and convergence rates of resulting support estimator

The behavior of the random set $C_{\hat{r}_0}(\mathcal{X}_n)$ as an estimator of S can be studied once the consistency of $\hat{r}_0$ has been proved. Two metrics between sets are usually considered in order to assess the performance of a support estimator. Specifically, let A and C be two closed, bounded, nonempty subsets of $\mathbb{R}^d$. The Hausdorff distance between A and C is defined by

$$d_H(A, C) = \max\left\{\sup_{a\in A} d(a, C), \sup_{c\in C} d(c, A)\right\},$$

where $d(a, C) = \inf\{\|a - c\| : c \in C\}$. Besides, if A and C are two bounded and Borel sets, then the distance in measure between A and C is defined by $d_\mu(A, C) = \mu(A\triangle C)$, where μ denotes the Lebesgue measure and $\triangle$, the symmetric difference, that is, $A\triangle C = (A\backslash C)\cup(C\backslash A)$. Hausdorff distance quantifies the physical proximity between two sets, whereas the distance in measure is useful to quantify their similarity in content. However, neither of these distances are completely useful for measuring the similarity between the shape of two sets. The Hausdorff distance between boundaries, $d_H(\partial A, \partial C)$, can be also used to evaluate the performance of the estimators (see Baíllo and Cuevas 2001; Cuevas and Rodríguez-Casal 2004; Rodríguez-Casal 2007 or Genovese et al. 2012).

In particular, if $\lim_{r\to r_0^+} d_H(S, C_r(S)) = 0$, then the consistency of $C_{\hat{r}_0}(\mathcal{X}_n)$ can be proved easily from Theorem 3. However, the consistency cannot be guaranteed if $d_H(S, C_r(S))$ does not go to zero as r goes to r_0 from above (as $\hat{r}_0$ does, see Proposition

1 below). This problem can be solved by considering the estimator $C_{r_n}(\mathcal{X}_n)$ where $r_n = \nu\hat{r}_0$ with $\nu \in (0,1)$ fixed. This ensures that, for n large enough, with high probability, $C_{r_n}(\mathcal{X}_n) \subset S$. From the practical point of view, the selection of ν is not a major issue because $\hat{r}_0$ is numerically approximated and the computed estimator always satisfies this property without multiplying by ν.

Theorem 4 *Let $\mathcal{X}_n$ be a random and i.i.d sample drawn according to a density f that satisfies $(f^L_{0,1})$ with compact, nonconvex and nonempty support S under (R). Let r_0 be the parameter defined in (1) and $\hat{r}_0$ defined in (3). Let $\{\alpha_n\} \subset (0,1)$ be a sequence converging to zero such that $\log(\alpha_n)/n \to 0$. Let be $\nu \in (0,1)$ and $r_n = \nu\hat{r}_0$. Then, eventually almost sure,*

$$d_H(S, C_{r_n}(\mathcal{X}_n)) \leq D\left(\frac{\log n}{n}\right)^{\frac{2}{d+1}}$$

for some positive constant D.

The same convergence rate holds for $d_H(\partial S, \partial C_{r_n}(\mathcal{X}_n))$ and $d_\mu(S \triangle C_{r_n}(\mathcal{X}_n))$.

5 Numerical implementation

The main numerical aspects of the estimation algorithm of r_0 in (1) are detailed in what follows. Although the method proposed in this work is fully data-driven, its practical implementation depends on the specification of the significance level of the test α. Choosing this value is clearly a much simpler problem than the specification of the shape index r_0.

From Theorem 3, with probability one, for a large enough n, the existence of the estimator $\hat{r}_0$ defined in (3) is guaranteed. However, in practice, this estimator might not exist for a specific sample $\mathcal{X}_n$ and a given value of the significance level α. Therefore, the influence of α must be taken into account from the practical point of view. The null hypothesis of $r-$convexity will be (incorrectly) rejected for $0 < r \leq r_0$ with probability at most α. This is not important from the theoretical point of view. Since we are assuming that $\alpha = \alpha_n$ goes to zero as the sample size increases this has not theoretical relevance. But, what should be done, for a given sample, if H_0 is rejected for *all* r (or at least *all* reasonable values of r)? In order to fix a minimum acceptable value of r, it is assumed that S (and, hence, its estimator) will have no connected components with probability content lesser than a value $\mathbf{p} \in (0,1)$. From an empirical approach, the connected components of the estimator will contain at least a proportion $\mathbf{p}$ of sample points. If $\mathbf{p}$ is sufficiently close to zero, too fragmented estimators (for instance, with isolated points or insignificant clusters) will not be considered even in the case that we reject H_0 for all r. In this latter case, the minimum value that ensures that all connected components of the estimator contain at least a proportion $\mathbf{p}$ of sample points will be taken. Therefore, this parameter $\mathbf{p}$ can be interpreted as a geometric stopping criterion that does not appear in theoretical results because the sequence α_n is assumed to tend to zero. Note that this shape assumption is very flexible. It does not limit too much the number of connected components. In fact, the estimator

could present, as maximum number of clusters, the largest integer less than or equal to $1/\mathbf{p}$. In particular, if $\mathbf{p}$ is close enough to zero, the number of connected components can reach very high values. In fact, it is expected that, when the sample size increases, $\mathbf{p} = \mathbf{p}_n$ tending to zero. An alternative procedure, very similar and computationally simpler, is to establish a maximum number of connected components instead of $\mathbf{p}$.

Although the definition of $\hat{r}_0$ depends on a test that must be applied several times in practice, remark that multiple testing does not play any role. It is possible to write $\hat{r}_0 = \sup\{r > 0 : \hat{V}_{n,r} \leq c_{n,\alpha}\}$. Since $\hat{V}_{n,r} \leq \hat{V}_{n,r'}$, for $r \leq r'$, dichotomy algorithms can be used to compute $\hat{r}_0$. The practitioner must select a maximum number of iterations I and two initial points r_m and r_M with $r_m < r_M$ such that the null hypothesis of r_M-convexity is rejected and the null hypothesis of r_m-convexity is accepted. According to the previous comments, it is assumed that the proportion of sample points in each connected component of $C_{r_m}(\mathcal{X}_n)$ must be at least $\mathbf{p}$. Choosing a value close enough to zero is usually sufficient to select r_m. According to Fig. 5 (second row, right), the maximal spacing in $C_{0.03}(\mathcal{X}_n)$ will be small enough to accept $0.03-$convexity. Therefore, taking $r_m \leq 0.03$ will be a good choice. However, if selecting this r_m is not possible because, for very low values of r, the hypothesis of $r-$convexity is still rejected, then r_0 is estimated as the positive closest value to zero r such that all connected components of $C_r(\mathcal{X}_n)$ contain at least a proportion $\mathbf{p}$ of sample points. On the other hand, if a large enough spacing for having a statistically significant test cannot be found in $H(\mathcal{X}_n)$, then we propose $H(\mathcal{X}_n)$ as the estimator for the support.

To sum up, the following inputs should be given: the significance level $\alpha \in (0, 1)$, a maximum number of iterations I, a proportion $\mathbf{p}$ and two initial values r_m and r_M. Given these parameters $\hat{r}_0$ will be computed as follows:

1. In each iteration and while the number of them is smaller than I:
   (a) $r = (r_m + r_M)/2$.
   (b) If the null hypothesis of $r-$convexity is not rejected, then $r_m = r$.
   (c) Otherwise, $r_M = r$.
2. Then, $\hat{r}_0 = r_m$.

Some technical aspects related to the computation of the maximal spacings must be also mentioned. In the proposed procedure, the null hypothesis needs to be tested I times. Since it involves the calculation of the maximal spacing, one may be aware of computational cost of the method. Nevertheless, as noted by Rodríguez-Casal and Saavedra-Nieves (2016), this maximal spacing does not need to be specifically determined and it is enough to check if there exists a point x such that

$$\{x\} \oplus \frac{c_{n,\alpha}^{1/d}}{\hat{f}_n^{1/d}(x)} A \subset C_r(\mathcal{X}_n) \backslash \mathcal{X}_n.$$

In this case, $\hat{V}_{n,r} \geq c_{n,\alpha}$ and, therefore, the null hypothesis of $r-$convexity will be rejected. Furthermore, note that if this disc exists, then $x \notin B_{c_{n,\alpha}^{x,w}}(X_k)$ where $c_{n,\alpha}^{x,w} = c_{n,\alpha}^{1/d} w_d^{-1/d} \hat{f}_n^{-1/d}(x)$ and X_k denotes the sample point such that $x \in Vor(X_k)$. Therefore, $\hat{f}_n(x) = f_n(X_k)$.

Then, the centers of the possible maximal balls that belong to the Voronoi tile with nucleus X_i $(1, \ldots, n)$ necessarily lie in $B_{c_{n,\alpha}^{X_i,w}}(X_i)^c \cap Vor(X_i)$. We will follow the next steps:

1. Determine the set of candidates for ball centers

$$D(r) = C_r(\mathcal{X}_n) \cap \bigcup_{X_i \in E(m)} (\partial B_{c_{n,\alpha}^{X_i,w}}(X_i) \cap Vor(X_i))$$

where $E(m) \subset \mathcal{X}_n$ denotes the extremes of the $m-$shape of $\mathcal{X}_n$ when $m = \min\left\{c_{n,\alpha}^{X_j,w} : X_j \in \mathcal{X}_n\right\}$, see Edelsbrunner (2014). If $x \in D(r)$, then we can guarantee that $B_{c_{n,\alpha}^{X_i,w}}(x) \cap \mathcal{X}_n = \emptyset$. Equivalently,

$$\{x\} \oplus \frac{c_{n,\alpha}^{1/d}}{\hat{f}_n^{1/d}(x)} A \subset C_r(\mathcal{X}_n) \backslash \mathcal{X}_n.$$

2. Calculate $M(r) = \max\{d(x, \partial C_r(\mathcal{X}_n) : x \in D(r)\}$.
3. If $M(r) \leq \hat{c}_{n,\alpha}$, then the null hypothesis of $r-$convexity is not rejected.

It should be noted that if, for all $X_i \in E(m)$, $x \notin \partial B_{c_{n,\alpha}^{X_i,w}}(X_i) \cap Vor(X_i)$ then $B_{c_{n,\alpha}^{X_i,w}}(x) \cap \mathcal{X}_n \neq \emptyset$. Therefore, these points can be discarded in order to determine $D(r)$. Furthermore, $E(m)$, $\partial C_r(\mathcal{X}_n)$ and $\partial B_{\hat{c}_{n,\alpha,r}^*}(\mathcal{X}_n)$ can be easily computed (at least for the bidimensional case). See Pateiro-López and Rodríguez-Casal (2010) for further details.

6 Simulation results

The behavior of the test of $r-$convexity is established in Sect. 3, and the performance of the estimator of r_0 described in Sect. 5 will be analyzed through a simulation study.

Three different supports S are considered in this section: $B_1[(0, 0)]\backslash B_{0.25}(0, 0)$, $B_2[(0, 0)]\backslash B_{0.25}(0, 0)$ and $B_{0.4}[(-0.5, 0.5)] \cup B_{0.4}[(0.5, -0.5)]$. All of them are $r-$convex for several $r > 0$. According to Fig. 4, r_0 is equal to 0.25 in the first two models and it takes the value 0.307 in the third model. Random samples have been generated by acceptance-rejection method on the three described supports. A standard multivariate normal distribution was considered for replicates generation in Models 1 and 2 and a mixture of two normal densities, for Model 3. The vector of weights of each normal component in the mixture is $(1/2, 1/2)$. The vectors of means are $(-0.5, 0.5)$ and $(0.5, -0.5)$, respectively, and the covariance matrices are $\Sigma = (\sigma_{ij})_{2\times 2}$, verifying that $\sigma_{ij} = 1/9$ if $i = j$ and $\sigma_{ij} = 1/15$ if $i \neq j$.

Regarding the performance of the $r-$convexity test, a total of 250 replicates of sample sizes 100, 500, 1000, and 2000 were generated for each of the simulation models. Tables 1, 2, and 3 contain the mean number of rejections when $\alpha = 0.1$ for Models 1, 2, and 3, respectively. It should be noted that different values of the parameter r have been considered. The exact value of r_0 was represented using bold

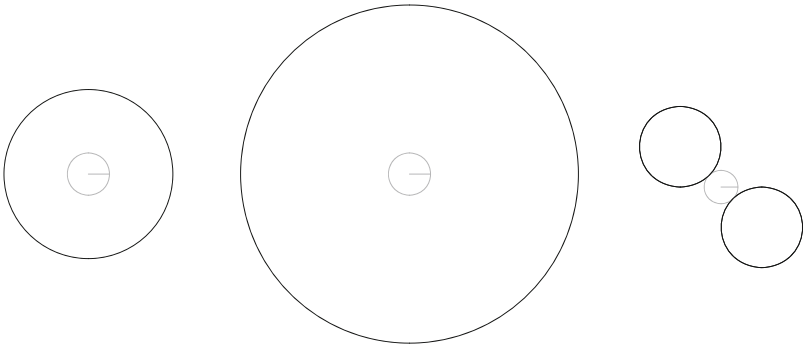

Fig. 4 Model 1: $S = B_1[(0,0)] \backslash B_{0.25}[(0,0)]$ where $r_0 = 0.25$ (left). Model 2: $S = B_2[(0,0)] \backslash B_{0.25}[(0,0)]$ where $r_0 = 0.25$ (center). Model 3: $S = B_{0.4}[(-0.5, 0.5)] \cup B_{0.4}[(0.5, -0.5)]$ where $r_0 = 0.307$ (left). The radius of the balls represented in gray color corresponds to the value of the shape index r_0

Table 1 Mean number of rejections for the $r-$convexity test over 250 replicates of Model 1 for different values of r and n when $\alpha = 0.1$. The value of r_0 is equal to 0.25

n/r	0.1	**0.25**	0.5	1
100	0.000	0.000	0.612	0.612
500	0.000	0.016	1.000	1.000
1000	0.000	0.012	1.000	1.000
2000	0.040	0.040	1.000	1.000

Table 2 Mean number of rejections for the $r-$convexity test over 250 replicates of Model 2 for different values of r and n when $\alpha = 0.1$.The value of r_0 is equal to 0.25

n/r	0.1	**0.25**	0.5	1
100	0.000	0.000	0.052	0.140
500	0.000	0.012	1.000	1.000
1000	0.000	0.092	1.000	1.000
2000	0.000	0.128	1.000	1.000

numbers for Models 1 and 2. For Model 3, the closest value to r_0 was also using bold numbers. Simulation results indicate that the test of r-convexity is well calibrated, as it is established in Theorem 2. Under the null hypothesis, the mean number of rejections tends, as the sample size increases, to the nominal level α. However, the $r-$convexity test exhibits better consistency behavior for Models 2 and 3 than for Model 1 although the shape of the supports in scenarios 1 and 2 is quite similar. The result in Aaron et al. (2017) on which test rests deals with convergence in distribution of extrema. Following Hall (1991), this convergence in distribution can be extremely slow. Therefore, bigger sample sizes should be considered for Model 1. Bootstrap calibration could be also investigated.

Regarding the estimation of r_0, the behavior of the algorithm proposed in Sect. 5 was analyzed when $\alpha = 0.01$, for the 250 replicates of the three models previously introduced. A maximum number of four connected components were allowed. Table 4 contains the empirical means (M) and the standard deviations (SD) of these estimations

Table 3 Mean number of rejections for the $r-$convexity test over 250 replicates of Model 3 for different values of r and n when $\alpha = 0.1$. The value of r_0 is equal to 0.307

n/r	0.1	**0.3**	0.5	0.75
100	0.000	0.000	0.000	0.020
500	0.040	0.048	0.128	0.98
1000	0.040	0.060	0.372	1.00
2000	0.088	0.092	0.932	1.00

Table 4 Means of 250 estimations for the parameter r_0 when $\alpha = 0.01$

n	100		500		1000		2000	
	M	SD	M	SD	M	SD	M	SD
Model 1	0.459	0.625	0.257	0.005	0.253	0.002	0.249	0.015
Model 2	2.700	0.773	0.292	0.249	0.258	0.006	0.253	0.008
Model 3	1.597	0.476	0.660	0.089	0.545	0.065	0.473	0.073

for Models 1, 2, and 3, respectively. A total of 214 estimations of r_0 were equal to ∞ for Model 1 when $n = 100$. This situation is repeated for Model 2 when $n = 100$ (for 223 samples) and $n = 500$ (11 replicates). These estimations have not been taken into account for computing the averages shown in Table 4. Finally, it is worth to mention that the asymptotic calibration of the test does not preclude observing the consistency of the estimator for the parameter r_0 whose estimation is the main goal of this work.

7 Extent of occurrence estimation

As an illustrative example, the new support estimator introduced in this work will be used for reconstructing the EOO of an assemblage of terrestrial invasive plants in two islands of the Azores Archipelago, Terceira and São Miguel. For this real dataset, we have shown that convexity assumption is very restrictive. According to Fig. 5 (first and second rows, left), sea areas are inside the classical estimator of the EOO. Obviously, it is overestimated given that terrestrial invasive plants do not occupy the Atlantic Ocean. The goal here is to reconstruct the EOO in this application overcoming the described limitation.

First, it is necessary to estimate the optimal value r_0 from the sample of 740 geographical locations. If we select the significance level α equal to 0.01 and $\mathbf{p} = 0.05$, the resulting estimator is $\hat{r}_0 = 0.057$. In Fig. 6, $C_{\hat{r}_0}(\mathcal{X}_{740})$ is shown. According to the results obtained, the EOO reconstruction has two different connected components corresponding to the two Azorean islands. In this case, $\hat{r}_0$ corresponds to the minimum value that guarantees that all connected components contain at least 37 sample points to avoid insignificant clusters of invasive plants. For smaller values of $\mathbf{p}$ (for instance, 0.02 that corresponds to a minimum number of geographical locations equal to 15 in each cluster), the number of connected components does not change. Unlike classical EOO estimator, sea areas are not inside the reconstruction. Therefore, a more realistic estimator of the EOO can be determined. It should be noted that the dataset may not

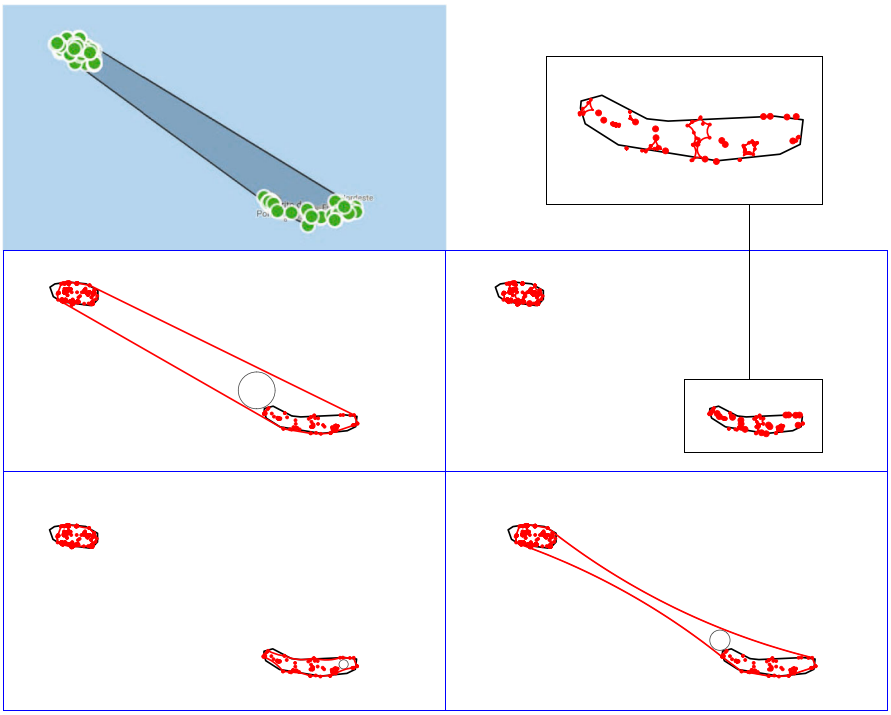

Fig. 5 In the first row, GeoCAT reconstruction of the EOO determined from the sample of 740 geographical locations in two Azorean islands (left). In the second row (red color), $H(\mathcal{X}_{740})$ (left) and $C_{0.03}(\mathcal{X}_{740})$ (right). In the third row (red color), $C_{0.3}(\mathcal{X}_{740})$ (left) and $C_5(\mathcal{X}_{740})$ (right)

be independent. Several sample points are collected at the same day, and they are very close to each other. This does not happen all days, but some data are clearly clustered. This may cause that $\hat{r}_0$ underestimate r_0. The condition on the size of the connected components prevents the estimator to take too small values in these situations, at it may happen in this illustrative example.

The new method, although designed for handling more complex situations, provides similar reconstructions to those corresponding to the convex hull in those cases where the classical reconstruction works appropriately. For showing this, we will focus on the geographical locations from São Miguel island. Separately, the EOO will be estimated from data corresponding to years 2015 and 2016. A total of 33 and 48 geographical locations are available in 2015 and 2016, respectively.

Figure 7 contains the EOO estimator in 2015 (left) and 2016 (center). In 2015, the resulting reconstruction of the EOO is equal to $H(\mathcal{X}_{33})$. In 2016, $\hat{r}_0 = 1.404$; however, the estimation of the EOO obtained, $C_{1.404}(\mathcal{X}_{48})$, is not so different from the convex hull. This last illustration suggests that, if more amount of data are available by year, this kind of analysis could be useful for studying the temporal changes in the spatial pattern of organisms, including invasive plants, on an area of interest.

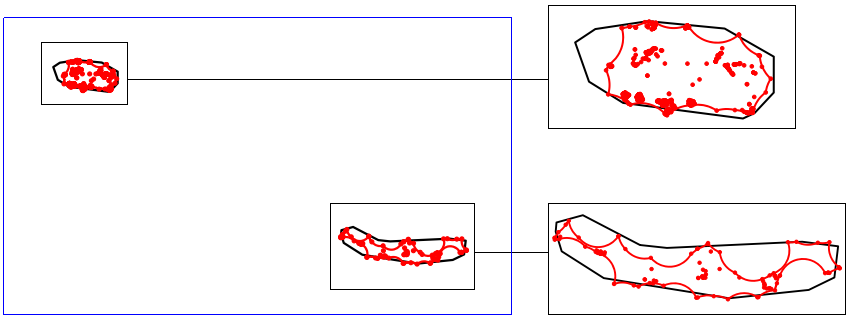

Fig. 6 EOO estimator, $C_{\hat{r}_0}(\mathcal{X}_{740})$ where $\hat{r}_0 = 0.057$

Fig. 7 EOO estimator in 2015, $H(\mathcal{X}_{33})$ (left); EOO estimator in 2016, $C_{\hat{r}_0}(\mathcal{X}_{48})$ where $\hat{r}_0 = 1.404$ (center); EOO estimators in 2015 (blue) and 2016 (gray) (right)

8 Conclusions and open problems

The main goal of this work is to propose a new data-driven method for reconstructing a $r-$convex support in a consistent way. The route designed to reach this goal can be summarized as follows: (1) Defining the optimal value of r, r_0, to be estimated, (2) establishing a nonparametric test to assess the null hypothesis that S is $r-$convex for a given $r > 0$, (3) defining the estimator of r_0 that strongly relies on the previous test (4) checking that the estimator of r_0 and the resulting support reconstruction are consistent and (5) studying the performance of the $r-$convexity test and the estimation algorithm of r_0 through simulations.

The definition of the estimator $\hat{r}_0$ depends on the $r-$convexity test established that, of course, could be used in an independent way. In many practical situations where the support is completely unknown and only a sample of points is available, it can be interesting to test if the corresponding support distribution is $r-$convex.

Furthermore, the behavior of the proposed support estimator was illustrated through the estimation of the EOO of an assemblage of terrestrial invasive plants in two Azorean islands, providing a novel tool for ERA. In this particular case, where convexity assumption on the EOO is too restrictive, our support estimator provides a more realistic and sophisticated reconstruction. Besides, we have also shown that when the classical convex reconstruction works appropriately, our estimator offers similar reconstructions. Furthermore, we have shown that estimating the EOO from annual (or any other time period) occurrences could be useful for detecting temporal changes in the spatial pattern of organisms.

Note that the resulting support estimator is spatially flexible. In other words, it is able to distinguish the different disconnected components of the support. Therefore, it

could be used for estimating the number of support connected components. Another relevant application deals with the integrated nested Laplace approximation (INLA) methodology introduced in Rue et al. (2009) and extended by Lindgren et al. (2011), establishing the stochastic partial differential equation (SPDE) approach. As noted by these authors, the application of a SPDE model requires the determination of a physical domain for the process where a discretization mesh is constructed. Different domains yield to different results, and our proposal can be viewed as a data-driven alternative to obtain a reasonable domain of interest.

Finally, another interesting problem and intimately related to the EOO reconstruction is to estimate the area of occupancy (AOO). The IUCN defined the AOO as the area within its extent of occurrence. Under $r-$convexity, we could estimate the AOO as the area of the $r-$convex hull of the sample points. However, this estimator suffers from the drawback of not being rate-optimal. Arias-Castro et al. (2019) proposed, under uniform distribution assumptions, an optimal volume estimator based on the sample $r-$convex hull using a sample splitting strategy that attains the minimax lower bound. Therefore, the problem of estimating the AOO could be studied from a different perspective in future.

9 Proofs

In this section, the proofs of the stated theorems are presented.

Proof of Theorem 1 Theorem 2 in Aaron et al. (2017) considers that f is Hölder continuous with respect to Lebesgue measure. Under $(f^L_{0,1})$, this condition is also satisfied.

Furthermore, Aaron et al. (2017) also assumed that there exists $k < d$ and $C_{\partial S} > 0$ such that $N(\partial S, \epsilon) \leq C_{\partial S}\epsilon^{-k}$ where $N(\partial S, \epsilon)$ denotes the inner covering number of ∂S. Under (R), Theorem 1 in Walther (1997) guaranteed that ∂S is a $\mathcal{C}^1$ $(d-1)-$dimensional submanifold. In this case, it is well known that the previous assumption is fulfilled for $k = d-1$. More details can be found in Aaron et al. (2017). □

Proof of Theorem 2 First, we will prove (a) and then, (b).

(a) Under H_0 $(C_r(S) = S)$, with probability one, $C_r(\mathcal{X}_n) \subset S$. Then,

$$\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n) \leq \sup\left\{\gamma : \exists x \in C_r(\mathcal{X}_n) \text{ such that } \{x\} \oplus \frac{\gamma}{\hat{f}_n(x)^{1/d}} A \subset S\backslash\mathcal{X}_n\right\}.$$

Let $G_n = \left\{\inf_{x\in C_r(\mathcal{X}_n)} \left(\frac{f(x)}{\hat{f}_n(x)}\right)^{1/d} \geq 1-\epsilon_n^+\right\}$ for some ϵ_n^+, see Lemma 1. If G_n holds, we can ensure that

$$\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n) \leq \sup\left\{\gamma : \exists x \in C_r(\mathcal{X}_n) \text{ such that } \{x\} \oplus \frac{(1-\epsilon_n^+)\gamma}{f(x)^{1/d}} A \subset S\backslash\mathcal{X}_n\right\}.$$

Therefore, under G_n, it is verified that $\Delta_n(\mathcal{X}_n) \geq (1-\epsilon_n^+)\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n)$. Consequently, $V_n(\mathcal{X}_n) \geq (1-\epsilon_n^+)^d \hat{V}_{n,r}$. Hence, if $\hat{V}_{n,r} > c_{n,\alpha}$, it is satisfied that

$V_n(\mathcal{X}_n)(1-\epsilon_n^+)^{1/d} \geq V_{n,r} > c_{n,\alpha}$. So, $V_n(\mathcal{X}_n) > c_{n,\alpha}(1-\epsilon_n^+)^d$. Then, we can write

$$\left\{\hat{V}_{n,r} > c_{n,\alpha}\right\} \cap G_n \subset \left\{V_n(\mathcal{X}_n) > c_{n,\alpha}(1-\epsilon_n^+)^d\right\} \cap G_n.$$

Therefore,

$$\mathbb{P}(\hat{V}_{n,r} > c_{n,\alpha}) \leq \mathbb{P}\left(V_n(\mathcal{X}_n) > c_{n,\alpha}(1-\epsilon_n^+)^d\right) + \mathbb{P}(G_n^c). \tag{4}$$

Next lemma, from Lemma 5 in Aaron et al. (2017), guarantees that $\mathbb{P}(G_n^c, \text{ i. o.}) = 0$. □

Lemma 1 (Aaron et al. (2017)) *Let $r > 0$ and let f be a density function that satisfies $(f_{0,1}^L)$ with compact and nonempty support S under (R). Let $\hat{f}_n$ be the corresponding density estimator introduced in Definition 2.6 whose kernel function is supposed to satisfy condition $(\mathcal{K}_\phi^p)$ and the sequence h_n of smoothing parameters fulfills $h_n = O(n^{-\zeta})$ for some $0 < \zeta < 1/d$. Then,*

(i) there exists a positive constant $\lambda_1 > 0$, which do not depend on r, such that for all $x \in C_r(\mathcal{X}_n)$, $(\hat{f}_n(x))^{1/d} \geq \lambda_1$, e.a.s.
(ii) there exists a sequence ϵ_n^+ such that, $\log(n)\epsilon_n^+$ tends to zero, for all $r \leq r_0$ and for $x \in C_r(\mathcal{X}_n)$,

$$\left(\frac{f(x)}{\hat{f}_n(x)}\right)^{1/d} \geq 1-\epsilon_n^+, \text{ e.a.s.}$$

Proof The proof is an straightforward consequence of Lemma 5 in Aaron et al. (2017). Under (R), if f is bounded from below on S, it is easy to show that set S is standard. Regarding conclusion (i) notice that for all r, $C_r(\mathcal{X}_n) \subset H(\mathcal{X}_n)$. □

Then, the second term in (4) is negligible and $\mathbb{P}(\hat{V}_{n,r} > c_{n,\alpha})$ can be bounded by

$$\begin{aligned}\mathbb{P}(U(\mathcal{X}_n) > &-(1-\epsilon_n^+)^d \log(-\log(1-\alpha)) + ((1-\epsilon_n^+)^d - 1)(\log(n) \\ &+(d-1)\log(\log(n)) + \log(\beta))).\end{aligned}$$

According to Theorem 1, $U(\mathcal{X}_n) \xrightarrow{d} U$ when $n \to \infty$. Furthermore, notice that U has a continuous distribution, so convergence in distribution implies that

$$\sup_u |\mathbb{P}(U(\mathcal{X}_n) \leq u) - \mathbb{P}(U \leq u)| \to 0.$$

Therefore, using that $\log(n)\epsilon_n^+$ tends to zero, we get that

$$\mathbb{P}(U(\mathcal{X}_n) > -\log(-\log(1-\alpha)) + o(1)) \to \alpha.$$

As a consequence,

$$\mathbb{P}(\hat{V}_{n,r} > c_{n,\alpha}) \leq \mathbb{P}(U(\mathcal{X}_n) > -\log(-\log(1-\alpha)) + o(1)) \to \alpha.$$

(b) An auxiliary result must be taken into account for completing the proof of (b).

Lemma 2 *Let $\mathcal{X}_n$ be a random and i.i.d sample drawn according to a density f that satisfies ($f^L_{0,1}$) with compact, nonconvex and nonempty support S under (R). Let r_0 be the parameter defined in (1). Then, for all $r > r_0$, there exists an open ball $B_\rho(x)$ such that $B_\rho(x) \cap S = \emptyset$ and*

$$\mathbb{P}\left(B_\rho(x) \subset C_r(\mathcal{X}_n),\ \textit{eventually}\right) = 1.$$

Proof Proof of Lemma 8.4 in Rodríguez-Casal and Saavedra-Nieves (2016) remains true if sample distribution is not uniform. Therefore, in this more general setting, it allows to guarantee that there exists a closed ball of radius $\rho > 0$ that, with probability one and for n large enough, is inside $C_r(\mathcal{X}_n)\backslash S$. □

From Lemma 1 (i),

$$\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n) \geq \lambda_1 \mathcal{R}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n),\ \text{e.a.s}$$

where λ_1 is a positive constant. Under H_1 (S is not $r-$convex, $S \subsetneq C_r(S)$), we will prove that,

$$\mathcal{R}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n) \geq \rho > 0,\ \text{e.a.s.}$$

Lemma 2 allows to guarantee that there exists a closed ball of radius $\rho > 0$ that, with probability one and for n large enough, is inside $C_r(\mathcal{X}_n)\backslash\mathcal{X}_n$. Consequently,

$$\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n) \geq \lambda_1 \mathcal{R}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n) \geq \lambda_1 \rho w_d^{1/d}.$$

Then,

$$\hat{\delta}(C_r(\mathcal{X}_n)\backslash\mathcal{X}_n) > \frac{\rho}{2}\lambda_1 w_d^{1/d},\ \text{e.a.s.}$$

The proof is finished taking into account that $c_{n,\alpha}$ tends to zero.

Proof of Theorem 3 Some auxiliary results are necessary. First we will prove that, with probability tending to one, $\hat{r}_0$ is at least as big as r_0.

Proposition 1 *Let f be a density function that fulfills condition ($f^L_{0,1}$) with compact, nonconvex and nonempty support S under (R). Let $\hat{f}_n$ be the corresponding density estimator introduced in Definition 3 whose kernel function is supposed to satisfy condition ($\mathcal{K}^p_\phi$) and the sequence h_n of smoothing parameters fulfills $h_n = O(n^{-\zeta})$ for*

some $0 < \zeta < 1/d$. *Let* r_0 *be the parameter defined in (1) and* $\hat{r}_0$ *defined in (3). Let* $\{\alpha_n\} \subset (0, 1)$ *be a sequence converging to zero. Then,*

$$\lim_{n\to\infty} \mathbb{P}(\hat{r}_0 \geq r_0) = 1.$$

Proof Equivalently, we will prove that

$$\lim_{n\to\infty} \mathbb{P}(\hat{r}_0 < r_0) = 0.$$

From the definition of $\hat{r}_0$, see (3), it is clear that

$$\mathbb{P}(\hat{r}_0 < r_0) \leq \mathbb{P}(\hat{V}_{n,r_0} > c_{n,\alpha_n})$$

where remember $\hat{V}_{n,r_0} = \hat{\delta}(C_{r_0}(\mathcal{X}_n)\backslash\mathcal{X}_n)^d$ and $c_{n,\alpha_n} = n^{-1}(-\log(-\log(1-\alpha_n)) + \log(n) + (d-1)\log\log(n) + \log\beta)$.
Since, with probability one, $C_{r_0}(\mathcal{X}_n) \subset S$, if $G_n = \left\{ \inf_{x\in C_{r_0}(\mathcal{X}_n)} \left(\frac{f(x)}{\hat{f}_n(x)}\right)^{1/d} \geq 1-\epsilon_n^+ \right\}$ remains true, we can ensure that, $\Delta_n(\mathcal{X}_n) \geq (1-\epsilon_n^+)\hat{\delta}(C_{r_0}(\mathcal{X}_n)\backslash\mathcal{X}_n)$. Therefore,

$$\{\hat{V}_{n,r_0} > c_{n,\alpha_n}\} \cap G_n \subset \{V_n(\mathcal{X}_n) > c_{n,\alpha_n}(1-\epsilon_n^+)^d\} \cap G_n.$$

Then,

$$\begin{aligned}\mathbb{P}(\hat{V}_{n,r_0} > c_{n,\alpha_n}) &= \mathbb{P}\left(\{\hat{V}_{n,r_0} > c_{n,\alpha_n}\} \cap G_n\right) + \mathbb{P}\left(\{\hat{V}_{n,r_0} > c_{n,\alpha_n}\} \cap G_n^c\right) \\ &\leq \mathbb{P}\left(V_n(\mathcal{X}_n) > c_{n,\alpha_n}(1-\epsilon_n^+)^d\right) + \mathbb{P}(G_n^c).\end{aligned}$$

Lemma 1 (ii) guarantees that $\mathbb{P}(G_n^c, \text{ i. o.}) = 0$. Therefore, the second term of the inequality is negligible and its is verified that

$$\limsup \mathbb{P}(\hat{V}_{n,r_0} > c_{n,\alpha_n}) \leq \mathbb{P}(V_n(\mathcal{X}_n) > (1-\epsilon_n^+)^d c_{n,\alpha_n}).$$

Consequently, $\mathbb{P}(\hat{V}_{n,r_0} > c_{n,\alpha_n})$ can be majorized by,

$$\begin{aligned}\mathbb{P}(U(\mathcal{X}_n) > &-(1-\epsilon_n^+)^d \log(-\log(1-\alpha_n)) + ((1-\epsilon_n^+)^d - 1)(\log(n) \\ &+(d-1)\log(\log(n)) + \log(\beta))).\end{aligned}$$

According to Theorem 1, $U(\mathcal{X}_n) \xrightarrow{d} U$ when $n \to \infty$. Furthermore, notice that U has a continuous distribution, so convergence in distribution implies that

$$\sup_u |\mathbb{P}(U(\mathcal{X}_n) \leq u) - \mathbb{P}(U \leq u)| \to 0.$$

Since $\alpha_n \to 0$ and $\log(n)\epsilon_n^+ \to 0$, we can prove

$$\mathbb{P}(U > -(1-\epsilon_n^+)^d \log(-\log(1-\alpha_n)) + ((1-\epsilon_n^+)^d - 1)(\log(n) + (d-1)\log(\log(n)) + \log(\beta))) \to 0.$$

This ensures that

$$\mathbb{P}(U(\mathcal{X}_n) > -(1-\epsilon_n^+)^d \log(-\log(1-\alpha_n)) + ((1-\epsilon_n^+)^d - 1)(\log(n) + (d-1)\log(\log(n)) + \log(\beta))) \to 0.$$

Therefore, $\mathbb{P}(\hat{r}_0 \geq r_0) \to 1$. □

It remains to prove that $\hat{r}_0$ cannot be arbitrarily larger that r_0. An auxiliary result must be proved.

Proposition 2 *Let $\mathcal{X}_n$ be a random and i.i.d sample drawn according to a density f that satisfies ($f_{0,1}^L$) with compact, nonconvex and nonempty support S under (R). Let r_0 be the parameter defined in (1) and $\{\alpha_n\} \subset (0,1)$ a sequence converging to zero such that $\log(\alpha_n)/n \to 0$. Then, for any $\epsilon > 0$,*

$$\mathbb{P}\left(\hat{r}_0 \leq r_0 + \epsilon, \ \textit{eventually}\right) = 1.$$

Proof Given $\epsilon > 0$ let be $r = r_0 + \epsilon$. According to Lemma 2, there exists $x_0 \in \mathbb{R}^d$ and $\rho > 0$ such that $B_\rho(x_0) \cap S = \emptyset$ and

$$\mathbb{P}\left(B_\rho(x_0) \subset C_r(\mathcal{X}_n), \ \text{eventually}\right) = 1.$$

Since, with probability one, $\mathcal{X}_n \subset S$ we have $B_\rho(x_0) \cap \mathcal{X}_n = \emptyset$. Then, $\{x_0\} \oplus \rho B_1[0] \subset C_r(\mathcal{X}_n) \backslash \mathcal{X}_n$. According to Lemma 1 (i), with probability one and for n large enough, there exists a constant $\lambda_1 > 0$ such that for all $x \in C_r(\mathcal{X}_n)$, it is verified that $\hat{f}_n^{1/d}(x) \geq \lambda_1$. Then, let γ be the positive constant $\lambda_1 \rho w_d^{1/d}$. Then, it is trivial to check that, with probability one and for n large enough,

$$\{x\} \oplus \frac{\gamma}{\hat{f}_n^{1/d}(x)} A \subset C_r(\mathcal{X}_n) \backslash \mathcal{X}_n.$$

Therefore, $\hat{\delta}(C_r(\mathcal{X}_n) \backslash \mathcal{X}_n) \geq \gamma > 0$ and, consequently, $\hat{V}_{n,r} = c_\gamma > 0$. Furthermore, since $C_r(\mathcal{X}_n) \subset C_{r'}(\mathcal{X}_n)$ for all $r' \geq r$, it is satisfied that $\hat{V}_{n,r'} \geq \hat{V}_{n,r} = c_\gamma > 0$. On the other hand, since $-u_{\alpha_n}/\log(\alpha_n) = \log(-\log(1-\alpha_n))/\log(\alpha_n) \to 1$, we have $c_{n,\alpha_n} \to 0$. Then, with probability one and for n large enough, we have $c_{n,\alpha_n} < c_\gamma$. Therefore, according the definition established in (3), $\hat{r}_0 \leq r$. □

Theorem 3 is a straightforward consequence of Propositions 1 and 2.

Proof of Theorem 4 Theorem 3 of Rodríguez-Casal (2007) ensures that, under (R) when $\mathfrak{r} = \lambda = \widetilde{r}$), then $\mathbb{P}(\mathcal{E}_n, \text{ eventually}) = 1$, where

$$\mathcal{E}_n = \left\{ d_H(S, C_{\widetilde{r}}(\mathcal{X}_n)) \leq D \left(\frac{\log n}{n} \right)^{2/(d+1)} \right\},$$

and D is some constant. Under the hypothesis of Theorem 4, this holds for any $\widetilde{r} \leq \min\{\mathfrak{r}, \lambda\}$. Fix one $\widetilde{r} \leq \min\{\mathfrak{r}, \lambda\}$ such that $\widetilde{r} < \nu r_0$ and define $\mathcal{R}_n = \{\widetilde{r} \leq r_n \leq r_0\}$. Since, by Theorem 3, $r_n = \nu \hat{r}_0$ converges in probability to νr_0 and $\widetilde{r} < \nu r_0 < r_0$, we have that $\mathbb{P}(\mathcal{R}_n) \to 1$. If the events $\mathcal{E}_n$ and $\mathcal{R}_n$ hold (notice that $\mathbb{P}(\mathcal{E}_n \cap \mathcal{R}_n) \to 1$), we have $C_{\widetilde{r}}(\mathcal{X}_n) \subset C_{r_n}(\mathcal{X}_n) \subset S$ and, therefore,

$$d_H(S, C_{r_n}(\mathcal{X}_n)) \leq d_H(S, C_{\widetilde{r}}(\mathcal{X}_n)) \leq D \left(\frac{\log n}{n} \right)^{2/(d+1)}.$$

This completes the proof of the first statement of Theorem 4. Similarly, it is possible to prove the result for the other error criteria considered in Theorem 4. □

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

References

Aaron C, Cholaquidis A, Fraiman R (2017) A generalization of the maximal-spacings in several dimensions and a convexity test. Extremes 20(3):605–634

Arias-Castro E, Pateiro-López B, Rodríguez-Casal A (2019) Minimax Estimation of the Volume of a Set under the Rolling Ball Condition. J Am Stat Assoc Theory Methods **114**, 1162–1173

Baíllo A, Chacón J E (2018) A survey and a new selection criterion for statistical home range estimation *arXiv preprint* arXiv:1804.05129

Baíllo A, Cuevas A (2001) On the estimation of a star-shaped set. Adv Appl Prob 33(4):717–726

Berrendero JR, Cuevas A, Pateiro-López B (2012) A multivariate uniformity test for the case of unknown support. Stat Comput 22(1):259–271

Burgman MA, Fox JC (2003) Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. Animal Conserv 6(1):19–28

Chevalier J (1976) Estimation du support et du contour du support d'une loi de probabilité Annales de l'IHP Probabilités et statistiques 12:339–364

Cuevas A, Fraiman R (2010) Set estimation. In: Kendall WS, Molchanov I (eds) New perspectives on stochastic geometry. Oxford University Press, pp 374–397

Cuevas A, Fraiman R, Pateiro-López B (2012) On statistical properties of sets fulfilling rolling-type conditions. Adv Appl Prob 44(2):311–329

Cuevas A, Rodríguez-Casal A (2004) On boundary estimation. Adv Appl Prob 36(2):340–354

De Haan L, Resnick S (1994) Estimating the home range. J Appl Prob 31(3):700–720

Deheuvels P (1983) Strong bounds for multidimensional spacings. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 64(4):411–424
Devroye L, Wise GL (1980) Detection of abnormal behavior via nonparametric estimation of the support. SIAM J Appl Math 38(3):480–488
Dümbgen L, Walther G (1996) Rates of convergence for random approximations of convex sets. Adv Appl Prob 28(2):384–393
Edelsbrunner H (2014) A short course in computational geometry and topology Springer, Number Mathematical methods
GBIF.org (27th May, (2019) Gbif occurrence Download https://doi.org/10.15468/dl.jtoo0d
Genovese CR, Perone-Pacifico M, Verdinelli I, Wasserman L (2012) The geometry of nonparametric filament estimation. J Am Stat Assoc 107(498):788–799
Hall P (1991) On convergence rates of suprema. Prob Theory Relat Fields 89(4):447–455
IUCN (2012). Iucn red list categories and criteria: Version 3.1 second edition iucn. *Gland, Switzerland: and Cambridge, UK: IUCN, iv + 32pp*
Janson S (1987) Maximal spacings in several dimensions. Annal Prob 15(1):274–280
Joppa LN, Butchart SH, Hoffmann M, Bachman SP, Akçakaya HR, Moat JF, Böhm M, Holland RA, Newton A, Polidoro B et al (2016) Impact of alternative metrics on estimates of extent of occurrence for extinction risk assessment. Conserv Biol 30(2):362–370
Lindgren F, Rue H, Lindström J (2011) An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. J R Stat Soc Series B (Stat Methodol) 73(4):423–498
Mandal DP, Murthy C (1997) Selection of alpha for alpha-hull in $\mathbb{R}^2$. Pattern Recognit 30(10):1759–1767
Martínez-Minaya J, Cameletti M, Conesa D, Pennino MG (2018) Species distribution modeling: a statistical review with focus in spatio-temporal issues. Stoch Environ Res Risk Assess 32(11):3227–3244
Pateiro-López B, Rodríguez-Casal A (2010) Generalizing the convex hull of a sample: the R package alphahull. J Stat Soft 34(1):1–28
Reitzner M (2003) Random polytopes and the efron-stein jackknife inequality. Annal Prob 31(4):2136–2166
Rodríguez-Casal A (2007) Set estimation under convexity type assumptions. Annales de l'IHP Probabilités et statistiques 43:763–774
Rodríguez-Casal A, Saavedra-Nieves P (2016) A fully data-driven method for estimating the shape of a point cloud. ESAIM: Prob Stat 20:332–348
Rondinini C, Wilson KA, Boitani L, Grantham H, Possingham HP (2006) Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecol Lett 9(10):1136–1145
Rue H, Martino S, Chopin N (2009) Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. J R Stat Soc Series b (Stat Methodol) 71(2):319–392
Schneider R (1988) Random approximation of convex sets. J Microsc 151(3):211–227
Schneider R (2014) Convex bodies: the Brunn-Minkowski theory. Cambridge University Press, Cambridge
Serra J (1983) Image analysis and mathematical morphology. Academic Press, London
Walther G (1997) Granulometric smoothing Annal Stat 2273–2299
Walther G (1999) On a generalization of blaschkes rolling theorem and the smoothing of surfaces. Math Methods Appl Sci 22(4):301–316