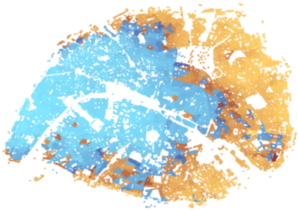


Xavier Bacon
xavier.bacon@ens-paris-saclay.fr
https://baconxavier.github.io/

Borelli Center
École Normale Supérieure de Paris-Saclay

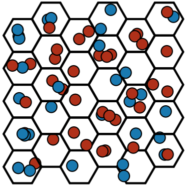
CCS 2025, September 02, 2025

Analysing Spatial Heterogeneity

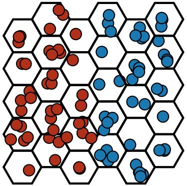
using Optimal Transport and applied to electoral results

in collaboration with Joseph Touzet (ÉNS Paris-Saclay)

Introduction



Weak Heterogeneity



High Heterogeneity

Minimal framework

- A finite set of areal units $\boldsymbol{\mathcal{X}} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots\}$
- Several groups (chemical or animal species, ethnic groups) distributed across $\boldsymbol{\mathcal{X}}$:

$$\boldsymbol{A} = \boldsymbol{A_1(x)}, \boldsymbol{A_2(x)}, \ldots$$

Goals

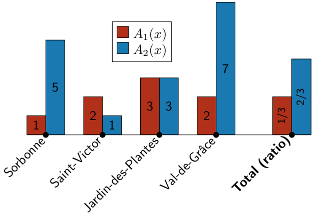
- Propose two indices (one global, one local) to measure the **spatial heterogeneity** of these distributions: are they well mixed or not?

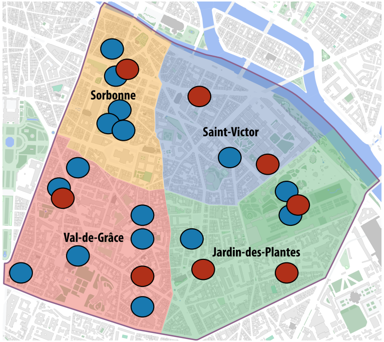
Similar works

- Dissimilarity Index (Duncan & Duncan, 1955), Isolation Index, Entropy-based measures Index (Theil, 1972), Multiscalar Lens Model (Olteanu et al., 2019), complex systems approaches (Agent-Based Modeling *via* Schelling, Network Analysis...).

xavier.bacon@ens-paris-saclay.fr

General context





Simulated election in Paris' Latin Quarter

Electoral data Modelling

- $\mathcal{X}$ = the set of polling stations (or districts).
- A_1 = votes cast for the candidate 1,
  A_2 = votes cast for the candidate 2, ...
- Distributions $\alpha_1, \alpha_2, \ldots$ correspond to the votes' ratio for the n-candidate, that is

$$\alpha_n(x) = \frac{A_n(x)}{\sum_y A_n(y)}$$

- Distance-cost matrix $C = C(x, y)$: straight-line distance between two polling stations.

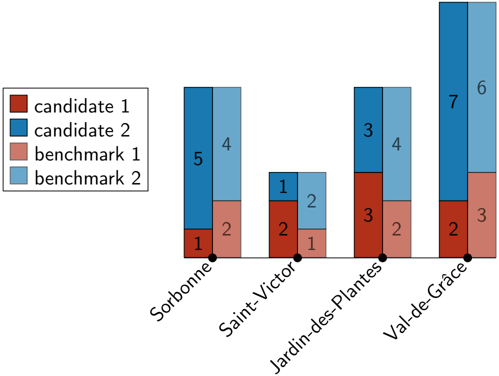
xavier.bacon@ens-paris-saclay.fr

A Global OT-Index

Global OT-Index for the n-candidate

$$D_n = \mathbf{OT}_C(\alpha_n, \bar{\alpha})$$

Candidates' scores and their respective benchmarks



- **Benchmark $\bar{\alpha}$:** represents the *perfect mixed* distribution of votes:

$$\bar{\alpha}(x) = \frac{\sum_n A_n(x)}{\sum_y \sum_n A_n(y)}.$$

- **Optimal Transport (OT):** a tool to compute a distance between two distributions.
- For the overall election:

$$\boldsymbol{D} = \sum_n p_n D_n$$

where p_n is the n-candidate's score.

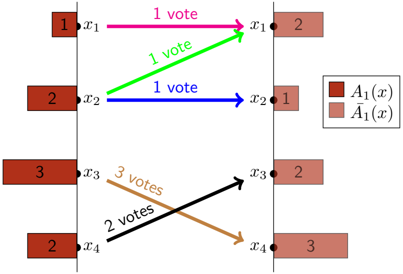
xavier.bacon@ens-paris-saclay.fr

Optimal Transport between α_n and $\bar{\alpha}$

$$\boldsymbol{D_n} = \mathsf{OT}_{\boldsymbol{C}}(\boldsymbol{\alpha_n}, \boldsymbol{\bar{\alpha}})$$
$$= \min\left\{\langle C \,|\, \gamma_n\rangle : \gamma_n \cdot \mathbb{1} = \alpha_n, \gamma_n^T \cdot \mathbb{1} = \bar{\alpha}\right\}$$

- Thanks to $C = C(x, y)$ which defines the " cost to move one vote from x to y ", D_n captured the geographic properties of $\boldsymbol{\mathcal{X}}$.
- γ_n dispatches votes from one polling station to another.
- D_n is a geographical index: spatial arrangements of areal units is taken into account.

$$\gamma_n = \frac{1}{8} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$



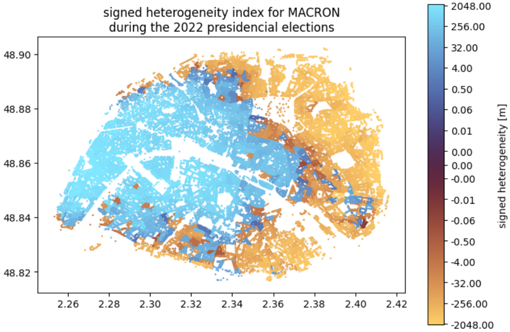
An example of transference plan γ_n

xavier.bacon@ens-paris-saclay.fr

A Local OT-Index

Local OT-Index

$$\omega_n(x) = \frac{1}{2\bar{\alpha}(x)} \langle C(x,\cdot) \mid \gamma_n^*(\cdot,x) - \gamma_n^*(x,\cdot) \rangle$$



- Optimal transference plans γ_n^* defines a flow between pairs of polling stations.
- Readily interpretable:
  C in m $\implies \omega_n$ and D_n in m
- French 2022 Presidential results of 4 candidates (among 12) *(Paris area)*:

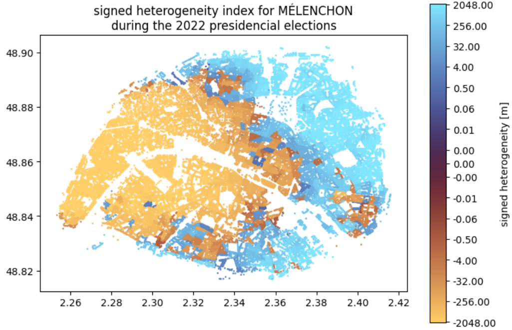
Candidate	OT-Index D_n	Score (%)
Macron	519 meters	35,49
Mélenchon	980 meters	29,93
Zemmour	1005 meters	8,2
Le Pen	295 meters	5,39

xavier.bacon@ens-paris-saclay.fr

A Local OT-Index

Local OT-Index

$$\omega_n(x) = \frac{1}{2\bar{\alpha}(x)} \langle C(x,\cdot) \,|\, \gamma_n^*(\cdot,x) - \gamma_n^*(x,\cdot) \rangle$$



- Optimal transference plans γ_n^* defines a flow between pairs of polling stations.
- Readily interpretable:
  C in m $\Longrightarrow$ ω_n and D_n in m
- French 2022 Presidential results of 4 candidates (among 12) *(Paris area)*:

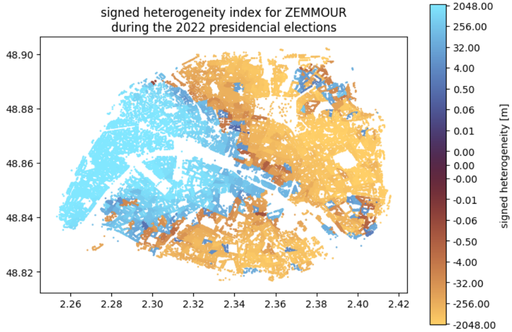
Candidate	OT-Index D_n	Score (%)
Macron	519 meters	35,49
Mélenchon	980 meters	29,93
Zemmour	1005 meters	8,2
Le Pen	295 meters	5,39

xavier.bacon@ens-paris-saclay.fr

A Local OT-Index

Local OT-Index

$$\omega_n(x) = \frac{1}{2\bar{\alpha}(x)} \langle C(x,\cdot) \,|\, \gamma_n^*(\cdot,x) - \gamma_n^*(x,\cdot) \rangle$$



- Optimal transference plans γ_n^* defines a flow between pairs of polling stations.
- Readily interpretable:
  C in m $\Longrightarrow \omega_n$ and D_n in m
- French 2022 Presidential results of 4 candidates (among 12) *(Paris area)*:

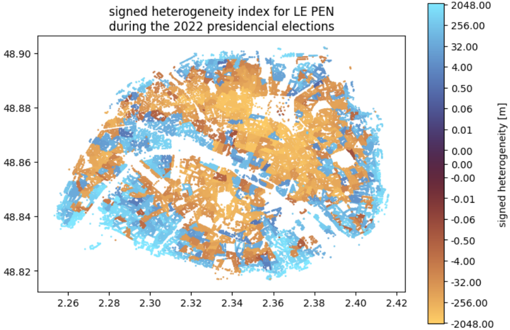
Candidate	OT-Index D_n	Score (%)
Macron	519 meters	35,49
Mélenchon	980 meters	29,93
Zemmour	1005 meters	8,2
Le Pen	295 meters	5,39

xavier.bacon@ens-paris-saclay.fr

A Local OT-Index

Local OT-Index

$$\omega_n(x) = \frac{1}{2\bar{\alpha}(x)} \langle C(x,\cdot) \mid \gamma_n^*(\cdot,x) - \gamma_n^*(x,\cdot) \rangle$$

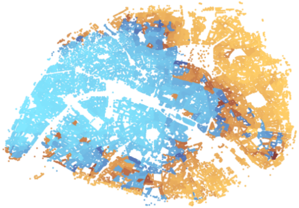


- Optimal transference plans γ_n^* defines a flow between pairs of polling stations.
- Readily interpretable:
  C in m $\Longrightarrow$ ω_n and D_n in m
- French 2022 Presidential results of 4 candidates (among 12) *(Paris area)*:

Candidate	OT-Index D_n	Score (%)
Macron	519 meters	35,49
Mélenchon	980 meters	29,93
Zemmour	1005 meters	8,2
Le Pen	295 meters	5,39

xavier.bacon@ens-paris-saclay.fr

Miscellaneous



- Future works: focus on special costs: concave (induces a few large displacements) or convex (induces a lot of small displacements).
- Code available: https://github.com/jolatechno/ot-heterogeneity
-

Thank you for your attention!

xavier.bacon@ens-paris-saclay.fr