A Wilcoxon-Mann-Whitney Spatial Scan Statistic for Real and Functional Data

Zaineb Smida, Lionel Cucala, Ali Gannoun and Ghislain Durif

Institut Montpelliérain Alexander Grothendieck
Université de Montpellier, France

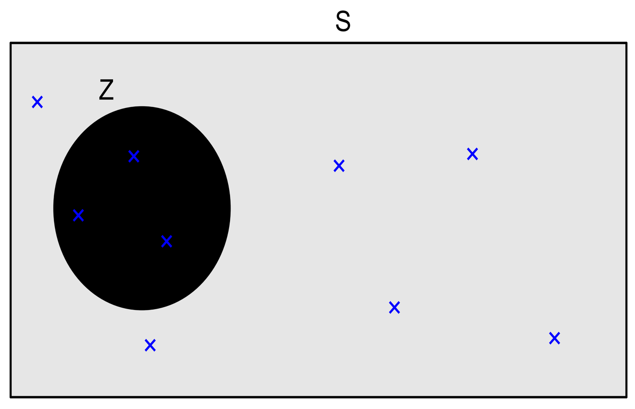
Introduction

- **Dataset:** Variable X observed in n spatial locations.
- **Goal:** Cluster detection.
- **Cluster** = spatial area Z in which X is significantly "different" from elsewhere.
- **Question:** Is there any significant cluster?

Scan method

- Statistical tests:

H_0 : "Absence of a cluster"
vs
$H_{1,Z}$: "Presence of a cluster Z".



- **Scan Statistic:** maximum of a concentration index over a set of potential clusters.
- **Significance:** Monte-Carlo procedure.

Method: Univariate case

- $Z \subset \mathcal{S}$: potential cluster and Z^c its complement.
- **Wilcoxon-Mann-Whitney statistic:**

$$SR(Z) = \sum_{i:s_i \in Z} R_i,$$

R_i : the rank of X_i.

- **Concentration index in Z:**

$$I_{\text{rank}}(Z) = \frac{SR(Z) - M(Z)}{\sqrt{V(Z)}},$$

$M(Z)$: the mean of $SR(Z)$ and $V(Z)$: the variance of $SR(Z)$ under H_0.

Method: Functional case

- $Z \subset \mathcal{S}$: potential cluster and Z^c its complement.
- **Wilcoxon-Mann-Whitney statistic:**

$$T_{\text{WMW}} = \frac{1}{n_Z n_{Z^c}} \sum_{\{i:s_i \in Z\}} \sum_{\{j:s_j \in Z^c\}} \frac{X_j - X_i}{\|X_j - X_i\|_{\mathcal{X}}},$$

n_Z: size of Z and n_{Z^c}: size of Z^c.

- **Concentration index in Z:**

$$\text{U}(Z) = (n_Z n_{Z^c}/n)^{1/2} T_{\text{WMW}}.$$

Significance of the tests

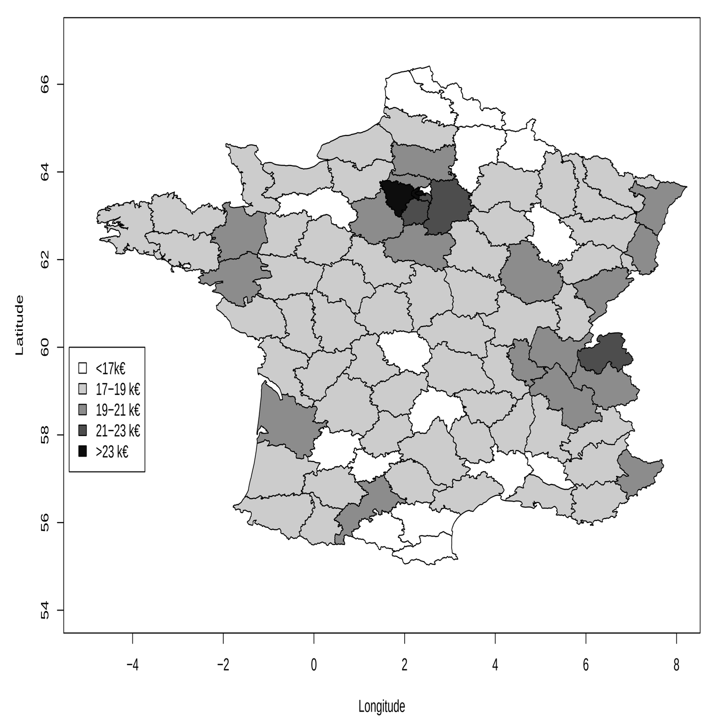
- T permutations of the X_i's : $\Lambda^{(1)}_{\text{WMW.SS}}, \cdots, \Lambda^{(T)}_{\text{WMW.SS}}$.
- **p-value:**

$$p_{value} = \frac{1 + \sum_{i=1}^{T} \mathbb{1}_{\{\Lambda^{(i)}_{\text{WMW.SS}} > \Lambda_{\text{WMW.SS}}\}}}{T+1}.$$

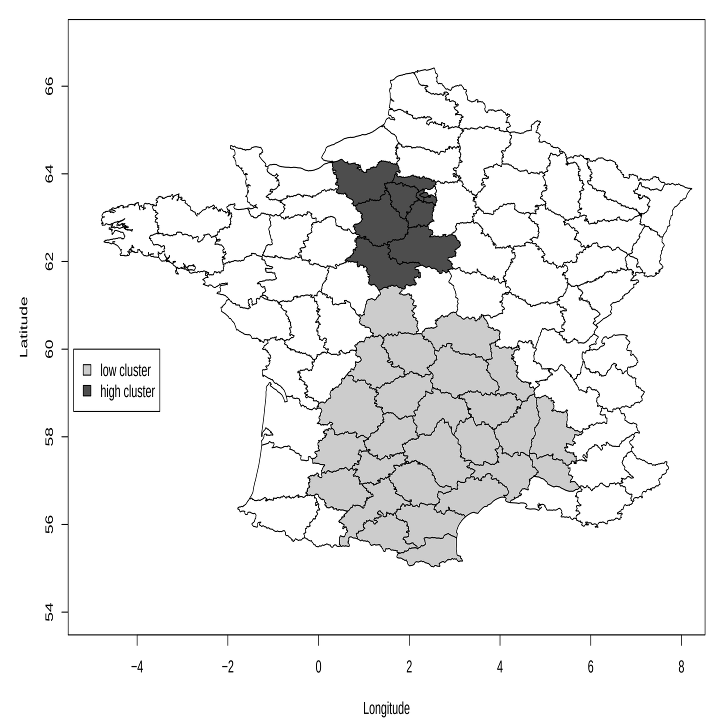
A Wilcoxon-Mann-Whitney spatial scan statistic for univariate data

- **Dataset:** $\{(X_i, s_i), i = 1, \cdots, n\}$, where $s_i \in S \subset \mathbb{R}^2$: spatial location
X_i: observation of real random variable X measured in location s_i (real marks).

- **Median income in France in 2012 [1]:**



- **MLC detected [1]:**



- **WMW univariate spatial scan statistic:**

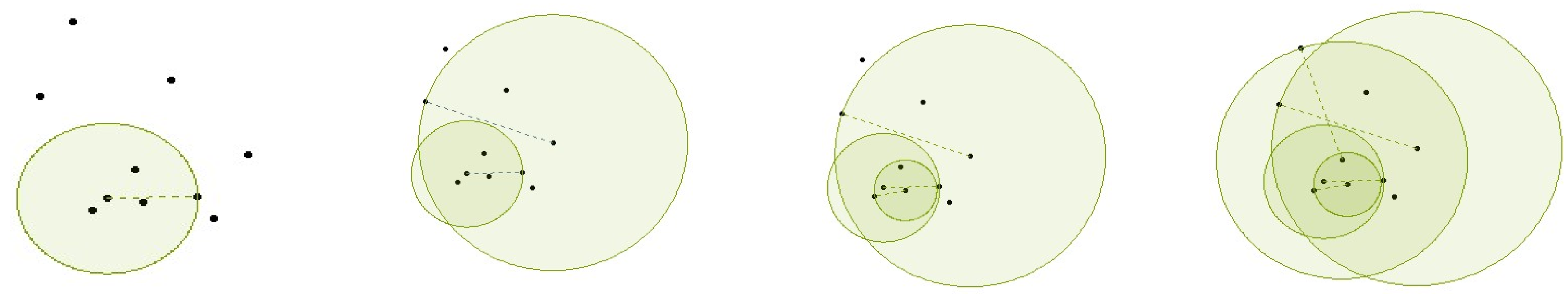
$$\Lambda_{\text{WMWUSS}} = \max_{Z \in \mathcal{D}} |I_{\text{rank}}(Z)| \,.$$

- **Most likely cluster (MLC):**

$$\hat{C} = \arg\max_{Z \in \mathcal{D}} |I_{\text{rank}}(Z)|.$$

A Set of potential clusters

- **Circular clusters:** $D_{i,j}$: disc with center s_i and passing through s_j.



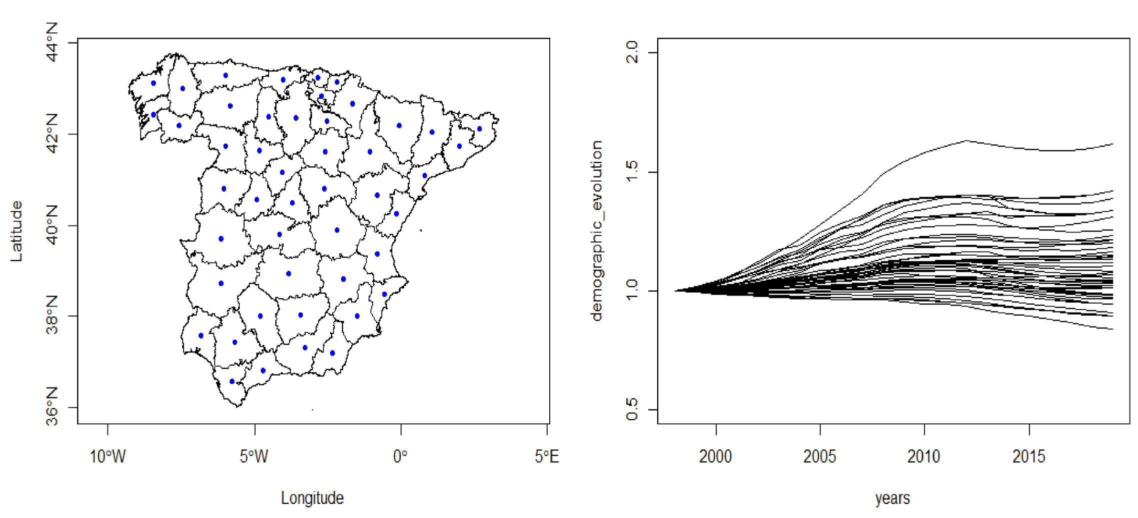
- **A set of potential clusters:**

$$\mathcal{D} = \{D_{i,j}, 1 \le i \le n, 1 \le j \le n\}.$$

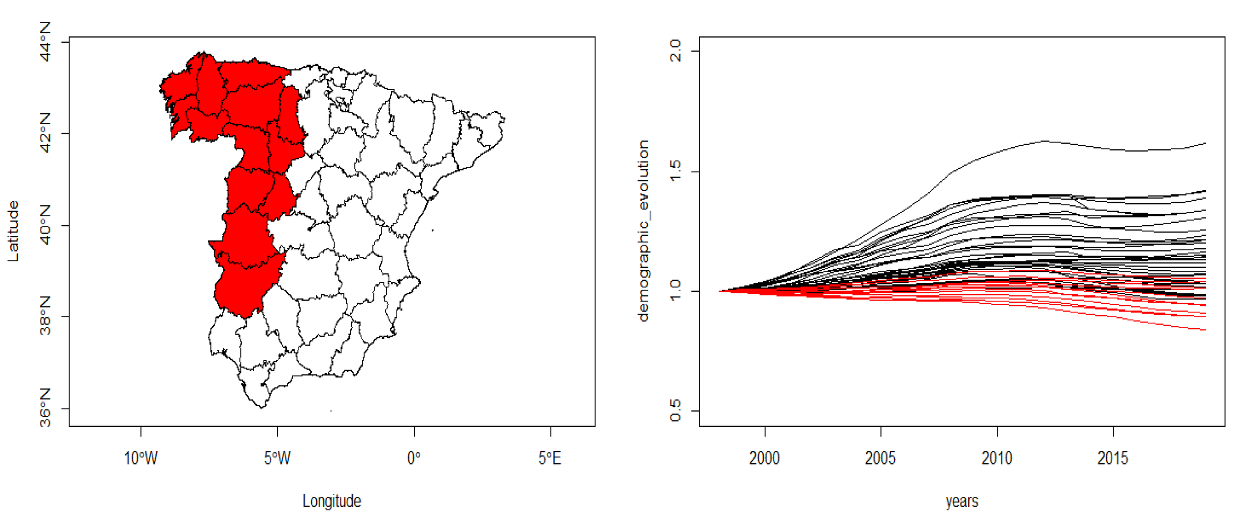
A Wilcoxon-Mann-Whitney spatial scan statistic for functional data

- **Dataset:** $\{(X_i, s_i), i = 1, \ldots, n\}$, where $s_i \in \mathcal{S} \subset \mathbb{R}^2$: spatial location
X_i: observation of $\mathcal{X}$-valued random element X measured in location s_i, $\mathcal{X}$: Hilbert space (functional marks).

- **Demographic evolution in Spain (from 1998 to 2019) [2]:**



- **MLC detected & its associated demographic curves [2]:**



- **WMW functional spatial scan statistic:**

$$\Lambda_{\text{WMWFSS}} = \max_{Z \in \mathcal{D}} \|U(Z)\|_{\mathcal{X}} \,.$$

- **Most likely cluster (MLC):**

$$\hat{C} = \arg\max_{Z \in \mathcal{D}} \|U(Z)\|_{\mathcal{X}}.$$

Conclusion and perspective

- **Conclusion:** Introduction of the nonparametric spatial scan statistics for real and functional data.
- **Perspective:** Develop another nonparametric spatial scan statistic based on the median test for functional data introduced in [3].

References

[1] Cucala, L. (2016). A Mann-Whitney scan statistic for continuous data. *Communications in Statistics - Theory and Methods*. **45**, 321–329.

[2] Smida, Z., Cucala, L., Gannoun, A. and Durif, G. (2022). A Wilcoxon-Mann-Whitney spatial scan statistic for functional data. *Computational Statistics & Data Analysis*. **167**, 107378.

[3] Smida, Z., Cucala, L., Gannoun, A. and Durif, G. (2022). A median test for functional data. *Journal of Nonparametric Statistics*. **34**, 520-553.

zaineb.smida@umontpellier.fr lionel.cucala@umontpellier.fr