EXTENT OF OCCURRENCE ESTIMATION FOR INVASIVE PLANTS

Paula Saavedra-Nieves and Alberto Rodríguez-Casal*

*Universidade de Santiago de Compostela

Innpar2D Workshop

10th December 2019

Azorean islands, Terceira and São Miguel

Figure: Geographical location of the Autonomous Region of the Azores.

NASA Terra satellite.

Extent of Occurrence for Invasive Plants in Portugal

5/28 invasive species in the database

Erigeron Karvinskianus

Pittosporum Undulatum

Agave Americana

Acacia Melanoxylon

Hedychium Gardnerianum

Figure: Terceira and São Miguel islands.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

International Union for Conservation of Nature

Extent of Occurrence for Invasive Plants in Portugal

Figure: 740 geographical locations for invasive plants in Terceira and São Miguel.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

International Union for Conservation of Nature

Extent of Occurrence for Invasive Plants in Portugal

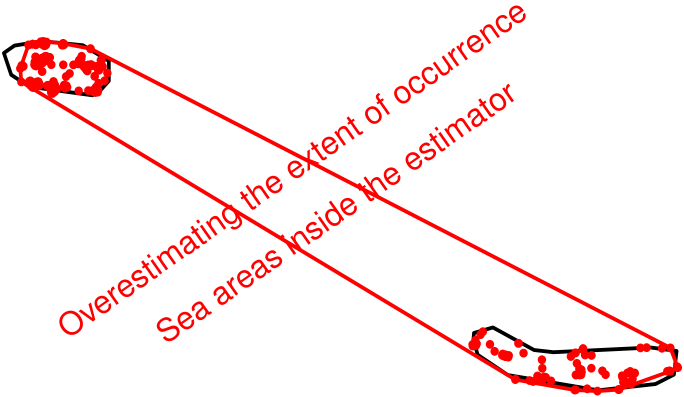


Figure: Convex hull of geographical locations.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

International Union for Conservation of Nature

Extent of Occurrence for Invasive Plants in Portugal

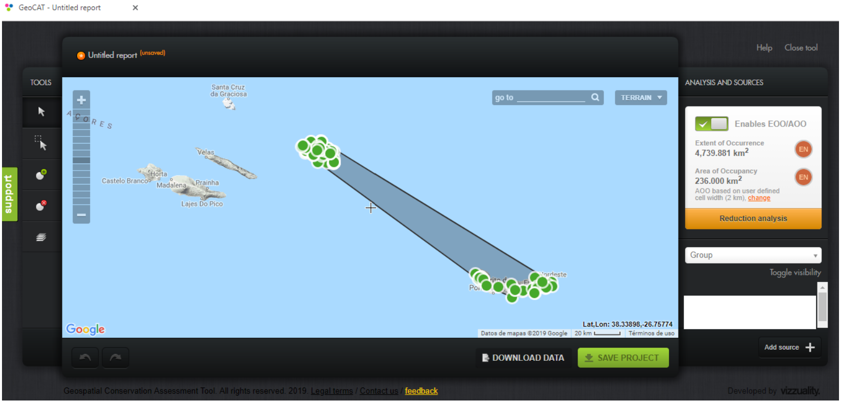


Figure: Geographical locations of invasive plants in Terceira and São Miguel islands.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

International Union for Conservation of Nature

http://geocat.kew.org/

Extent of Occurrence for Invasive Plants in Portugal

The convex hull of a sample is the intersection of all halfspaces that contain it.

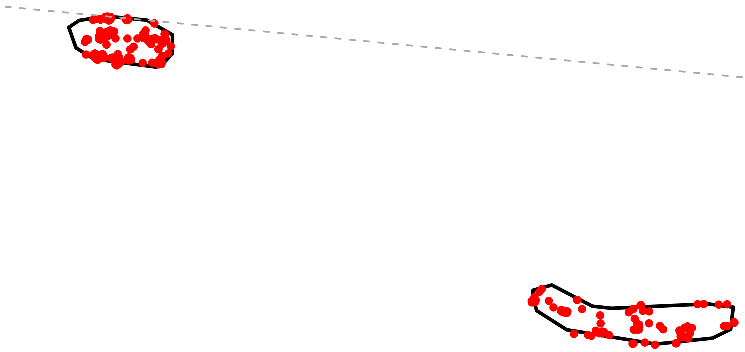


Figure: Convex hull of sample points.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Extent of Occurrence for Invasive Plants in Portugal

The convex hull of a sample is the intersection of all halfspaces that contain it.

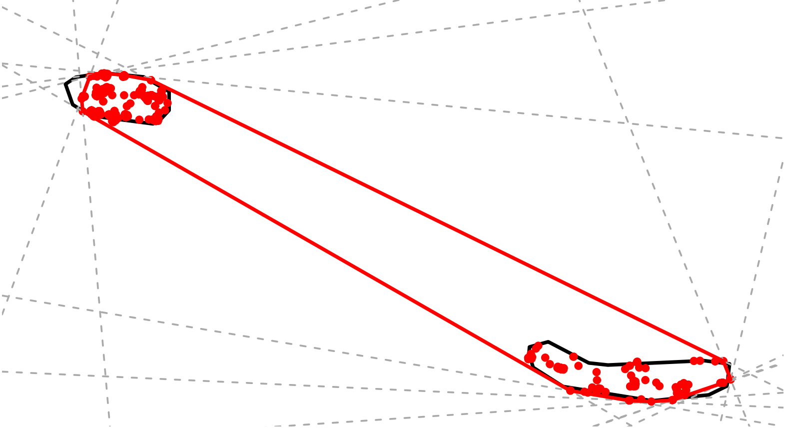


Figure: Convex hull of sample points.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Extent of Occurrence for Invasive Plants in Portugal

The $r-$convex hull of sample points for $r > 0$ can be calculated as the intersection of the complementaries of balls with radius bigger or equal than $r > 0$ containing the sample.

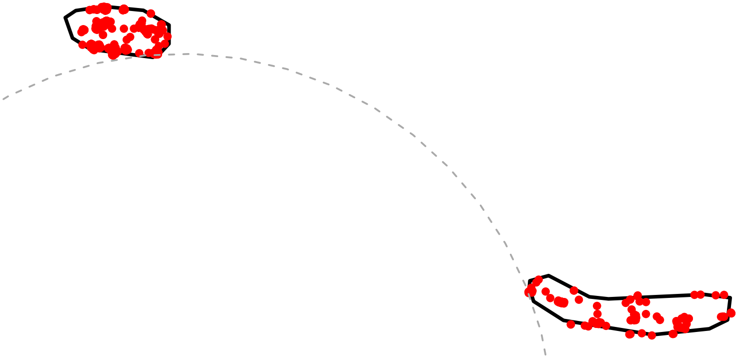


Figure: $r-$convex hull with $r = 0.3$.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Extent of Occurrence for Invasive Plants in Portugal

The $r-$convex hull of sample points for $r > 0$ can be calculated as the intersection of the complementaries of balls with radius bigger or equal than $r > 0$ containing the sample.

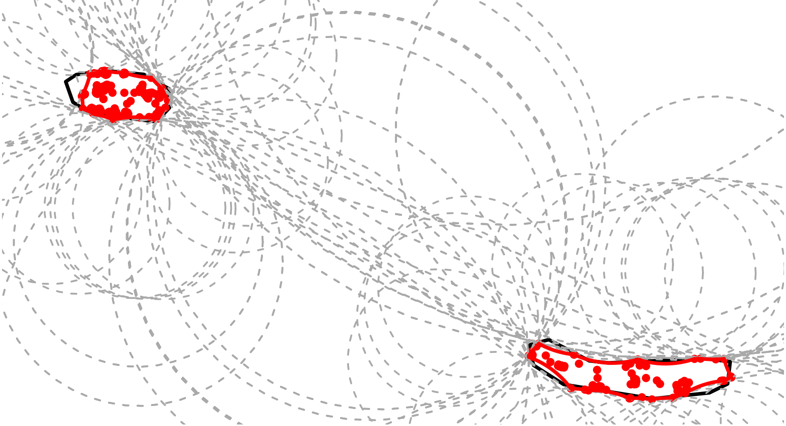


Figure: $r-$convex hull with $r = 0.3$.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Extent of Occurrence for Invasive Plants in Portugal

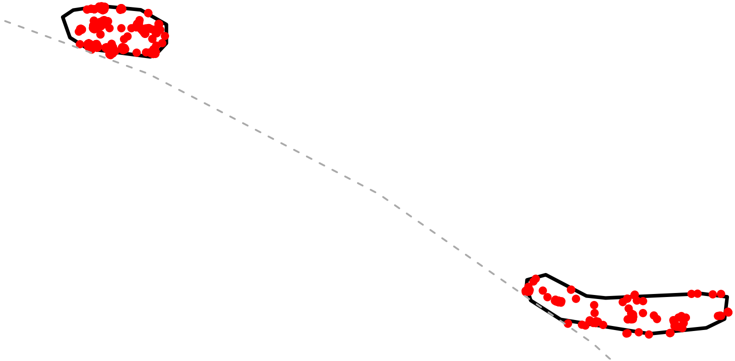


Figure: $r-$convex hull with $r = 5$.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Rodríguez-Casal, A. and Saavedra-Nieves, P. (2016). A fully data-driven method for estimating the shape of a point cloud. ESAIM: Probability and Statistics, 332-348.

Extent of Occurrence for Invasive Plants in Portugal

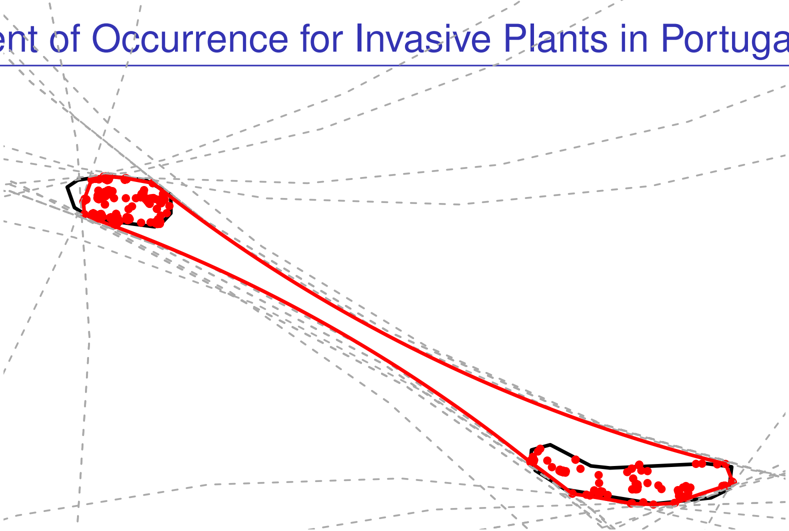


Figure: $r-$convex hull with $r = 5$

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Rodríguez-Casal, A. and Saavedra-Nieves, P. (2016). A fully data-driven method for estimating the shape of a point cloud. ESAIM: Probability and Statistics, 332-348.

Extent of Occurrence for Invasive Plants in Portugal

Is it possible to estimate the shape index *r* from data?

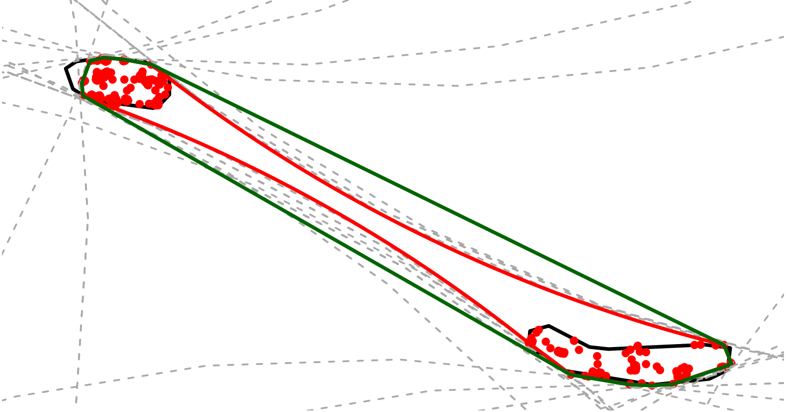


Figure: $r-$convex hull with $r = 5$ (red) and convex hull (darkgreen).

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Rodríguez-Casal, A. and Saavedra-Nieves, P. (2016). A fully data-driven method for estimating the shape of a point cloud. ESAIM: Probability and Statistics, 332-348.

Extent of Occurrence for Invasive Plants in Portugal

Is it possible to estimate the shape index *r* from data?

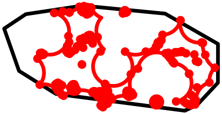
Figure: $r-$convex hull with $r = 0.03$.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Rodríguez-Casal, A. and Saavedra-Nieves, P. (2016). A fully data-driven method for estimating the shape of a point cloud. ESAIM: Probability and Statistics, 332-348.

Extent of Occurrence for Invasive Plants in Portugal

Is it possible to estimate the shape index r from data?



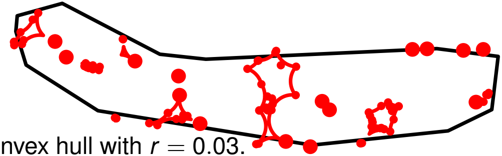


Figure: $r-$convex hull with $r = 0.03$.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Rodríguez-Casal, A. and Saavedra-Nieves, P. (2016). A fully data-driven method for estimating the shape of a point cloud. ESAIM: Probability and Statistics, 332-348.

Table of contents

- Support estimation under $r-$convexity assumption.
  - Interpretation and influence of the shape index $r > 0$.
- A new automatic method to estimate the optimal value of r.
  - Theoretical results.
- Analysis of the resulting support estimator.
  - Convergence rates.
- Real data analysis.
  - Estimating the extent of occurrence for invasive plants.

Support estimation

Main task: Reconstructing the support S for an absolutely continuous probability measure from a random sample of points, $\mathcal{X}_n = \{X_1, ..., X_n\}$.

Our goal: Proposing a new support estimator for reconstructing S that is $r-$convex and usually unknown.

How?

Step 1. Estimating the optimal value of the unknown parameter r, $\hat{r}$, from $\mathcal{X}_n$.

Step 2. Analyzing the $\hat{r}-$convex hull of $\mathcal{X}_n$ as an estimator for S.

Support estimation

Our scenario: We will assume that S is $r-$convex which is a more flexible and general geometric condition than convexity*.

Definition: A set $A \subset \mathbb{R}^d$ is said to be $r-$convex, for $r > 0$, if

$$A = C_r(A),$$

where

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

is the $r-$convex hull of A and $B_\epsilon(x)$, the open ball of radius $\epsilon > 0$ centered at x.

*If A is convex and closed then it is also $r-$convex for all $r > 0$ (see Walther, 1999).

Walther, G. (1999). On a generalization of Blaschke's rolling theorem and the smoothing of surfaces. Mathematical Methods in the Applied Sciences, 22, 301-316.

Support estimation

How could we interpret the parameter r?:

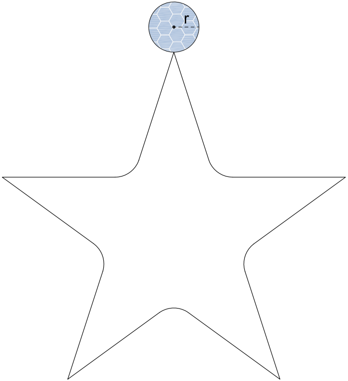


Figure: A ball of radius r rolls freely in $\overline{A^c}$.

*If A is $r-$convex then A ball of radius r rolls freely in $\overline{A^c}$ (see Cuevas et al., 2012).

Cuevas, A., Fraiman, R. and Pateiro-López, B. (2012). On statistical properties of sets fulfilling rolling-type conditions. Advances in Applied Probability, 44, 311–329.

A new method for selecting r

Our goal: We will estimate $r_0 = \sup\{r > 0 : C_r(S) = S\}$ under (f_L):

(f_L) $\mathcal{X}_n$ is generated from a density f that is bounded from below and Lipschitz continuous restricted to its bounded support S.

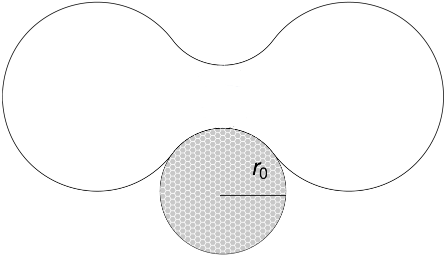


Figure: The value r_0 could be seen as a shape parameter.

r_0 is the optimal value to be estimated:

- If S is $r-$convex then it is r^*-convex for all $0 < r^* \leq r$. So, $C_{r^*}(S) \subset C_r(S) = S$.

A new method for selecting r

Our goal: We will estimate $r_0 = \sup\{r > 0 : C_r(S) = S\}$ under (f_L):

(f_L) $\mathcal{X}_n$ is generated from a density f that is bounded from below and Lipschitz continuous restricted to its bounded support S.

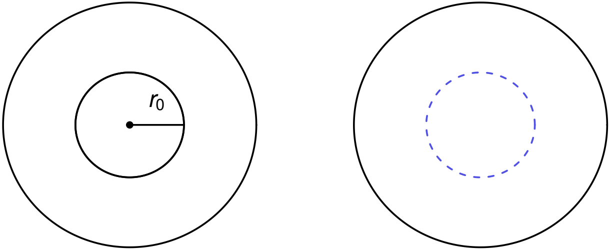


Figure: $C_r(S)$ (right) is considerably bigger than the circular ring S (left) if $r > r_0$.

r_0 is the optimal value to be estimated:

- If S is $r-$convex then it is r^*-convex for all $0 < r^* \leq r$. So, $C_{r^*}(S) \subset C_r(S) = S$.
- If $r > r_0$ then $C_r(S)$ and S could be very different

A new method for selecting r

Is always $r_0 = \sup\{r > 0 : C_r(S) = S\}$ a maximum? It is a maximum under (R_λ^r).

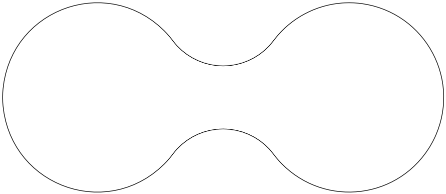


Figure: (R_λ^r) A ball of radius $\lambda > 0$ rolls freely in S and a ball of radius $r > 0$ rolls in $\overline{S^c}$.

If $r = \lambda$, the boundary is smooth and the curvature is bounded by $1/r$, see Walther (1997).

Walther, G. (1997). Granulometric smoothing. Annals of Statistics, 25, 2273–2299.

A new method for selecting r

How will we estimate the shape parameter r_0? Testing $r-$convexity for a given $r > 0$!

If S is not $r-$convex for a large r, the maximal spacing in $C_r(\mathcal{X}_n)$ will be larger than the maximal spacing in S.

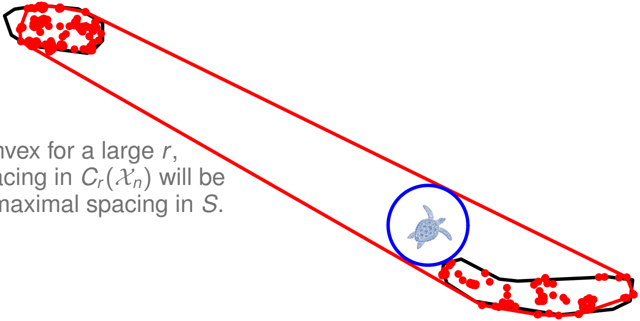


Figure: Largest ball (blue) inside the convex hull of $\mathcal{X}_n$ that does not intersect $\mathcal{X}_n$.

- Given $r > 0$, we have proposed a test:

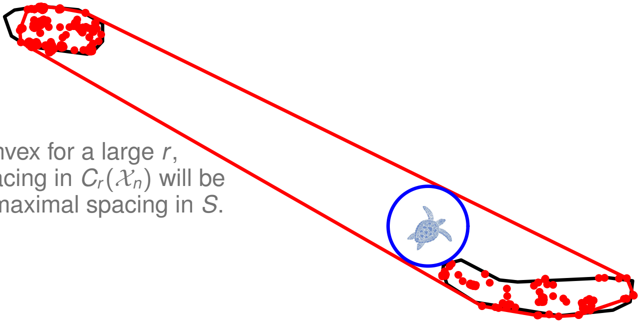
$$H_0 : S \text{ is } r- \text{ convex versus } H_1 : S \text{ is not } r-\text{convex}.$$

- This test rejects H_0 when it detects a large spacing in $C_r(\mathcal{X}_n)$.
- Under (f_L) and (R_λ^r), the test rejects H_0 if r is too large.

Janson, S. (1987). Maximal Spacings in Several Dimensions. Annals of Probability, 1, 274–280.

A new method for selecting r

How will we estimate the shape parameter r_0? Testing $r-$convexity for a given $r > 0$!



If S is not $r-$convex for a large r, the maximal spacing in $C_r(\mathcal{X}_n)$ will be larger than the maximal spacing in S.

Figure: Largest ball (blue) inside the convex hull of $\mathcal{X}_n$ that does not intersect $\mathcal{X}_n$.

- Given $r > 0$, we have proposed a test:

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

- This test rejects H_0 when it detects a large spacing in $C_r(\mathcal{X}_n)$.
- Under (f_L) and (R_λ^r), the test rejects H_0 if r is too large.

Berrendero, J. R., Cuevas, A. and Pateiro-López, B. (2012). A multivariate uniformity test for the case of unknown support. Statistics and Computing, 22, 259-271.

A new method for selecting r

How will we estimate the shape parameter r_0? Testing $r-$convexity for a given $r > 0$!

If S is not $r-$convex for a large r, the maximal spacing in $C_r(\mathcal{X}_n)$ will be larger than the maximal spacing in S.

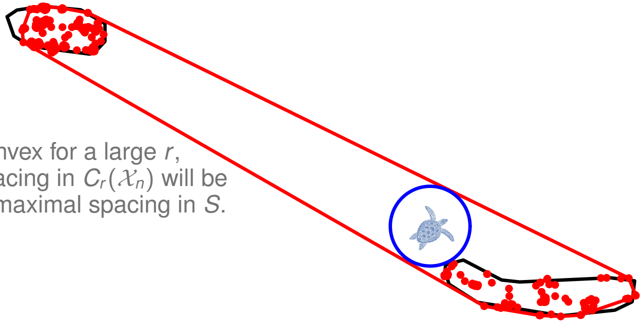


Figure: Largest ball (blue) inside the convex hull of $\mathcal{X}_n$ that does not intersect $\mathcal{X}_n$.

- Given $r > 0$, we have proposed a test:

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

- This test rejects H_0 when it detects a large spacing in $C_r(\mathcal{X}_n)$.
- Under (f_L) and (R_λ^r), the test rejects H_0 if r is too large.

Rodríguez-Casal, A. and Saavedra-Nieves, P. (2016). A fully data-driven method for estimating the shape of a point cloud. ESAIM: Probability and Statistics, 332-348.

A new method for selecting r

How will we estimate the shape parameter r_0? Testing $r-$convexity for a given $r > 0$!

If S is not $r-$convex for a large r, the maximal spacing in $C_r(\mathcal{X}_n)$ will be larger than the maximal spacing in S.

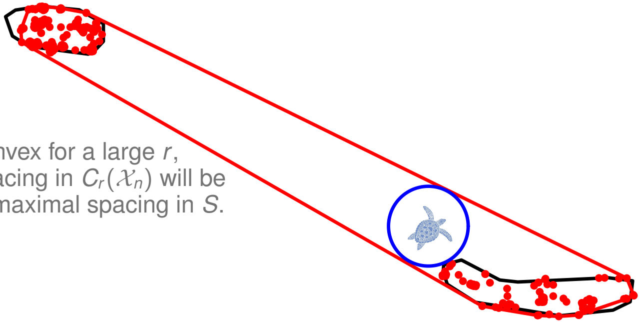


Figure: Largest ball (blue) inside the convex hull of $\mathcal{X}_n$ that does not intersect $\mathcal{X}_n$.

- Given $r > 0$, we have proposed a test:

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

- This test rejects H_0 when it detects a large spacing in $C_r(\mathcal{X}_n)$.
- Under (f_L) and (R_λ^r), the test rejects H_0 if r is too large.

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

A new method for selecting r

How will we estimate the shape parameter r_0? Testing $r-$convexity for a given $r > 0$!

If S is really $r-$convex, the maximal spacings in $C_r(\mathcal{X}_n)$ and S will be similar.

Figure: Largest ball (blue) inside the estimator $C_{0.3}(\mathcal{X}_n)$ that does not intersect $\mathcal{X}_n$.

- Given $r > 0$, we have proposed a test:

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

- This test rejects H_0 when it detects a large spacing in $C_r(\mathcal{X}_n)$.
- Under (f_L) and (R_λ^r), the test rejects H_0 if r is too large.

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

A new method for selecting r

How will we estimate the shape parameter r_0? Testing $r-$convexity for a given $r > 0$!

Figure: Largest ball (blue) inside the known support S that does not intersect $\mathcal{X}_n$.

- Given $r > 0$, we have proposed a test:

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

- This test rejects H_0 when it detects a large spacing in $C_r(\mathcal{X}_n)$.
- Under (f_L) and (R_λ^r), the test rejects H_0 if r is too large.

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

A new method for selecting r

Our estimator:

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

Dependence on the significance level α of the test:

- It is not important from the theoretical point of view:
  - It is assumed that $\alpha = \alpha_n$ goes to zero as the sample size increases.
- In practice, $\hat{r}_0$ could be too small for a fixed sample:
  - The level of fragmentation of the estimator was bounded.

Consistency of the new method for selecting r

Theorem: Let α_n be a sequence converging to zero. Under (R_λ^r), (f_L) and some additonal assumptions,

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

Theorem: Let α_n be a sequence converging to zero such that $\log(\alpha_n)/n \to 0$. Under (R_λ^r), (f_L) and some additonal assumptions,

$$\hat{r}_0 \to r_0 \text{ in probability.}$$

Consistency of the new method for selecting r

Theorem: Let α_n be a sequence converging to zero. Under (R^r_λ), (f_L) and some additonal assumptions,

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

Theorem: Let α_n be a sequence converging to zero such that $\log(\alpha_n)/n \to 0$. Under (R^r_λ), (f_L) and some additonal assumptions,

$$\hat{r}_0 \to r_0 \text{ in probability.}$$

*Could we consider $C_{\hat{r}_0}(\mathcal{X}_n)$ as an estimator for the support S?

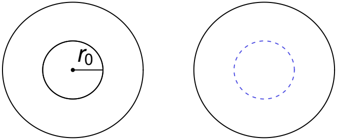


Figure: If $r > r_0$, $C_r(S)$ (right) can be considerably bigger than the circular ring S (left).

Consistency of the new support estimator

Theorem: Let α_n be a sequence converging to zero such that $\log(\alpha_n)/n \to 0$. Under (R_λ^r), (f_L) and some additonal assumptions, let be $\nu \in (0,1)$ and $r_n = \nu \hat{r}_0$. Then,

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

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

$$\mu(S \triangle C_{r_n}(\mathcal{X}_n)) = O_P\left(\left(\frac{\log n}{n}\right)^{\frac{2}{d+1}}\right).$$

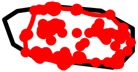
* **This estimator achieves the same convergence rates as the convex hull for convex sets, see Dümbgen and Walther (1996).**

Dümbgen, L. and Walther, G. (1996). Rates of convergence for random approximations of convex sets. Advances in Applied Probability, 28, 384-393.

Rodríguez-Casal, A. and Saavedra-Nieves, P. (2014). A fully data-driven method for estimating the shape of a point cloud. ESAIM: Probability and Statistics, 20, 332-348.

Real data analysis

Estimation of the extent of occurrence from the 740 geographical locations:



5/28 invasive species in the database

Erigeron Karvinskianus

Pittosporum Undulatum

Agave Americana

Acacia Melanoxylon

Hedychium Gardnerianum

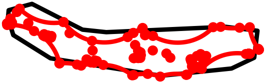
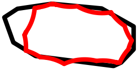


Figure: $\hat{r}_0-$convex hull of $\mathcal{X}_{740}$. The significance level for the test was fixed equal to 0.01.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Real data analysis

Estimation of the extent of occurrence from the 740 geographical locations:



5/28 invasive species in the database

Erigeron Karvinskianus

Pittosporum Undulatum

Agave Americana

Acacia Melanoxylon

Hedychium Gardnerianum

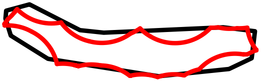
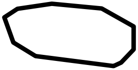


Figure: $\hat{r}_0-$convex hull of $\mathcal{X}_{740}$. The significance level for the test was fixed equal to 0.01.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Real data analysis

Estimation of the extent of occurrence in São Miguel island by year:



5/28 invasive species in the database

Erigeron Karvinskianus

Pittosporum Undulatum

Agave Americana

Acacia Melanoxylon

Hedychium Gardnerianum

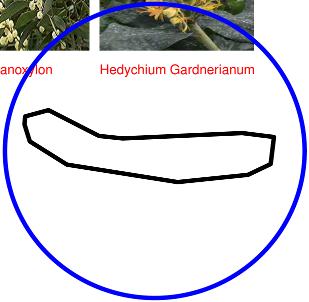


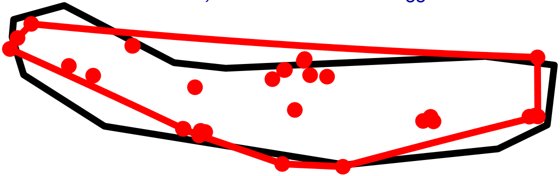
Figure: São Miguel island.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Real data analysis

Estimation of the extent of occurrence in São Miguel island by year:

2015, convex hull of $\mathcal{X}_{33}$



2016, $C_{1.5}(\mathcal{X}_{48})$

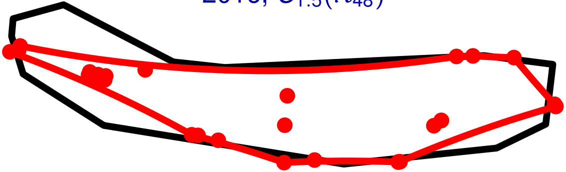


Figure: Estimations for the São Miguel island. The significance level for the test was fixed equal to 0.01.

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Real data analysis

Estimation of the extent of occurrence in São Miguel island by year:

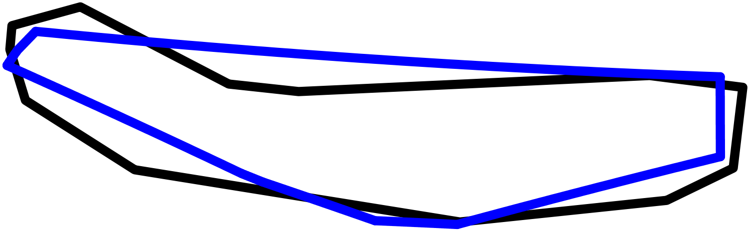


Figure: Evolution of invasive species in São Miguel island: 2015 (blue) and 2016 (gray).

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

Real data analysis

Estimation of the extent of occurrence in São Miguel island by year:

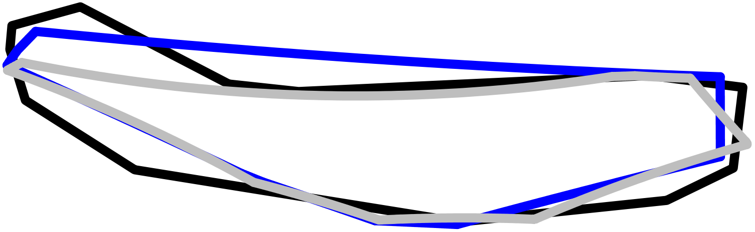


Figure: Evolution of invasive species in São Miguel island: 2015 (blue) and 2016 (gray).

GBIF.org (27th May 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.jtoo0d.

THANKS

EXTENT OF OCCURRENCE ESTIMATION FOR INVASIVE PLANTS

Paula Saavedra-Nieves and Alberto Rodríguez-Casal*

*Universidade de Santiago de Compostela

Innpar2D Workshop
10th December 2019