

# Analyse spatiale

## Statistiques spatiales

Juste Raimbault<sup>1,2,3,4,\*</sup>

\* `juste.raimbault@ign.fr`

<sup>1</sup>LASTIG, Univ Gustave Eiffel, IGN-ENSG

<sup>2</sup>CASA, UCL

<sup>3</sup>UPS CNRS 3611 ISC-PIF

<sup>4</sup>UMR CNRS 8504 Géographie-cités

Cours Analyse Spatiale - DeSIGeo

01/03/2022

## 1 Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique

## 2 Coordonnées et distances

## 3 Analyse d'un semis de points

## 4 Mesures de densité et concentration

- Densité
- Concentration

## 5 Corrélation spatiale

## 6 Régression géographique pondérée

## 7 Pour aller plus loin

## 1 Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique

## 2 Coordonnées et distances

## 3 Analyse d'un semis de points

## 4 Mesures de densité et concentration

- Densité
- Concentration

## 5 Corrélation spatiale

## 6 Régression géographique pondérée

## 7 Pour aller plus loin

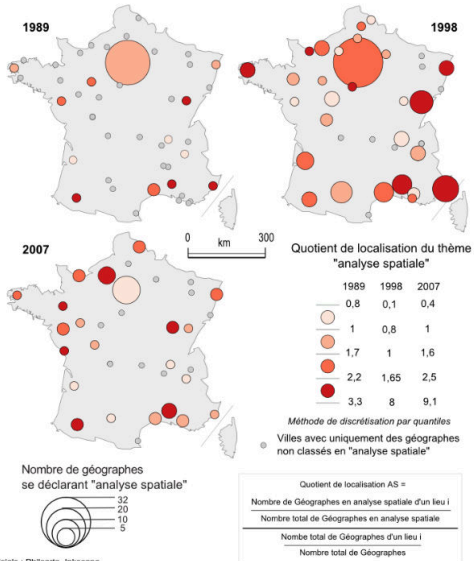
*“Etude formalisée de la configuration et des propriétés de l'espace produit et vécu par les sociétés humaines” [Pumain and Saint-Julien, 2010]*

→ **Théories et méthodes** pour l'analyse et la mesure des effets de la situation géographique

- Décrire les configurations de certains objets: naturels (sources, feux, séismes, ...), des villes, des aménités, des infrastructures, ...
- Synthétiser ("modéliser") leur structure spatiale: semis concentrés, réguliers, aléatoires, organisation centre-périphérie, réseau, structure hiérarchique, ...
- Relier leur localisation relative à leur caractéristiques *en partie*
- Inclure les positions relatives, les effets de voisinage et la distance géographique dans les analyses quantitatives.

# Analyse spatiale en géographie

**Géographes se classant "analyse spatiale"**  
Répartition spatiale en 1989, 1998 et 2007 (effectifs et quotient de localisation)  
source : Répertoire des géographes français



[Cuyala, 2014]

- objectif explicatif, principalement par la dimension “horizontale” des interactions (“vertical”: sociétés-environnement)
- relations entre lieux engendrées par les relations entre acteurs sociaux spatialisés
- hypothèse d'une certaine autonomie et spécificité de la dimension spatiale de l'organisation sociale (interférant avec les dimensions culturelle, environnementale, politique, économique, ...)
- lien entre formes et processus: existence de régularités dans la morphogenèse des systèmes géographique, indépendantes de la variété des milieux physiques
- ces régularités sont intégrées dans des modèles pour expliquer le fonctionnement et l'évolution des systèmes spatiaux

## 1 Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique

## 2 Coordonnées et distances

## 3 Analyse d'un semis de points

## 4 Mesures de densité et concentration

- Densité
- Concentration

## 5 Corrélation spatiale

## 6 Régression géographique pondérée

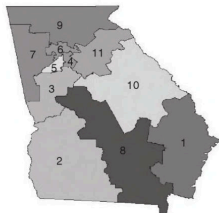
## 7 Pour aller plus loin



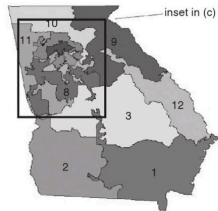
*Difficultés récurrentes lors de la prise en compte de l'espace :*

- *Modifiable Areal Unit Problem*
- Echelle des objets et processus
- Non-stationarité spatiale
- Données incertaines et/ou bruitées
- Général et particulier

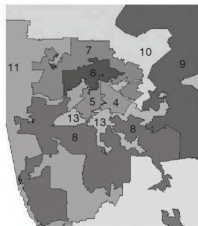
# Modifiable Areal Unit Problem (MAUP)



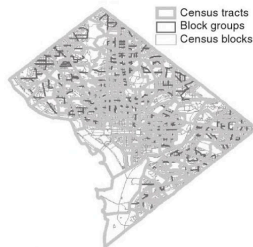
(a)



(b)

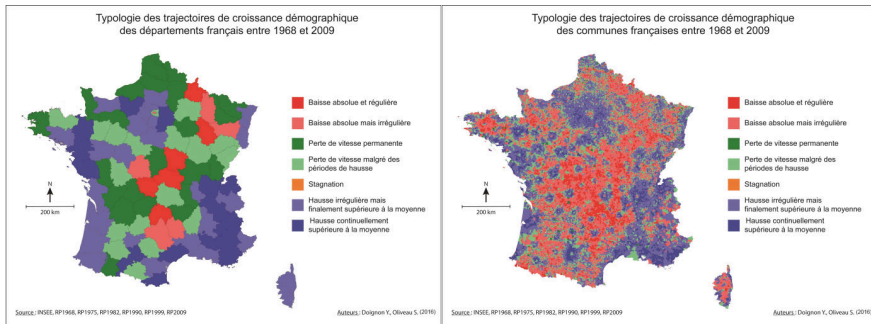


(c)



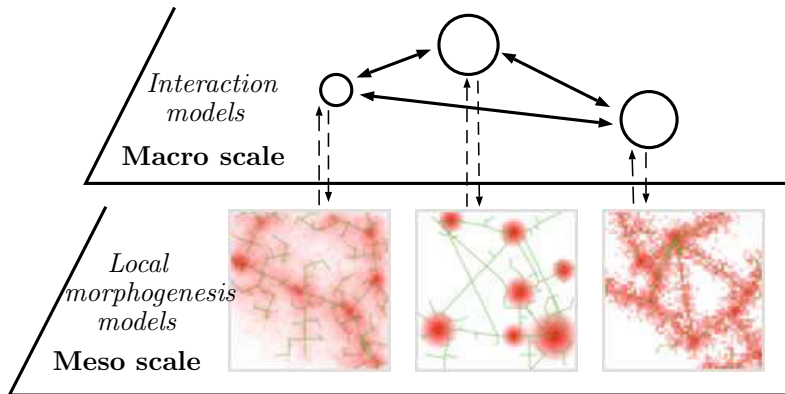
(d)

# MAUP : exemple



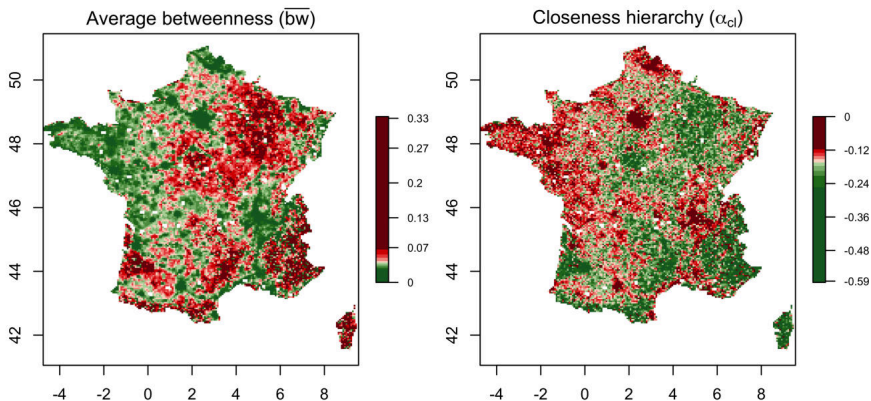
[Oliveau and Doignon, 2016]

*Processus spécifiques aux échelles* [Raimbault, 2021]



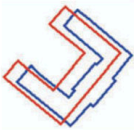
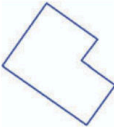
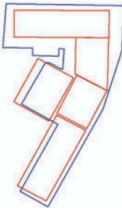
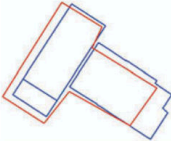
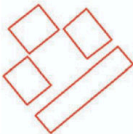

# Non-stationnarité spatiale

*Exemple : variations des indicateurs topologiques du réseau routier*  
[Raimbault, 2019]



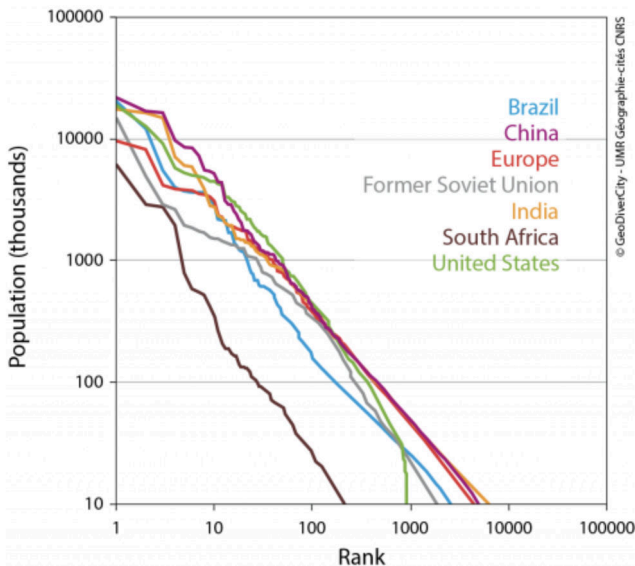
# Données incertaines ou bruitées

## Qualité des données dans OpenStreetMap [Fan et al., 2014]

Relation	1:1	1:0	1:n
Illustration			
Relation	n:1	0:1	n:m
Illustration			

# Général et particulier

*Hiéarchie urbaine comme une propriété universelle des systèmes urbains*  
[Pumain et al., 2015]



## 1 Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique

## 2 Coordonnées et distances

## 3 Analyse d'un semis de points

## 4 Mesures de densité et concentration

- Densité
- Concentration

## 5 Corrélation spatiale

## 6 Régression géographique pondérée

## 7 Pour aller plus loin



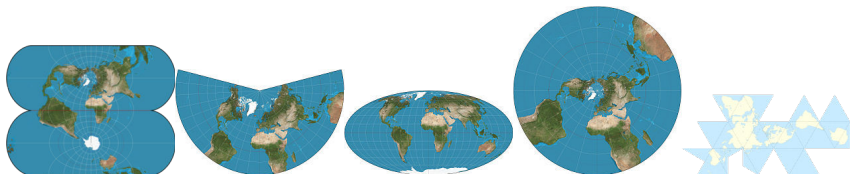
Coordonnées d'un point sur la surface terrestre déterminées par :

- **Système géodésique de référence** : RGF93, ETRS89, WGS84
- **Ellipsoïde** : Clarke1880, GRS 1980
- **Système de projection** : Mercator, Lambert, Mollweide, etc.

# Projections

Exemples de systèmes de projection :

- Universal Transverse Mercator : cylindrique conforme (conserve angles)
- Lambert (II, 93) : conique conforme
- Mollweide : pseudo-cylindrique équivalente (conserve surfaces)
- Azimutale, Fuller (polyhédrale) ...



**Formats raster** : geotif, images (avec fichiers annexes de référencement), formats compressés, ...

**Formats vecteur** :

- shapefile (ESRI) : binaire, simple couche
- geodatabase (ESRI) : binaire, multiples couches
- GeoJSON: JSON, simple couche
- GML (Open Geospatial Consortium), KML (Google): XML, simple
- Geopackage (OGC): Sqlite, multiple
- PostGIS, Spatialite : extensions SIG aux systèmes de bases de données



## Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique



## 2 Coordonnées et distances



## Analyse d'un semis de points



## Mesures de densité et concentration

- Densité
- Concentration



## Corrélation spatiale



## Régression géographique pondérée



## Pour aller plus loin

**Coordonnées projetées :**  $(\lambda, \phi, r) \rightarrow (x, y, z)$

**Analyse spatiale en pratique :** données spatiales en 2D  $(x, y)$

**Définition :**  $E$  ensemble non vide muni de  $d : E \times E \rightarrow \mathbb{R}_+$  tel que  $d$  soit une distance, i.e. vérifiant les trois propriétés :

- Symétrie :  $\forall x, y \in E, d(x, y) = d(y, x)$
- Séparation :  $\forall x, y \in E, d(x, y) = 0 \iff x = y$
- Inégalité triangulaire :  $\forall x, y, z \in E, d(x, y) \leq d(x, z) + d(z, y)$

**Topologie :** boules ouvertes, ouverts, fermés

**Définition :**  $E$  espace vectoriel sur  $\mathbb{K}$  muni d'une norme  $N$  avec :

- Séparation :  $\forall x \in E, N(x) = 0 \implies x = 0_E$
- homogénéité :  $\forall (\lambda, x) \in \mathbb{K} \times E, N(\lambda x) = |\lambda| N(X)$
- Sous-additivité :  $\forall x, y \in E, N(x + y) \leq N(x) + N(Y)$

**Exercice :** Définir une distance sur un espace vectoriel normé

Pour  $x, y \in \mathbb{R}^n$

$$d_p(x, y) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

**Question :** inégalité triangulaire ?

**Application pratique :** distance euclidienne ( $p = 2$ ), distance de Manhattan ( $p = 1$ )



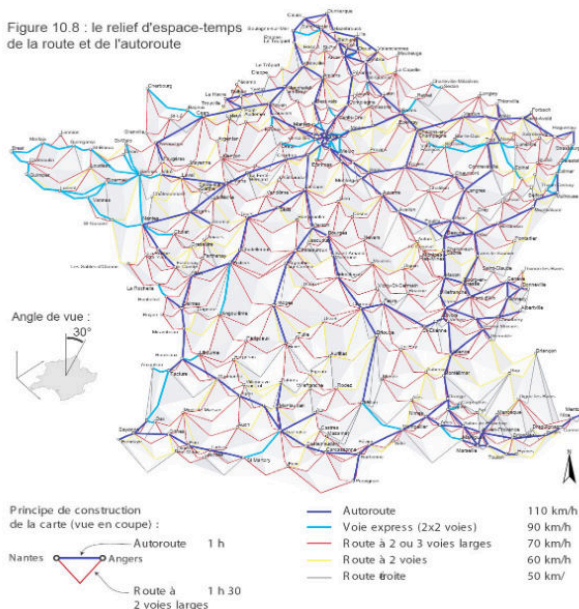
*Great-circle distance* : à partir des coordonnées géographiques  $(\lambda, \phi)$

$$d_{ij} = R \cdot \arccos(\sin \phi_i \sin \phi_j + \cos \phi_i \cos \phi_j \cos(\lambda_i - \lambda_j))$$

**Exercice** : à  $\lambda$  (longitude) constant, à partir de quel ordre de grandeur de  $\delta\phi = |\phi_i - \phi_j|$  une distance euclidienne sur des coordonnées non projetées devient incohérente ?

# Distance-temps et réseaux

Figure 10.8 : le relief d'espace-temps  
de la route et de l'autoroute





## Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique



## Coordonnées et distances



## 3 Analyse d'un semis de points



## Mesures de densité et concentration

- Densité
- Concentration



## Corrélation spatiale



## Régression géographique pondérée



## Pour aller plus loin

Barycentre du nuage de points

$$\vec{\mu} = \frac{\sum_{i=1}^n w_i \vec{x}_i}{\sum_{i=1}^n w_i}$$

**Exercice :** *Minimise la distance au carré à tous les points*

Point minimisant la somme des distances à tous les points

$$\vec{m} = \operatorname{argmin}_{\vec{x}} \sum_{i=1}^n d(\vec{x}_i, \vec{x})$$

# Point médian pondéré

Point minimisant la somme des distances aux éléments du stock  $w_i$  aux points  $\vec{x}_i$

$$\vec{m} = \operatorname{argmin}_{\vec{x}} \frac{\sum_{i=1}^n w_i \cdot d(\vec{x}_i, \vec{x})}{\sum_{i=1}^n w_i}$$

**Exercice 1 :** existence et unicité du point médian

**Exercice 2 :** localisation du point médian en termes de quantiles

**Exercice 3 :** définition dans le cas d'un champ continu?

Distance-type pondérée

$$\sigma = \sqrt{\frac{\sum_{i=1}^n w_i [(x_i - x_\mu)^2 + (y_i - y_\mu)^2]}{\sum_{i=1}^n w_i}}$$

*Mesure de l'étendue spatiale de la distribution autour du barycentre*

**Première partie du TP en R :** calculer/représenter l'évolution des points moyens, médian, dispersion, pour la population des aires urbaines françaises entre 1830 et 2000

**Données :** aires urbaines dans le temps long, Géographie-cités, [Pumain and Riandey, 1986]





## Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique



## Coordonnées et distances



## Analyse d'un semis de points



## 4 Mesures de densité et concentration

- Densité
- Concentration



## Corrélation spatiale



## Régression géographique pondérée



## Pour aller plus loin

## 1 Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique

## 2 Coordonnées et distances

## 3 Analyse d'un semis de points

## 4 Mesures de densité et concentration

- Densité
- Concentration

## 5 Corrélation spatiale

## 6 Régression géographique pondérée

## 7 Pour aller plus loin

Indice pour quantifier un niveau d'agrégation d'un stock  $x_i$   
[Le Néchet, 2011]

$$\bar{d} = \sum_{i,j} \frac{P_i P_j}{(\sum_k P_k)^2} d_{ij}$$

*Généralisation : fonction croissante arbitraire de la distance*

# Distance au plus proche voisin

Comparaison de la distance moyenne au plus proche voisin à la distance attendue dans le cas aléatoire, pour une densité globale  $\lambda$

$$v = 2\sqrt{\lambda} \cdot \frac{1}{N} \sum_i \min_{j \neq i} d_{ij}$$

Rapport entre densité locale et globale pour tester la concentration de points.

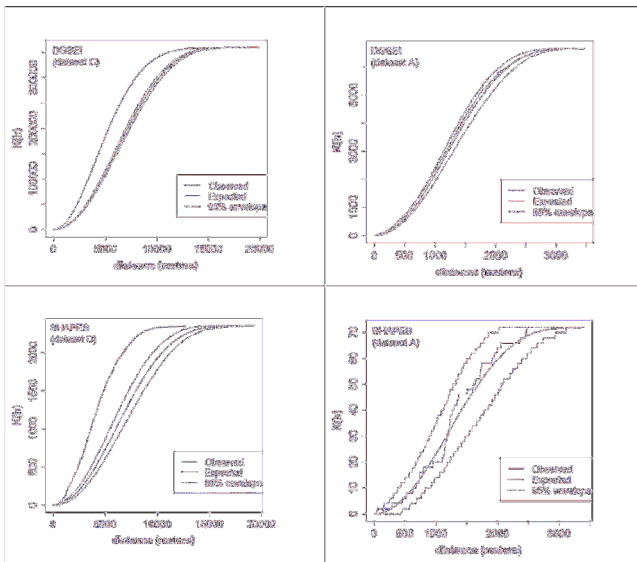
$$K(r) = \frac{1}{\lambda} \sum_{i \neq j} \frac{\mathbb{1}(d_{ij} < r)}{n}$$

avec  $\lambda$  densité globale de points

→ test statistique pour comparer  $K(r)$  et le nombre de points attendus avec une distribution aléatoire  $\pi \cdot r^2$

# Exemple

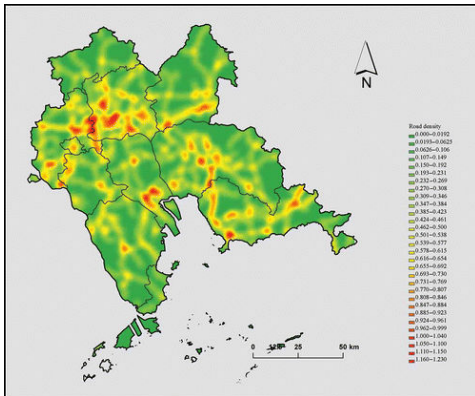
*Application à la concentration des accidents cyclistes*  
[Vandenbulcke et al., 2017]



# Kernel Density Estimation

Estimation d'une densité continue à partir du nuage de point, comme une agrégation de noyaux (Rosenblatt-Parzen) de distance caractéristique  $h$

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{d(x, x_i)}{h}\right)$$



## 1 Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique

## 2 Coordonnées et distances

## 3 Analyse d'un semis de points

## 4 Mesures de densité et concentration

- Densité
- Concentration

## 5 Corrélation spatiale

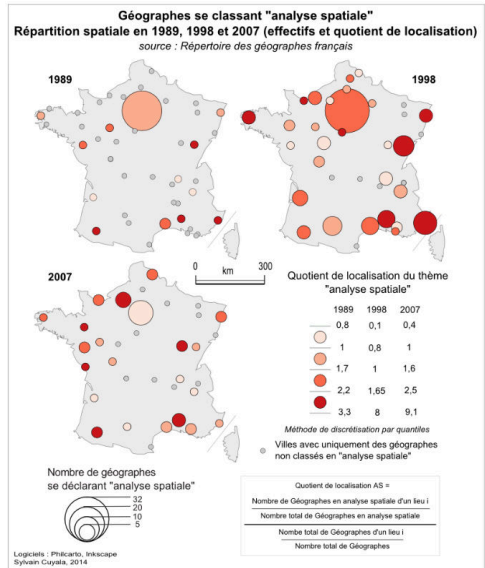
## 6 Régression géographique pondérée

## 7 Pour aller plus loin



Rapport entre la part locale d'une activité  $j$  et sa part globale, pour l'unité spatiale  $i$

$$S_{ij} = \frac{A_{ij} / \sum_j A_{ij}}{\sum_i A_{ij} / \sum_{i,j} A_{i,j}}$$



Indice de dissimilarité entre des stocks (utilisé pour quantifier des inégalités)

$$G = \frac{\sum_i \sum_j |x_i - x_j|}{2n \sum_i x_i}$$

*Indice non spatialisé*

Indice pour quantifier la régularité de la distribution statistique

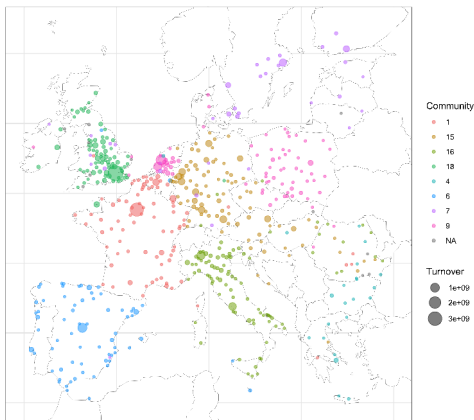
$$H(x) = - \sum_{i=1}^n p_i \log(p_i)$$

*Indice non spatialisé*

# Grouper des points dans l'espace

Exemples d'algorithmes pour classifier des points spatiaux :

- k-means
- DBSCAN (basé sur la densité) et extension
- Détection de communautés dans des réseaux de proximité



*Détection de communautés dans un réseau de liens industriels (filiales) [Raimbault et al., 2020]*



## Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique



## Coordonnées et distances



## Analyse d'un semis de points



## Mesures de densité et concentration

- Densité
- Concentration



## Corrélation spatiale



## Régression géographique pondérée



## Pour aller plus loin

# Autocorrélation: Indice de Moran

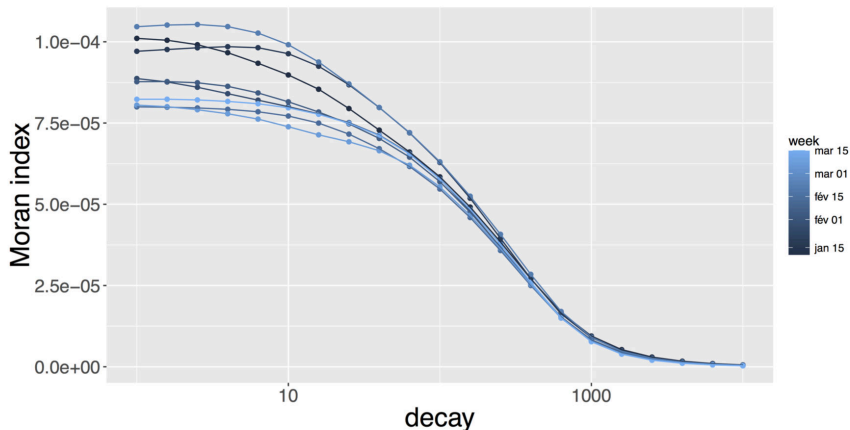
Indice d'autocorrélation spatiale pour un champ  $x_i$  étant donné des poids spatiaux  $w_i$  permettant de définir un voisinage

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \times \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

→ positif pour des configurations agrégées, nul pour des configurations aléatoires, négatifs pour des configurations “échiquier”

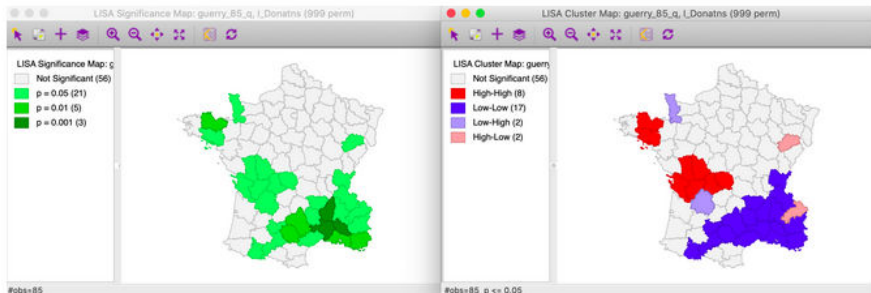
# Application : régimes d'autocorrélation

*Extraction des échelles typiques des processus dans un système multi-échelle [Bergeaud and Raimbault, 2020]*



# LISA : indices localisés

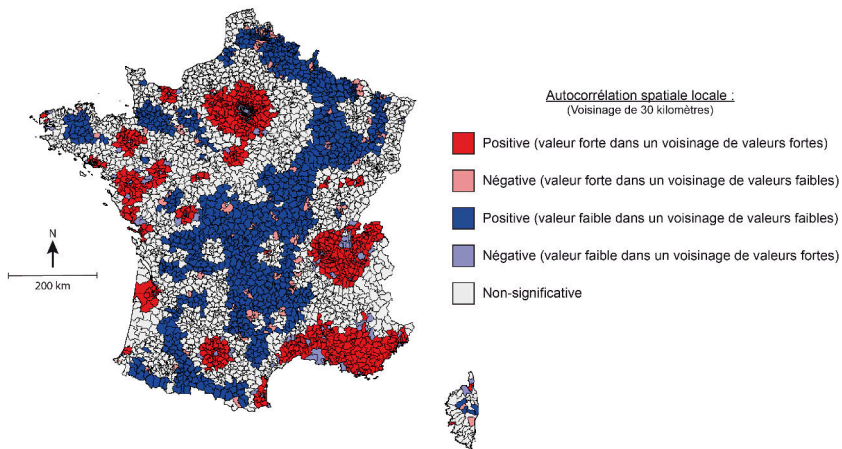
Statistique locale d'autocorrélation:  $I_i \propto (x_i - \bar{x}) \cdot \sum_j w_{ij}(x_j - \bar{x})$



*Exemple d'analyse LISA intégrée au logiciel Geoda : significativité statistique et clusters de corrélation [Anselin et al., 2010]*



## L'autocorrélation spatiale locale de la croissance démographique des cantons français entre 1968 et 2009



Source : INSEE, RP1968, RP1975, RP1982, RP1990, RP1999, RP2009  
Logiciel : GeoDa, 10 000 permutations, significativité à 95%

Auteurs : Doignon Y., Oliveau S. (2014)

**Deuxième partie du TP en R :** application des indices de densité/concentration, corrélation spatiale

**Données :** différents types d'aménités obtenus d'OpenStreetMap  
<https://demo.openstreetmap.fr/map>



## Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique



## Coordonnées et distances



## Analyse d'un semis de points



## Mesures de densité et concentration

- Densité
- Concentration



## Corrélation spatiale



## Régression géographique pondérée



## Pour aller plus loin

# Régression géographique pondérée

*Comment inclure des effets de voisinage et prendre en compte la non-stationnarité spatiale dans des modèles statistiques ?*

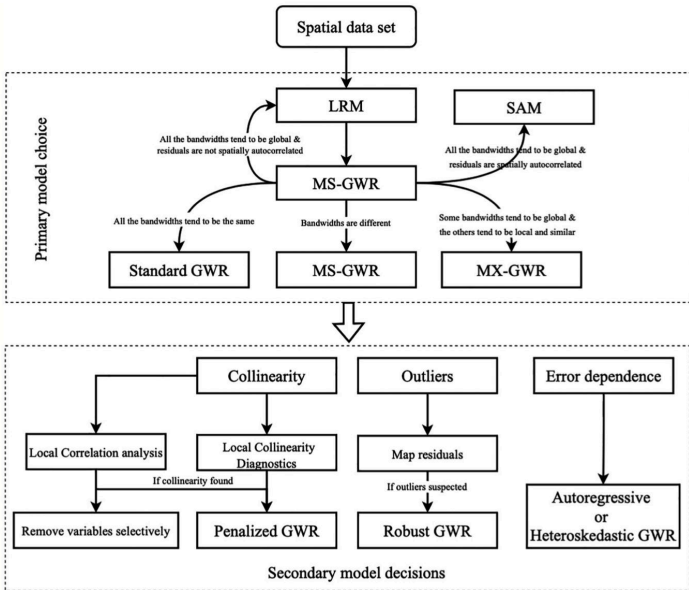
Modèle GWR basique pour les variables  $y_i$  aux positions  $\vec{u}_i$  et variables explicatives  $x_{ik}$

$$y_i = \beta_0(\vec{u}_i) + \sum_k \beta_k(\vec{u}_i) x_{ik} + \varepsilon_i$$

avec les observations pondérées par un poids spatial  $w_i(r)$  en fonction de la distance à  $\vec{u}_i$

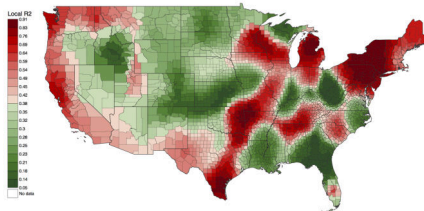
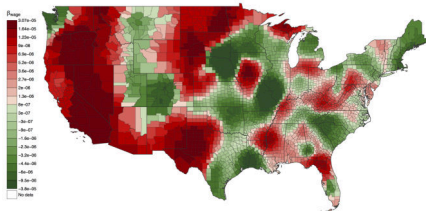
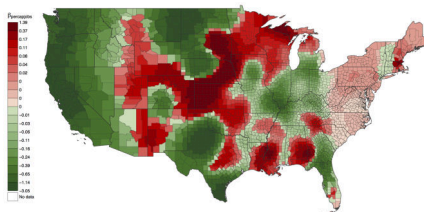
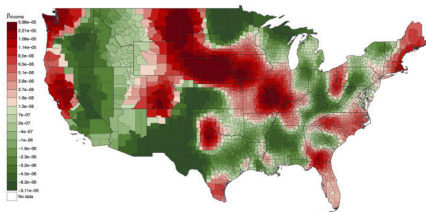
- Différents kernels pour les poids (gaussien, exponentiel, puissance, bisquare)
- Estimation de la taille de kernel optimale par optimisation de l'AIC par exemple

# Méthode complète d'application



# Exemple

*Application avec échelle spatiale optimale (critère AIC) aux déterminants des prix du carburants aux US [Bergeaud and Raimbault, 2020]*



**Troisième partie du TP en R : analyse GWR**

**Données :** caractéristiques socio-économiques et prix du carburant pour les Comtés au US



## Introduction

- Cadre théorique
- Particularités des phénomènes spatiaux
- Information géographique



## Coordonnées et distances



## Analyse d'un semis de points



## Mesures de densité et concentration

- Densité
- Concentration



## Corrélation spatiale



## Régression géographique pondérée



## Pour aller plus loin



- Extrapolation, kriging
- Processus de points spatiaux
- Modèles statistiques avec auto-corrélation spatiale
- Apprentissage supervisé, non-supervisé, semi-supervisé
- Modèles de simulation
- ...



Anselin, L., Syabri, I., and Kho, Y. (2010).

Geoda: an introduction to spatial data analysis.

*In Handbook of applied spatial analysis*, pages 73–89. Springer.



Bergeaud, A. and Raimbault, J. (2020).

An empirical analysis of the spatial variability of fuel prices in the united states.

*Transportation Research Part A: Policy and Practice*, 132:131–143.



Cai, X., Wu, Z., and Cheng, J. (2013).

Using kernel density estimation to assess the spatial pattern of road density and its impact on landscape fragmentation.

*International Journal of Geographical Information Science*, 27(2):222–230.



Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., et al. (2021).

A route map for successful applications of geographically weighted regression.

*Geographical Analysis.*



Cuyala, S. (2014).

*Analyse spatio-temporelle d'un mouvement scientifique. L'exemple de la géographie théorique et quantitative européenne francophone.*

PhD thesis, Université Paris 1 Panthéon-Sorbonne.



Fan, H., Zipf, A., Fu, Q., and Neis, P. (2014).

Quality assessment for building footprints data on openstreetmap.

*International Journal of Geographical Information Science*,  
28(4):700–719.



Le Néchet, F. (2011).

Consommation d'énergie et mobilité quotidienne selon la configuration des densités dans 34 villes européennes.

*Cybergeog: European Journal of Geography.*



l'Hostis, A. (2003).

Théorie des graphes et représentations des distances: chronocartes et autres représentations.



Oliveau, S. and Doignon, Y. (2016).

La diagonale se vide? analyse spatiale exploratoire des décroissances démographiques en france métropolitaine depuis 50 ans.

*Cybergeog: European Journal of Geography.*



Pumain, D. and Riandey, B. (1986).

Le fichier de l'ined: "urbanisation de la france".

*Espace Populations Sociétés*, 11(2):269–278.



Pumain, D. and Saint-Julien, T. (2010).

*Analyse spatiale: les localisations.*

Armand Colin.



Pumain, D., Swerts, E., Cottineau, C., Vacchiani-Marcuzzo, C., Ignazzi, C. A., Bretagnolle, A., Delisle, F., Cura, R., Lizzi, L., and Baffi, S. (2015).

Multilevel comparison of large urban systems.

*Cybergeo: European Journal of Geography.*



Raimbault, J. (2019).

An urban morphogenesis model capturing interactions between networks and territories.

In *The mathematics of urban morphology*, pages 383–409. Springer.



Raimbault, J. (2021).

Strong coupling between scales in a multi-scalar model of urban dynamics.

*arXiv preprint arXiv:2101.12725.*



Raimbault, J., Zdanowska, N., and Arcaute, E. (2020).

Modeling growth of urban firm networks.

*arXiv preprint arXiv:2009.05528.*



Vandenbulcke, G., Int Panis, L., and Thomas, I. (2017).

On the location of reported and unreported cycling accidents: A spatial network analysis for brussels.

*Cybergeo: European Journal of Geography.*



Wong, D. W. (2004).

The modifiable areal unit problem (maup).

*In WorldMinds: Geographical perspectives on 100 problems*, pages 571–575. Springer.