

Statistique spatiale avec R

Séminaire ODR

Thibault LAURENT
thibault.laurent@tse-fr.eu

Toulouse School of Economics, CNRS

4 Septembre 2018

1 Introduction

2 Analyse exploratoire de données spatiales

3 Outils spécifiques à la statistique spatiale

4 Modélisation spatiale

5 Méthodes d'interpolation de données surfaciques

Remerciements et références

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Remerciements :

- Christine Thomas-Agnan et l'équipe Statistique de TSE.

Références :

- A. Baddeley (2010), "Analysing spatial point patterns in R", CSIRO.
- R. Bivand, E. Pebesma et V. Gómez-Rubio (2013), "Applied spatial data analysis with R", Springer-Verlag, 2^e édition.
- P.J. Diggle et P.J. Ribeiro (2007), "Model-based geostatistics", Springer-Verlag.
- J. LeSage et R.K. Pace (2009), "Introduction to spatial econometrics", Chapman et Hall.
- C. Thomas-Agnan (2018), "Analyse statistique des données spatiales", UT1-Capitole.

- Un tableau de données :

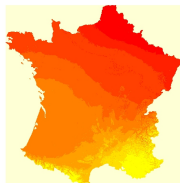
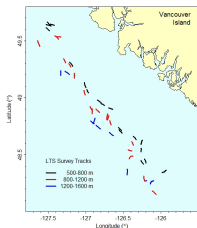
| Individus | Variables | | |
|-----------|-----------|-----|----------|
| | 1 | ... | p |
| 1 | x_{11} | ... | x_{1p} |
| ⋮ | ⋮ | ... | ⋮ |
| n | x_{n1} | ... | x_{np} |

- Spécificité en statistique spatiale : chaque individu est associé à une entité géographique.
- Objectifs : utiliser à la fois l'information statistique et géographique.

⇒ Apparition de nouvelles méthodes statistiques qui vont dépendre de la nature "spatiale" des données.

Entité géographique : le point de vue de l'informaticien

Représentation des données sous 2 modes : vecteurs (polygone, point ou ligne brisée) ou raster (pixel).



Entité géographique : le point de vue du statisticien

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Soit D le domaine d'étude dans l'espace et s la localisation géographique. Cressie (1993) suggère la classification des données spatiales suivante :

- Données de type géostatistique : s varie continuellement dans D , mais les observations sont faites en un nombre fini de points déterministes s_i de D .
- Données de type économétrie spatiale : comme en géostatistique, la position n'est pas aléatoire. L'unité géographique est de type surfacique.
- Données de type semis de points : la localisation de la donnée est modélisée comme aléatoire.

Exemple de données de type géostatistique

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Domaines d'applications : écologie, environnement, climatologie, industrie pétrolière, etc.



Etude de la contamination du sol sur une zone inondable. Données observées en chaque point : concentration de zinc, cuivre, etc.

Entité géographique : le point de vue du statisticien

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Soit D le domaine d'étude dans l'espace et s la localisation géographique. Cressie (1993) suggère la classification des données spatiales suivante :

- Données de type géostatistique : s varie continuellement dans D , mais les observations sont faites en un nombre fini de points déterministes s_j de D .
- Données de type économétrie spatiale : comme en géostatistique, la position n'est pas aléatoire. L'unité géographique est de type surfacique.
- Données de type semis de points : la localisation de la donnée est modélisée comme aléatoire.

Exemple de données de type économétrie spatiale

Statistique spatiale avec R

T. Laurent

Introduction

Analyse exploratoire

Cartographie

Graphiques interactifs

Visualisation

Outils spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation spatiale

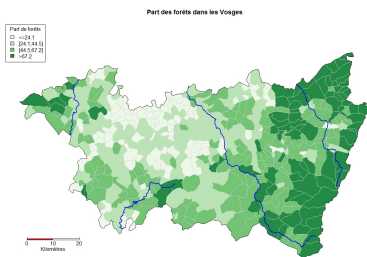
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Domaines d'applications : statistique publique, secteur immobilier, etc.



Mesurer l'impact de la nature des sols sur la qualité des eaux (ANR ModULand). Données agrégées au niveau de la commune.

Entité géographique : le point de vue du statisticien

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Soit D le domaine d'étude dans l'espace et s la localisation géographique. Cressie (1993) suggère la classification des données spatiales suivante :

- Données de type géostatistique : s varie continuellement dans D , mais les observations sont faites en un nombre fini de points déterministes s_j de D .
- Données de type économétrie spatiale : comme en géostatistique, la position n'est pas aléatoire. L'unité géographique est de type surfacique.
- Données de type semis de points : la localisation de la donnée est modélisée comme aléatoire.

Exemple de données de type semis de points

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

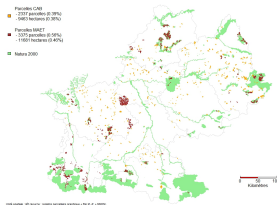
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Domaines d'applications : foresterie, épidémiologie, sismologie, écologie, agriculture, etc.



Etude de la concentration spatiale des exploitations agricoles converties à l'Agriculture Biologique (ANR ModULand).

Autres types de données spatiales : données spatio-temporelles

Statistique spatiale avec R

T. Laurent

Introduction

Analyse exploratoire

Cartographie

Graphiques interactifs

Visualisation

Outils spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation spatiale

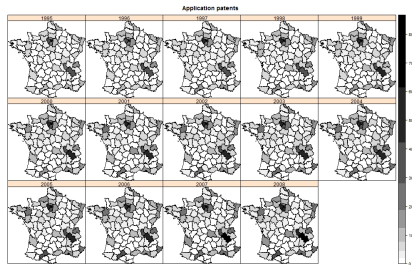
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Domaines d'applications : économie, écologie, etc.



Evaluer l'impact des dépenses en R&D sur l'innovation avec des données de panel (Laurent et Moussa, 2015). Package **spacetime** (Pebezma *et al.*, 2012).

Autres types de données spatiales : données de flux

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

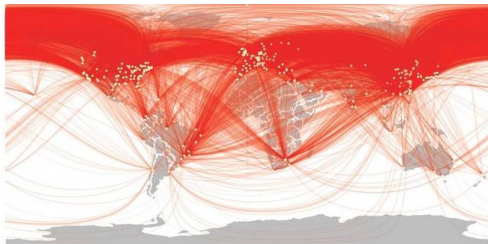
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Domaines d'applications : transport aérien, geomarketing, etc.



Etude de la dépendance spatiale des liaisons aériennes entre aéroports
(travail en collaboration avec Airbus).

Les logiciels disponibles

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Cartographie :

- QGIS : <http://www.qgis.org/fr/site/>

Statistique spatiale :

- environ 200 packages R (Septembre 2018) dont 17 installés avec le logiciel. R. Bivand référence la liste des packages sur : <http://cran.r-project.org/web/views/Spatial.html>
- Matlab et la toolbox d'économétrie spatiale écrite par James P. LeSage (<http://www.spatial-econometrics.com/>)
- PySAL: librairie de fonctions sur l'analyse spatiale écrite en Python (<https://pysal.readthedocs.io/en/latest/>)

1 Introduction

2 Analyse exploratoire de données spatiales

- Cartographie
- Graphiques interactifs
- Autres outils de visualisation

3 Outils spécifiques à la statistique spatiale

4 Modélisation spatiale

5 Méthodes d'interpolation de données surfaciques

Qu'est ce qu'un "Coordinate Reference System" (CRS) ?

- Choix d'une représentation de la terre : ellipsoïde Globale (WGS 84, Clarke 1880, IAG GRS80, etc.) ou Locale (NTF, Nouvelle Triangulation de la France).
- Choix du système de coordonnées géographiques (points de références pour longitude, latitude, etc.) pour exprimer les entités géographiques en longitude et latitude (λ, ϕ) (en degrés/minutes/secondes ou degrés décimaux).
- Choix d'un système de projection des coordonnées pour minimiser les déformations, pouvoir calculer des distances kilométriques, etc.

Exemple

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

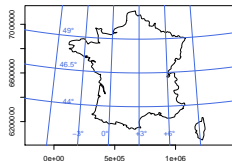
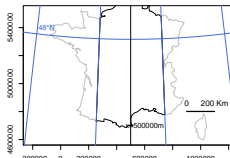
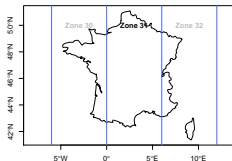
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Selon la source des données (IGN, Insee, etc.), les CRS
(Coordinate Reference System) peuvent varier...



Packages R pour la cartographie

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique
Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique
Econométrie spatiale

Semis de points

Interpolation

- **rgdal** (faisant appel aux bibliothèques GDAL et PROJ.4): import/export de fichiers de données spatiales, conversion de coordonnées spatiales d'un CRS vers un autre.
- **sp** et **sf** (package récent qui adopte la norme "Simple Features" et se rapproche d'une utilisation Tidyverse) : définition et méthodes de classes d'objet `Spatial`.
- **raster** : images de type raster.
- **rgeos** : fonctions qui permettent de manipuler des entités spatiales (union/intersection de polygones, buffer, calcul d'aires, etc.)

Dessiner des cartes

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs
Visualisation

Outils
spécifiques

Géostatistique
Econométrie spatiale
Semis de points

Modélisation
spatiale

Géostatistique
Econométrie spatiale
Semis de points

Interpolation

- Représenter l'information spatiale (polygone, point, ligne, pixel) et statistique (qualitative ou quantitative) en même temps : packages **sp** et **sf**.
- Traitement particulier pour les données de type raster : package **raster** permet d'optimiser les temps de calcul.
- Le statisticien peut influencer la lecture de la carte par le choix de la représentation (carte choroplètes ou bubbles), le choix de la méthode de discrétisation (méthode des quantiles, des K -moyennes, classes d'amplitude égales, etc., package **classInt**), le choix des couleurs (package **RColorBrewer**).

Choix de la discrétisation

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

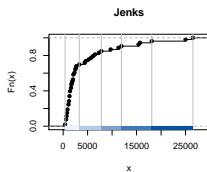
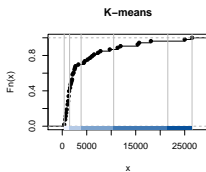
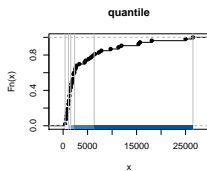
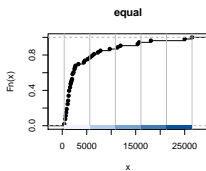
Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation



Choix des couleurs

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

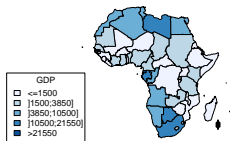
Géostatistique

Econométrie spatiale

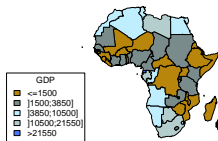
Semis de points

Interpolation

GDP per inhabitant



GDP per inhabitant



Représentation du PIB moyen par habitant dans les pays africains en 2013.
Source : CIA.

Carte choroplèthe VS Bubbles

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

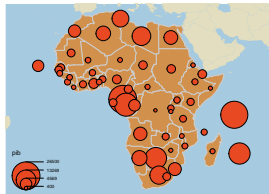
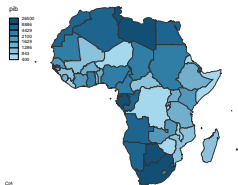
Modélisation
spatiale

Géostatistique

Econométrie spatiale

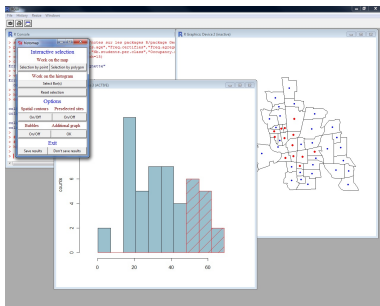
Semis de points

Interpolation



Graphiques interactifs (1)

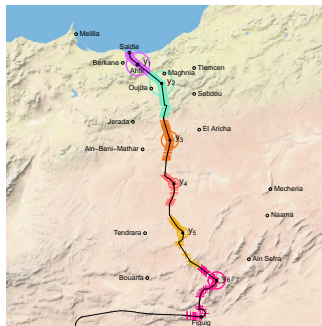
T. Laurent, A. Ruiz-Gazen et C. Thomas-Agnan (2012),
“**GeoXp**: an R package for Exploratory Spatial Data Analysis”,
Journal of Statistical Software.



GeoXp : une interface Tcl/Tk, une carte et un graphique. Téléchargeable depuis <https://github.com/cran/GeoXp>

Autres outils de visualisation (1)

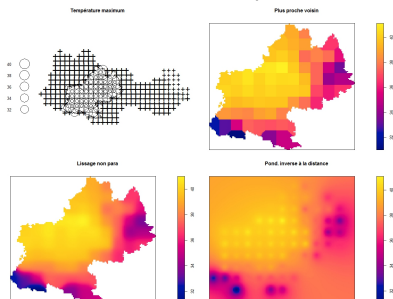
Utiliser des fonds de couverture aérienne, faire des animations :
packages **ggmap**, **OpenStreetMap**, **animation**, **shiny**, etc.



Exemple : positionnement optimal d'un nouveau poste de garde sur la frontière Maroc/Algérie (Daouia et Van Keilegom, 2018). Voir animation sur http://www.thibault.laurent.free.fr/code/daouia/daouia_VanKeilegom.html

Autres outils de visualisation (2)

Utiliser des méthodes d'interpolation ou lissage des données spatiales (packages **spatstat**, **akima**).



Interpolation/lissage de données météo 25×25 km sur l'ensemble de la région Midi-Pyrénées : méthode du plus proche voisin, lissage non paramétrique et lissage proportionnel à l'inverse de la distance. Projet dans le cadre de l'ANR ModULand (Chakir *et al.*, 2016).

1 Introduction

2 Analyse exploratoire de données spatiales

3 Outils spécifiques à la statistique spatiale

- Géostatistique
- Econométrie spatiale
- Semis de points

4 Modélisation spatiale

5 Méthodes d'interpolation de données surfaciques

Décomposition d'un champ aléatoire

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

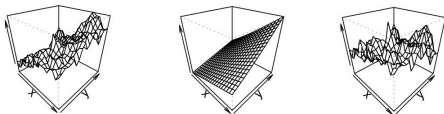
Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

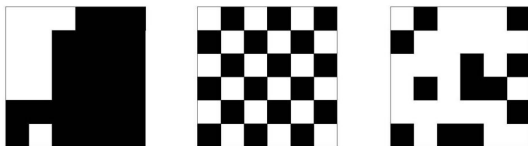


$$X_s = \mathbb{E}(X_s) + (X_s - \mathbb{E}(X_s))$$

- **Hétérogénéité spatiale** : la répartition marginale de X_s varie avec s ($\mathbb{E}(X_s)$ non constante).
- **Autocorrélation spatiale** : les variables X_s et X_t sont d'autant plus corrélées que la distance entre s et t est petite (existence du moment d'ordre 2; fonction d'autocovariance $R(s, t) = \text{Cov}(X_s, X_t)$).

Autocorrélation spatiale

Exemple lorsque X_s prend ses valeurs dans $\{0, 1\}$:



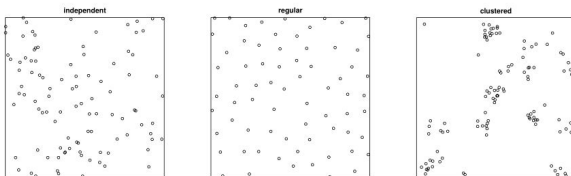
Autocorrélation spatiale : forte positive, forte négative, absence.

Outils pour apprécier l'autocorrélation spatiale :

- le **variogramme** (empirique) en géostatistique.
- le **diagramme de Moran** (+ tests) en économétrie spatiale. Nécessite la création d'une matrice de voisinage.

Homogénéité et interaction spatiale

Pour les **semis de points** :



- **Homogénéité** (notion d'ordre 1) : le nombre moyen de points par unité de surface est-il constant au travers du domaine ?
- **Interaction** (notion d'ordre 2) : y-a-t-il de l'attraction ou de la répulsion entre les points ?

Le variogramme

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

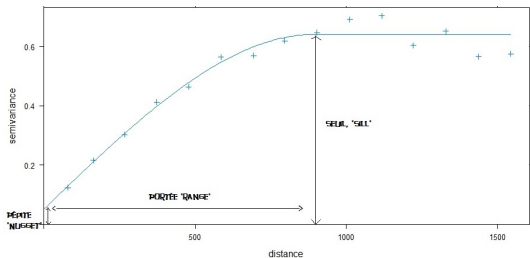
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Le variogramme d'un champ X_s vaut :
$$V(X_s, X_t) = \text{Var}\{X_s - X_t\} (=2\text{Var}X_s - 2R(s, t))$$
- Exemple ("variogramme dit exponentiel") :
$$V(u) = \tau^2 + \sigma^2\{1 - \exp(-\frac{u}{\phi})\}, \text{ avec } u = \|s - t\|$$



Effet de $\text{pédite}=\tau^2$, $\text{seuil}=\sigma^2$ ($\text{Var}X_s$), ϕ contrôle la portée.

Variogramme empirique

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

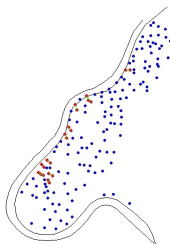
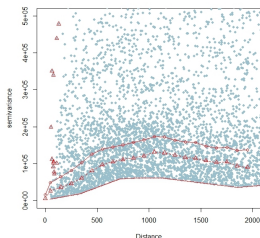
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- 1 Soit le nuage de points (u_{ij}, v_{ij}) où $u_{ij} = \|s_i - s_j\|$ et $v_{ij} = \frac{1}{2}(x_{s_i} - x_{s_j})^2$.
- 2 **Variogramme empirique** : pour différentes valeurs de u , on fait la moyenne des v_{ij} t.q. $u_{ij} \in [u - \frac{h}{2}; u + \frac{h}{2}]$.



Exemple de variogramme empirique + version robuste (Cressie, 1993) avec **GeoXp**. Choix de h : au moins 30 paires inclus dans $[u - \frac{h}{2}; u + \frac{h}{2}]$.

Choix d'un variogramme

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

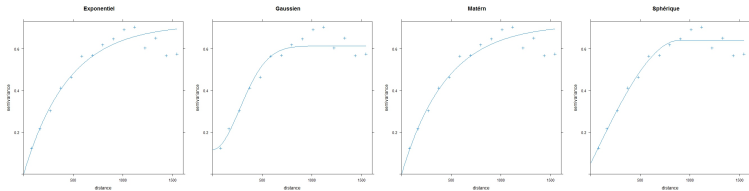
Géostatistique

Econométrie spatiale

Semis de points

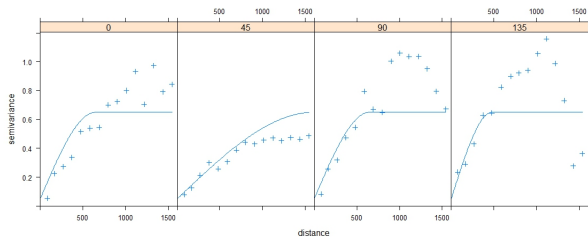
Interpolation

- Fonction `vgm` (package **gstat**) : choix de la forme du variogramme (exponentiel, sphérique, gaussien ou Matérn) + choix des valeurs initiales de l'effet de pépite, portée et seuil.
- Fonction `fit.variogram` : algorithme (Pebezma, 2004) qui calcule les paramètres optimaux pour que le variogramme s'ajuste le mieux (au sens des moindres-carrés) au variogramme empirique.



Variogramme directionnel

Cas **anisotrope** : le variogramme dépend de la direction de \vec{st} .



Exemple de variogramme directionnel avec le package **gstat** (fonction `variogram` avec option `alpha` pour définir la direction).

Voisins / Poids / Matrice W

Statistique spatiale avec R

T. Laurent

Introduction

Analyse exploratoire

Cartographie

Graphiques interactifs

Visualisation

Outils spécifiques

Géostatistique

Econométrie spatiale

Semis de points

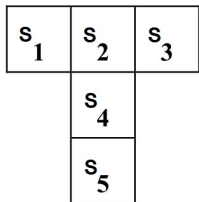
Modélisation spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation



Voisins de type
"rook"

$s_1 : s_2$

$s_2 : s_1, s_3, s_4$

$s_3 : s_2$

$s_4 : s_2, s_5$

$s_5 : s_4$

Poids de type
"row-normalized"

$s_1 : 1$

$s_2 : 0.33, 0.33, 0.33$

$s_3 : 1$

$s_4 : 0.5, 0.5$

$s_5 : 1$

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0.33 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Voisins / Poids / Matrice W

Statistique spatiale avec R

T. Laurent

Introduction

Analyse exploratoire

Cartographie

Graphiques interactifs

Visualisation

Outils spécifiques

Géostatistique

Econométrie spatiale

Semis de points

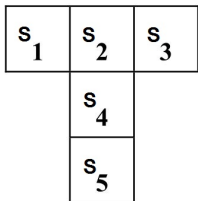
Modélisation spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation



Voisins de type
"queen"

s_1 : s_2 , s_4

s_2 : s_1 , s_3 , s_4

s_3 : s_2 , s_4

s_4 : s_1 , s_2 , s_3 , s_5

s_5 : s_4

Poids de type
"row-normalized"

s_1 : 0.5, 0.5

s_2 : 0.33, 0.33, 0.33

s_3 : 0.5, 0.5

s_4 : 0.2, 0.2, 0.2, 0.2

s_5 : 1

$$W = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 & 0 \\ 0.33 & 0 & 0.33 & 0.33 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0.2 & 0.2 & 0.2 & 0 & 0.2 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Matrices de voisinage avec R (package `spdep`)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- 1 On crée une liste de voisins (objet `nb`) :
 - Méthodes basées sur la géométrie des unités spatiales :
 - contiguïté des polygones : s_i et s_j sont voisins s'ils partagent une frontière commune (fonction `poly2nb`).
 - triangulation de Delaunay (fonction `tri2nb`).
 - Méthodes basées à partir d'un calcul de distance :
 - les k plus proches voisins (fonction `knn2nb`).
 - voisins si distance inférieure à une distance seuil c (fonction `dnearneigh`).
- 2 On choisit les poids à associer aux voisins (objet `listw`) : poids binaires, normalisés en ligne, etc.

Visualisation de la structure de voisinage sous R (1)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

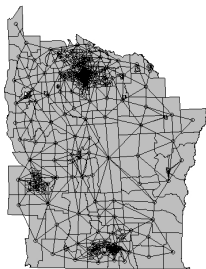
Géostatistique

Econométrie spatiale

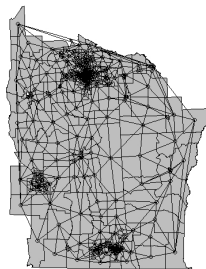
Semis de points

Interpolation

Matrice de contigüité



Triangularisation de Delaunay



Pour des données surfaciques (ici des sous-comptés (“counties”) de l’état de New-York), on considère les centroïdes ou autres points de référence (localisation des hôtels de ville, etc.) pour appliquer une triangularisation de Delaunay.

Visualisation de la structure de voisinage sous R (2)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

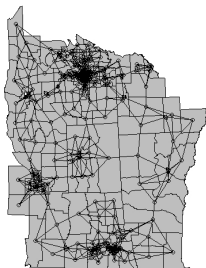
Géostatistique

Econométrie spatiale

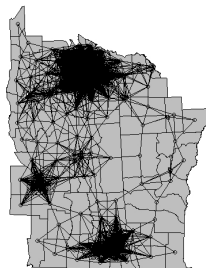
Semis de points

Interpolation

4 plus proches voisins



Distance seuil

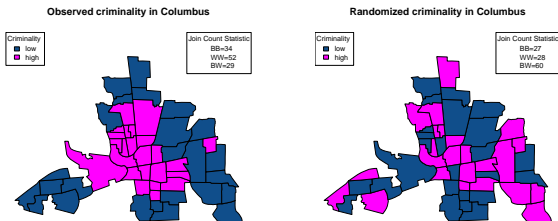


Pour le choix des distances : distance euclidienne à vol d'oiseau, distance réelle parcourue, temps de parcours, etc.

Test de “Join Count” (package `spdep`)

x_i est la valeur dans $\{0, 1\}$ ($=\{W, B\}$) de X sur le site s_i ; w_{ij} est la valeur de la matrice de voisinage W entre le site s_i et s_j .

- 1 On définit les statistiques observées suivantes (Cliff et Ord, 1981) : $BB = \frac{1}{2} \sum_{i,j} w_{ij} x_i x_j$, $WW = \frac{1}{2} \sum_{i,j} w_{ij} (1 - x_i)(1 - x_j)$ et $BW = \frac{1}{2} \sum_{i,j} w_{ij} (x_i - x_j)^2$.
- 2 Sous hypothèse de “non-free sampling” (tirage sans remise), on a le choix entre un test exact (`joincount.test`) ou un test de permutation . (`joincount.test` et `joincount.mc`).



Test de "Join Count" : exemple

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

On teste ici $H_0 = \{\text{absence d'autocorrélation spatiale}\}$.

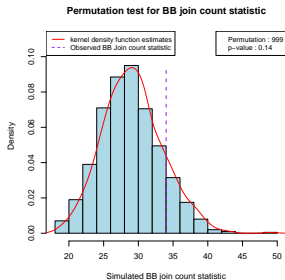


Tableau de résultat des tests exacts

| | Joincount | Expected | Variance | z-value |
|----|-----------|----------|----------|---------|
| BB | 34.00 | 29.34 | 18.64 | 1.08 |
| WW | 52.00 | 26.99 | 17.65 | 5.95 |
| BW | 29.00 | 58.67 | 26.04 | -5.81 |

Remarque: les tests sur BB et WW donnent des résultats opposés.

Tests pour variables continues (package `spdep`)

x_j est la valeur dans \mathbb{R} de X sur le site s_j .

- Le test du I de Moran. Soit $I = \frac{n}{\sum_{ij} w_{ij}} \frac{\sum_{ij} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$, on a le choix entre un test exact (Cliff et Ord, 1981) ou un test de permutation (fonctions `moran.test` et `moran.mc`).
- Autres tests possibles : le C de Geary (`geary.test`), le G de Getis-Ord (`globalG.test`)

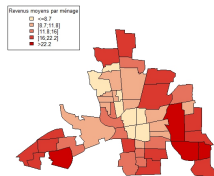
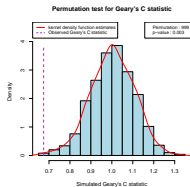
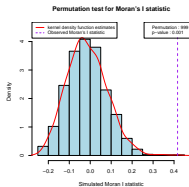


Diagramme de Moran

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

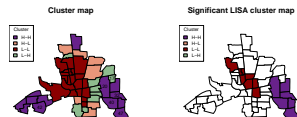
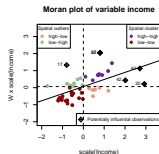
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Nuage de points de WX en fonction de X .
- Si W normalisée en ligne, WX est la "moyenne de X observée sur les voisins".
- Découpage en 4 quadrants autour de (\bar{X}, \bar{X}) : HH (High-High), LL (Low-Low), HL et LH.
- Autocorrélation spatiale globale forte : pente de la droite de régression (I de Moran) élevé.
- LISA (Anselin, 1995) : détection d'autocorrélation spatiale locale (I_i et test locaux).



Processus de Poisson

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

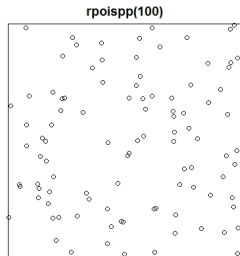
Econométrie spatiale

Semis de points

Interpolation

Soit un domaine D et $\{x_1, \dots, x_n\}$ un semis de points suivant un processus de Poisson homogène :

- il existe un réel $\lambda > 0$ tel que pour tout borélien A de D , $N(A)$ (le nombre de points observés sur A) suit une loi de Poisson de moyenne $\lambda |A|$, où $|A|$ désigne l'aire de A .
- sachant que $N(A) = n$, les n points du processus qui sont dans A forment un échantillon de la loi uniforme sur A .
- fonction `rpoispp` de **spatstat**.



CSR (Complete Spatial Randomness) simulé : les points sont indépendants et sont répartis sur D ($[0, 1] \times [0, 1]$ par défaut) au hasard.

Estimation de l'intensité

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

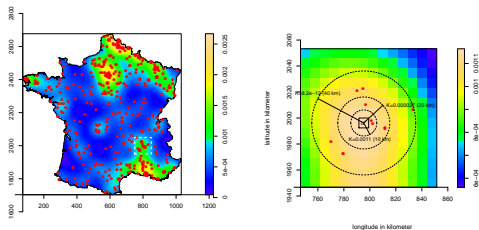
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Dans le cas d'un processus homogène, $\hat{\lambda}(u) = \frac{N(D)}{|D|}$.
- Dans le cas d'un processus inhomogène (*i.e.* présence d'hétérogénéité spatiale), on peut utiliser un estimateur non paramétrique : $\lambda^*(u) = \sum_{i=1}^n e(x_i)k(x_i - u)$, où k est le noyau gaussien et $e(x_i)$ un terme de correction de bord.

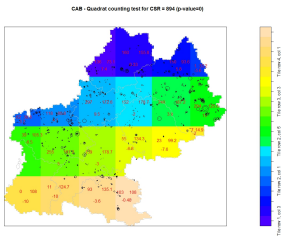


Lissage non paramétrique de l'intensité des phénomènes spatiaux non identifiés (fonction `density.ppp` de **spatstat**). Projet en collaboration avec le CNES.

Test d'homogénéité spatiale pour semis de points (1)

Test basé sur les quadrats (quadrat.test de **spatstat**):

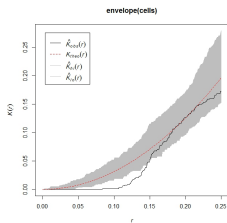
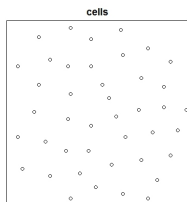
- 1 On découpe D en m quadrats.
- 2 On compte le nombre de points observés et attendus si les données étaient distribués selon un processus de Poisson homogène (“Complete Spatial Randomness”).
- 3 On applique un test de χ^2 à $m - 1$ degrés de liberté.



Test d'homogénéité spatiale pour semis de points (2)

Test basé sur des simulations :

- 1 On réalise M simulations d'un processus de Poisson d'intensité $\lambda = \frac{N(D)}{|D|}$.
- 2 On calcule une caractéristique d'ordre 2 du processus pour chaque simulation. Exemple : la fonction K de Ripley estimée par $\hat{K}(r) = \frac{|D|}{N(D)(N(D)-1)} \sum_{i,j} \mathbf{1}(d_{i,j} \leq r) e_{i,j}$.
- 3 On trace les enveloppes de ces courbes.



La fonction K associée aux données n'est pas incluse dans l'enveloppe ce qui amène à rejeter l'hypothèse de CSR (fonctions `Kest` et `envelope`).

1 Introduction

2 Analyse exploratoire de données spatiales

3 Outils spécifiques à la statistique spatiale

4 Modélisation spatiale

- Géostatistique
- Econométrie spatiale
- Semis de points

5 Méthodes d'interpolation de données surfaciques

Le Krigeage

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

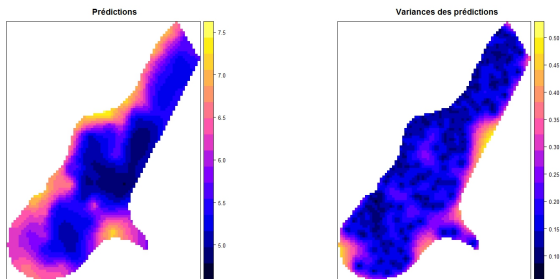
Interpolation

Soit y_1, \dots, y_n les valeurs observées d'une variable Y sur les sites s_1, \dots, s_n .

- Objectif : estimation d'une valeur non observée y_0 sur le site s_0 .
- Estimateur de la forme $\hat{Y}(s_0) = \sum_{i=1}^n \lambda_i Y(s_i)$, avec $\sum_i \lambda_i = 1$.
- $\hat{\lambda}_i$ obtenus en résolvant un système d'équations qui diffèrent selon l'hypothèse faite sur Y :
 - Krigeage simple : Y est stationnaire de moyenne connue.
 - Krigeage ordinaire : Y stationnaire de moyenne inconnue.
 - Krigeage universel : Y non-stationnaire (avec une tendance).
- Nécessite le choix d'un modèle de variogramme γ .

Exemple de Krigeage

- Calcul des prédictions + variances de prédictions (contrairement aux méthodes de lissage non paramétrique et interpolation spatiale).
- Fonctions `krige` de **gstat** et `spplot` de **sp** pour la représentation cartographique



Exemple de krigeage ordinaire en utilisant un variogramme sphérique.

Modèles d'économétrie spatiale (package `spdep`)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Contexte “usuel” :

- On cherche à expliquer une variable Y par un ensemble de variables exogènes X .
- L'analyse des résidus des modèles classiques (par exemple régression MCO) révèle la présence d'autocorrélation spatiale dans les résidus (diagramme et test de Moran pour résidus).
- Objectif : prendre en compte l'autocorrélation spatiale dans le modèle.
- Choix de modèles possibles : LAG et SDM (fonction `lagsarlm`), SEM (fonction `errorsarlm`), CAR (`spautolm`), etc.

Exemple : Laurent et Moussa (2015)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

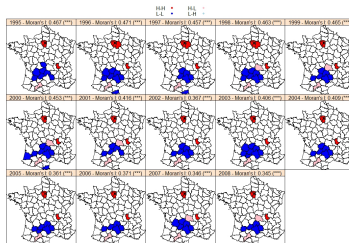
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Variable à expliquer Y : le nombre de brevets déposés par département à l'EPO (*European Patent Office*).
- Variables explicatives : dépenses en R&D dans les établissements privés, densité de population, PIB, etc.



Forte autocorrélation spatiale de la variable dépendante.

Modèle de type “Spatial Lag”

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

$$Y = \rho WY + X\beta + \epsilon, \text{ avec } \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

- Ce modèle prend en compte dans l'espérance de Y à la fois les variables exogènes X et la moyenne de Y dans les zones voisines.
- ρ ($\rho \in [0, 1[$), β , et σ sont estimés par maximum de vraisemblance ou la méthode des moments généralisés.
- Différents tests sur les paramètres et la présence d'autocorrélation spatiale dans les résidus (Wald, rapport de vraisemblance, multiplicateurs de Lagrange).
- Décomposition des effets marginaux (LeSage et Pace, 2012) en effets directs, indirects et totaux (fonction impacts).

Diffusion spatiale des spillovers

Statistique spatiale avec R

T. Laurent

Introduction

Analyse exploratoire

Cartographie

Graphiques interactifs

Visualisation

Outils spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

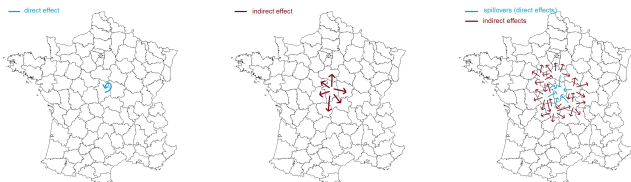


Illustration de la diffusion spatiale des spillovers. A du sens uniquement pour les modèles de type LAG et SDM.

| W-Order | Direct | indirect | total |
|--------------------|---------------|--------------|---------------|
| W^0 | 0.35032 (***) | 0 | 0.35032 (***) |
| W^1 | 0 | 0.03792 (*) | 0.03792 (**) |
| W^2 | 0.000838 | 0.00326 | 0.00410 |
| W^3 | 0.000030 | 0.00041 | 0.00044 |
| W^4 | 0.000005 | 0.00004 | 0.00005 |
| $\sum_{q=0}^4 W^q$ | 0.3511 (***) | 0.04158 (**) | 0.3927 (***) |

Prédiction en économétrie spatiale

Statistique spatiale avec R

T. Laurent

Introduction

Analyse exploratoire

Cartographie

Graphiques interactifs

Visualisation

Outils spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Goulard, Laurent et Thomas-Agnan (2017) :

| Predictor | Out-of-sample predictors formulae |
|-----------|--|
| BP | $\hat{Y}_O^{BP} = \hat{Y}_O^{TC} - \hat{Q}_{OO}^{-1} \hat{Q}_{OS} \times (Y_S - \hat{Y}_S^{TC})$ |
| TC | $\hat{Y}_O^{TC} = [(I - \hat{\rho}W)^{-1} X \hat{\beta}]_O$ |
| TS^1 | $\hat{Y}_O^{TS^1} = X_O \hat{\beta} + \hat{\rho} W_{OS} Y_S$ |
| BP_W | $\hat{Y}_O^{BPW} = \hat{Y}_O^{TC} + \hat{\Sigma}_{OS} W'_{OS} (W_{OS} \hat{\Sigma}_{SS} W'_{OS})^{-1} (W_{OS} Y_S - W_{OS} \hat{Y}_S^{TC})$ |
| BP_N | $\hat{Y}_O^{BPN} = \hat{Y}_O^{TC} - \hat{Q}_{OO}^{-1} \hat{Q}_{OJ} (Y_J - \hat{Y}_J^{TC})_J$ for J set of indices of neighbors of O |
| TC^1 | $\hat{Y}_O^{TC^1} = \text{row } o \text{ of } \{I_{n_S+1} - \hat{\rho} \begin{pmatrix} W_{SS} & W_{So} \\ W_{oS} & W_{oo} \end{pmatrix}\}^{-1} \begin{pmatrix} X_S \\ X_o \end{pmatrix} \hat{\beta}$ |
| BP^1 | $\hat{Y}_O^{BP^1} = \hat{Y}_O^{TC^1} - \hat{Q}_{oo}^{-1} \hat{Q}_{oS} (Y_S - \hat{Y}_S^{TC^1})$ |
| BP^1_W | $\hat{Y}_O^{BP^1_W} = \hat{Y}_O^{TC^1} + \hat{\Sigma}_{oS} W'_{oS} (W_{oS} \hat{\Sigma}_{SS} W'_{oS})^{-1} (W_{oS} Y_S - W_{oS} \hat{Y}_S^{TC^1})$ |
| BP^1_N | $\hat{Y}_O^{BP^1_N} = \hat{Y}_O^{TC^1} - \hat{Q}_{oo}^{-1} \hat{Q}_{oJ} (Y_J - \hat{Y}_J^{TC^1})_J$ for J set of indices of neighbors of o |

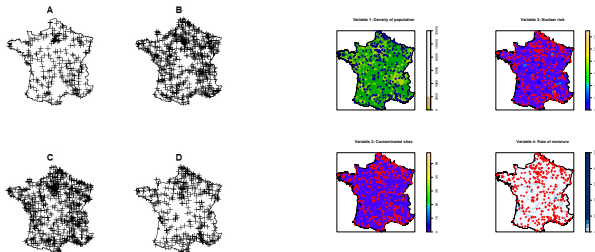
Modélisation de l'intensité

Processus de Poisson inhomogène : $\lambda(u) = e^{Z(u)\beta}$

- Les covariables Z doivent être observées sur les semis de points et d'autres locations $u \in D$ (d'où la nécessité de lisser ou interpoler les variables explicatives).

- Le processus peut être **marqué** :

$$y = \{(x_1, m_1), \dots, (x_n, m_n)\}, x_i \in D, m_i \in M.$$



Modélisation de l'intensité des PAN de type Poisson inhomogène avec des covariables de type anthropologie/environnement.

Résultats

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

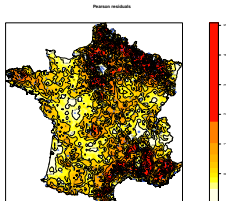
■ Estimation des coefficients du modèle :

| | coefficients | std | p-value | sign. |
|------------------|--------------|-------|---------|-------|
| (Intercept) | -8.998 | 0.220 | <0.001 | *** |
| log(pop) | 0.308 | 0.052 | <0.001 | *** |
| log(1 + conta) | 0.630 | 0.074 | <0.001 | *** |
| log(1 + nuclear) | 0.406 | 0.119 | <0.001 | *** |
| wetlands | 0.026 | 0.011 | 0.015 | * |

■ Prédictions : $\hat{\lambda}(u) = e^{Z(u)}\hat{\beta}$

■ Résidus : $\lambda^*(u) - \hat{\lambda}(u)$

■ Cartographie des résidus :



Ecriture générale d'un modèle de processus ponctuels

Statistique spatiale avec R

T. Laurent

Introduction

Analyse exploratoire

Cartographie

Graphiques interactifs

Visualisation

Outils spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Modèles de type : $\lambda(u, \mathbf{x}) = \exp(\psi^T B(u) + \phi^T C(u, \mathbf{x}))$
(Baddeley et Turner, 2000), avec $\theta = (\psi, \phi)$ les paramètres à estimer.

- $B(u)$ dépend seulement de location u et représente la tendance ou l'effet des covariables.
- $C(u, \mathbf{x})$ représente les interactions stochastiques et la dépendance entre les points du processus.
- Estimation des paramètres par pseudo-maximum de la vraisemblance.
- Les modèles de Gibbs : Poisson, Strauss, Geyer, etc.
- Fonction ppm de **spatstat**.

Modélisation de l'intensité des parcelles en CAB

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

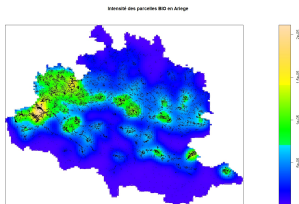
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

Projet en cours (ANR ModULand) : prendre en compte le fait que les agriculteurs se mettant au bio peuvent tenter leurs voisins de s'y mettre.



Observations de zones avec une forte intensité spatiale de parcelles en conversion AB. Objectif : prendre en compte des covariables environnementales (zones protégées, etc.) et le phénomène d'attraction. Modèles envisagés de type Geyer.

Conclusion Partie 1

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

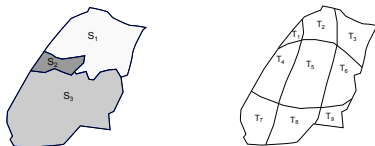
Interpolation

- Une multitude de domaines d'application.
- Nécessite des compétences pluridisciplinaires (notamment en géomatique).
- Plusieurs méthodes possibles (pas forcément incompatibles) en fonction de la nature "spatiale" des données et des objectifs.
- Logiciel R couvre la grande majorité des champs d'application de la statistique spatiale
- Projets en cours : utilisation du bootstrap spatial, modélisation de données de flux.

- 1 Introduction
- 2 Analyse exploratoire de données spatiales
- 3 Outils spécifiques à la statistique spatiale
- 4 Modélisation spatiale
- 5 Méthodes d'interpolation de données surfaciques**

Problématique et références

Objectif : transférer les données observées sur des zones surfaciques (**source**) vers d'autres zones (**target**).



Références:

- Do Van Huyen, Christine Thomas-Agnan, Anne Vanhems (2015). Spatial reallocation of areal data: a review, *RERU*.
- Do Van Huyen, Christine Thomas-Agnan, Anne Vanhems (2015). Accuracy of areal interpolation methods: Count data, *Spatial Statistics*.
- Do Van Huyen, Christine Thomas-Agnan, Anne Vanhems (2015) Testing areal interpolation methods with US census 2010 data, *Région et Développement*.

Notations (unités spatiales)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

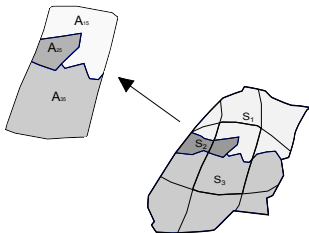
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- $\{S_i\}_{i=1,2,\dots,I}$: sources,
- $\{T_j\}_{j=1,2,\dots,J}$: targets,
- $\{A_{ij}\}$: intersection entre S_i et T_j , A_{ij} peut être vide ($i = 1, 2, \dots, I, j = 1, 2, \dots, J$).



Notations (unités statistiques)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- **Variable d'intérêt:** Y . On connaît Y sur les sources y_s et on souhaite la transférer sur les targets y_t . Si on note Ω le domaine entier et $\{\Omega_k\}_k$ les sous-domaines de Ω , alors Y est dit:
 - extensive si $Y_\Omega = \sum_k Y_{\Omega_k}$ (ex : population),
 - intensive si il existe des w_k t.q. $Y_\Omega = \sum_k w_k Y_{\Omega_k}$ où $\exists k$ t.q. $w_k \neq 1$ (ex : densité de population)
- **Information auxiliaire:** X (réseaux routiers, etc.) accessible au niveau des targets, ou bien des intersections (variable quantitative).
- **Zones de contrôle :** l'information auxiliaire est qualitative (usage des sols).

Nombre de logements sur 8 communes (1)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

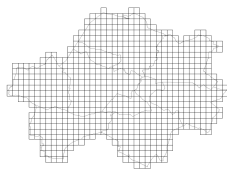
Modélisation
spatiale

Géostatistique

Econométrie spatiale

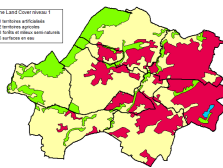
Semis de points

Interpolation



Matrice Land Cover (niveau 1)

- 1. surfaces agricoles
- 2. surfaces agricoles
- 3. forêts et milieux semi-naturels
- 5. surfaces en eau



Nombre de logements sur 8 communes (2)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

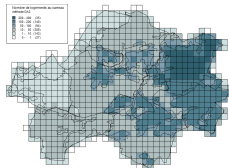
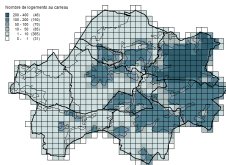
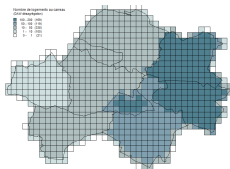
Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation



Hypothèses au choix

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

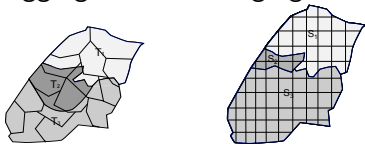
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Aggrégation ou désagrégation ?



- $Y_{A_{ij}} = Y_{S_i}$, Y est-elle uniforme sur les sources (variable intensive) ?
- $Y_{A_{ij}} = Y_{T_j}$, Y est-elle uniforme sur les targets (variable intensive) ?
- La distribution de Y est-elle connue ? Par exemple, si Y est une variable de comptage, Y peut suivre une loi de Poisson, et si Y est continue, une loi normale.
- Propriété pycnophylactique?

Les principales méthodes

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- “Point in polygon method” (**PIP**).
- “Dasymetric methods”: allocation proportionnelle à une variable auxiliaire (ex: surface, longueur des routes) (**DAW**, **DAX**).
- Régression sur variable(s) auxiliaire(s) (ex: longueur des routes, classification des sols) (**REG**).
- Lissage avec ou sans la propriété pycnophylactique.

Point-in polygon PIP

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

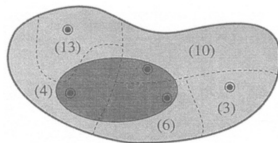
Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Référence : Sadahiro (2000).
- Seulement pour des données de comptage.
- Un point “représentatif” est choisi pour chaque source. La valeur source associée à un point est assignée à la target auquel le point appartient.



$$4 + 10 + 6 = 20$$

- Ne satisfait pas la propriété pycnophylactique.

Méthode proportionnelle à l'aire (DAW)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Référence : Sadahiro (1999), Bloom *et al.* (1996), etc. (méthode implémentée par défaut sur MapInfo).
- Si $|A|$ dénote la surface de la région A ,
 - 1 si Y est extensive (ex : donnée de population) , la donnée source est allouée aux intersections proportionnellement aux aires :

$$\hat{Y}_{A_{st}} = \frac{|A_{st}|}{|S_s|} Y_{S_s}$$

- 2 si Y est intensive (ex : densité de population), Y est supposée uniforme sur les sources:

$$\hat{Y}_{A_{st}} = Y_{S_s}$$

- Satisfait la propriété pycnophylactique.

Méthode dasymétrique avec variable auxiliaire X (DAX)

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Référence : Voss *et al.* (1999), Reibel (2004)
- Idée similaire à DAW en remplaçant l'information sur l'aire par X (NB : X , supposée connue au niveau des intersections).
 - 1 si Y et X sont extensives, on fait l'hypothèse que la donnée source y peut être affectée au niveau des intersections de façon proportionnelle à x : $\hat{Y}_{A_{st}} = \frac{x_{A_{st}}}{x_{S_s}} y_{S_s}$
où $x_{S_s} = \sum_t x_{A_{st}}$.
 - 2 si Y est intensive, on doit utiliser la propriété qui permet de passer d'une variable extensive à intensive.
- Satisfait la propriété pycnophylactique.

Méthode dasymétrique sur des zones de contrôle (DAC)

- Référence : Mennis and Hultgren (2006).
- Utilise une information subjective sur des zones de contrôle en faisant l'hypothèse que les intersections sont emboîtées dans les zones de contrôle
- Si Y est une donnée de comptage :

$$\hat{Y}_{A_{st}} = y_{S_s} \frac{|A_{st}| \hat{D}_c(st)}{\sum_l |A_{sl}| \hat{D}_c(sl)}$$

où \hat{D}_c est une valeur de densité sur les zones de contrôle c (donnée par un expert ou estimée statistiquement) :

$$\hat{D}_c = \frac{\sum_s y_s}{\sum_s |S_s|}, \quad s \in c,$$

où $s \in c$ peut avoir plusieurs significations (inclusion, centroid, pourcentage de couverture).

Régression sans variable auxiliaire

- Référence : Goodchild *et al.* (1992)
- Objectif : sous certaines hypothèses, dériver un système d'équation qui lie Y sur les sources à Y sur les targets et utiliser un OLS pour le résoudre.

1 si Y est extensive, l'hypothèse est que Y peut être allouée aux intersections proportionnellement à leur surface :

$$Y_{S_s} = \sum_t \hat{Y}_{A_{st}} = \sum_t \frac{|A_{st}|}{|T_t|} \hat{Y}_{T_t}$$

2 si Y est intensive avec des poids de surface, l'hypothèse est que Y soit uniforme sur les targets

$$y_{S_s} = \sum_t \frac{|A_{st}|}{|S_s|} \hat{Y}_{A_{st}} = \sum_t \frac{|A_{st}|}{|S_s|} \hat{Y}_{T_t}$$

Régression avec variables auxiliaires

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Référence : Flowerdew et Green (1992).
- Y intensive. Au moins deux variables auxiliaires.
- Sous hypothèse que Y est uniforme sur les targets:
 - 1 Ecrire la relation entre Y sur les sources et Y sur les targets: $Y_{S_s} = \sum_t w_{st} Y_{A_{st}} = \sum_t w_{st} Y_{T_t}$ où w_{st} sont connus.
 - 2 Faire a régression de Y sur les targets sur les X sur les targets $Y_T = X_T \beta$ pour obtenir la régression de Y sur les sources vers X sur les targets $Y_S = WX_T \beta$, estimer les coefficients.
 - 3 A partir des coefficients estimés et des X sur les targets, obtenir les valeurs ajustées des Y sur les targets.

Autres méthodes

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

- Modèle de Poisson pour des données de comptage avec information auxiliaire au niveau des target (EM algorithm).
- Modèle de régression incluant de l'autocorrélation spatiale (Murakami and Tsutsumi, 2011).
- Méthodes bayésiennes avec autocorrélation spatiale (Mugglin *et al.*, 2000) ou sans (Mugglin et Carlin, 1998).
- Méthode d'interpolation pycnophylactique (Tobler, 1979).
- Krigeage de type Poisson (Goovaerts, 2006).

Résumé des méthodes

Statistique
spatiale avec
R

T. Laurent

Introduction

Analyse
exploratoire

Cartographie

Graphiques
interactifs

Visualisation

Outils
spécifiques

Géostatistique

Econométrie spatiale

Semis de points

Modélisation
spatiale

Géostatistique

Econométrie spatiale

Semis de points

Interpolation

| MET | Target var. Y | | Auxiliary var. X | | | Control zones |
|----------------|----------------------------|--|--------------------------|--|----------------------|------------------|
| | Nature | Add. assumptions | Dimension | Nature | Support | |
| DAW | Extensive Intensive | Homogeneous on sources | none | none | none | none |
| DAX | Extensive Intensive | none | 1 | Extensive or intensive | intersection | none |
| DAC | Extensive Intensive | Homogeneous on controls Homogeneous on controls | 1 | categorical | control | yes |
| REG wo aux. | Extensive Intensive | Homogeneous on targets | none | none | none | none |
| REG w. aux | Extensive Intensive | none weight area | ≥ 1 ≥ 1 | Extensive or intensive Extensive or intensive | target target | none none |
| PIP | Extensive | none | none | none | none | none |