

Prendre en compte l'hétérogénéité spatiale avec le lissage

Manuel de statistique spatiale

Auriane Renaud, Arlindo Dos Santos et François Sémécurbe

Pôle Analyses Urbaines

17 mai 2018

Cadre de cette présentation

- Dans cette présentation, l'approche cartographique est privilégiée pour introduire les concepts du lissage spatial.
- L'intérêt premier est de transmettre une vision intuitive et spatiale du lissage tout en abordant les problématiques géographiques sous-jacentes auxquelles il est relié.
- De ce fait, nous n'évoquerons quasiment pas ici, le lissage comme outil d'estimation de l'intensité d'un processus ponctuel.

Pour en savoir plus...

Pour une présentation plus approfondie du lissage, un chapitre lui est consacré dans le manuel de statistique spatiale .

Carte de densité de population

- Avec l'avènement des États-nations à la fin du 18^e siècle, la répartition des populations dans l'espace est devenue un enjeu majeur de connaissance.
- Les statisticiens de l'époque ont forgé progressivement les outils cartographiques que nous utilisons à présent couramment pour représenter l'hétérogénéité spatiale de nos statistiques.
- Un type de représentation a en particulier transformé en profondeur notre façon de se représenter le peuplement: la carte de densité. Celle-ci transforme une donnée ponctuelle en une donnée surfacique.

Une question de vocabulaire

En géographie, la **densité** représente le nombre d'occurrences d'un phénomène dans l'emprise spatiale d'un territoire. En statistique spatiale, on emploie plutôt le terme d'**intensité**. Dans cette présentation, nous utilisons indistinctement les deux termes.

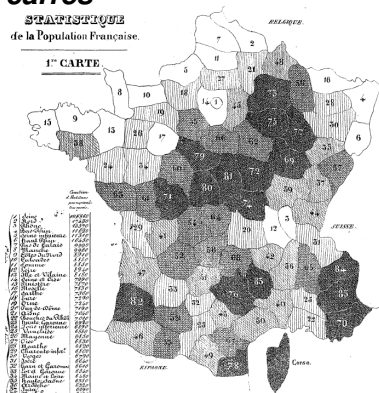
Carte de densité de population - Exemples

On suppose que la population est répartie uniformément sur l'emprise spatiale des unités territoriales.

Habitants par myriamètres carrés

STATISTIQUE
de la Population Française.

1^{re} CARTE.

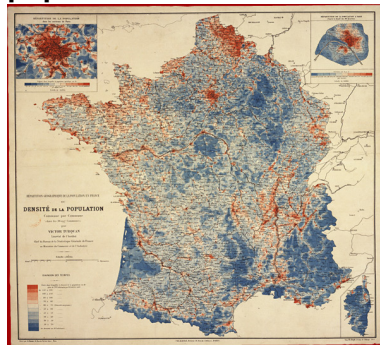


Carte

du Comte A. D'Angéville tirée de *l'Essai sur la statistique* (1836)

Pôle Analyses Urbaines

Répartition géographique de la population en France



Carte de Victor Turquan (1887)

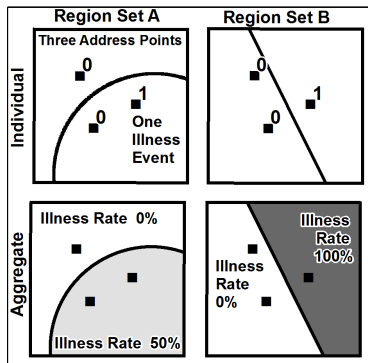


Le Maup - Théorie

Modifiable Areal Unit Problem (Openshaw et Taylor, 1979) : les cartes de densité ont néanmoins un défaut, les informations qu'elles dévoilent, **sont tributaires du découpage utilisé.**

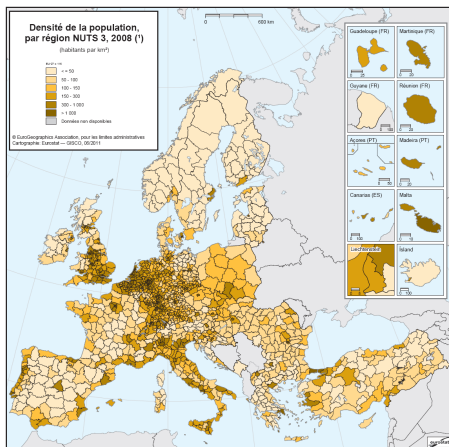
Le Maup se décompose en deux sous-problèmes interdépendants:

- l'effet d'échelle : les résultats sont fonctions de la taille des mailles ;
- l'effet de zonage : les résultats sont fonctions de la forme des mailles.



CC BY 3.0

Le Maup - En pratique

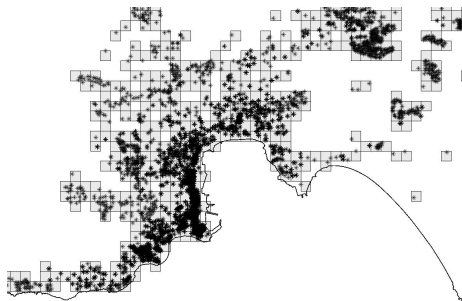


(*) La densité de la population est le rapport entre la population (moyenne) et la superficie terrestre. La superficie terrestre est la superficie totale sans les eaux intérieures. Pour Belgique, Danemark, Allemagne, France, Chypre, Pologne et Portugal, la superficie totale a été utilisée à la place de la superficie terrestre. Pologne, par région NUTS 2, Royaume-Uni, 2007.

«L'État garantissait l'ordre social. Mais, en tant que "conteneur" , il offrait une unité territoriale pour l'ensemble des statistiques dont avait besoin la science sociale empirique pour étudier les processus sociaux et économiques. Les catégories étatiques devinrent les principales catégories opératoires de la science sociale empirique. » (John Agnew, 2014)

Le carroyage - théorie

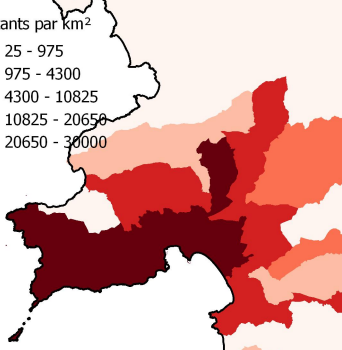
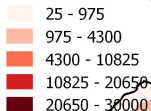
- Le carroyage est une grille régulière qui sert de découpage territorial
- L'effet d'échelle est maîtrisé: l'utilisateur définit la taille de ses carreaux



le carroyage - pratique

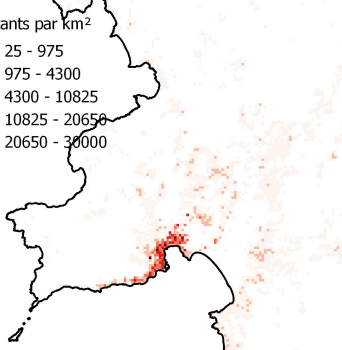
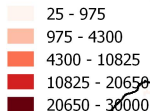
Densité de population

habitants par km²



Densité de population

habitants par km²



Les effets de zonage ne sont pas contrôlés

Les résultats dépendent du choix de l'orientation et de l'origine de la grille utilisée.

Le lissage - théorie

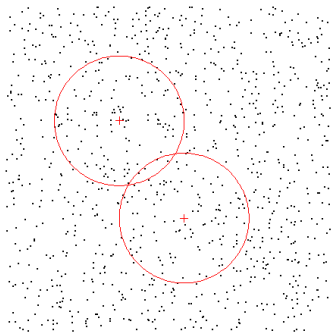
Le lissage est une méthode d'estimation locale de la densité qui s'affranchit des zonages:

- il peut être compris comme une généralisation de la carte de densité qui contrôlerait les effets du découpage géographique.
- il exploite un découpage territorial «mobile» (par exemple un disque de rayon r) qui s'adapte à chaque point observé.
- pour un point d'estimation $u = (u_x, u_y)$ le lissage consiste à faire une moyenne pondérée locale en accordant plus de poids aux points analysés proches du point d'estimation et moins aux points éloignés.

Le lissage - théorie

Utilisation du noyau le plus simple: le disque uniforme

- Pour un point d'observation u , on dénombre $N_r(u)$ les points analysés contenu dans le disque centré sur u de rayon r .
- La densité est obtenue au point u par la formule simple : $\frac{N_r(u)}{\pi r^2}$
- Dans l'exemple de droite, le rayon de lissage est de 400m. Pour le point d'estimation u situé en haut, $N_r(u) = 123$, la densité est 244 points par km^2 . Pour le point d'estimation du bas, $N_r(u) = 113$ et la densité est 224 points par km^2 .

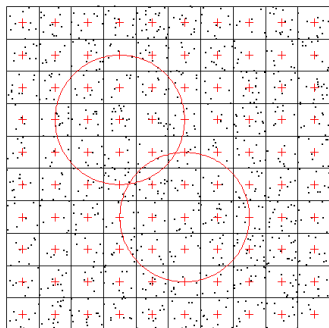


Les croix rouges sont les points d'observations, les cercles rouges l'emprise des disques de rayon $r = 400$ et en noir les points analysés.

Le lissage - pratique

En pratique on discrétise l'espace d'observation de telle sorte que la taille des carreaux soit petite par rapport à la taille du rayon de lissage et donner ainsi l'illusion du continu.

Les points d'estimations sont les centroides d'un carroyage.

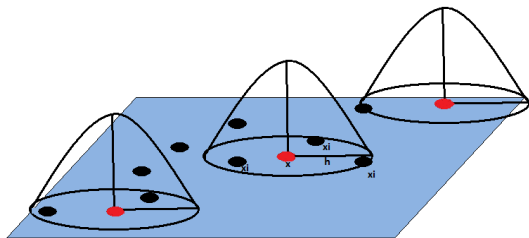


L'information est associée aux carreaux pour permettre des représentations cartographiques

54	82	88	97	101	97	87	85	73	61
69	103	123	128	130	127	118	118	118	87
72	110	111	120	128	124	126	111	125	96
74	99	109	123	115	112	125	115	125	88
72	100	118	122	120	106	120	119	116	84
59	97	101	108	116	115	128	128	116	86
66	89	115	117	117	113	126	148	137	93
64	101	113	112	114	119	130	149	142	103
68	101	113	120	114	111	123	128	126	91
54	77	88	89	83	84	75	93	86	58

Le lissage : Le noyau

Une façon de généraliser le lissage est de recourir à des noyaux non uniformes. La pondération décroît avec la distance au point d'estimation. Les principaux noyaux utilisés sont le noyau **biweight** et **gaussien**.

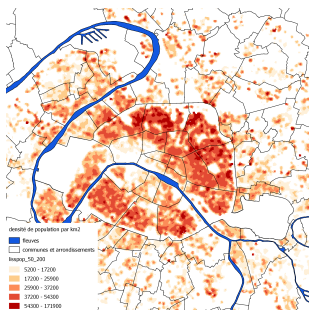


Definition

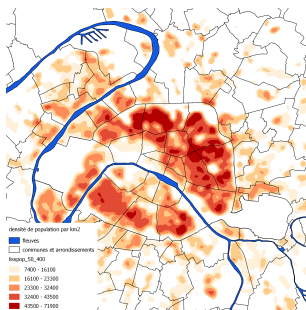
$$\lambda(u) = \sum_{i=1}^n \frac{1}{r^2} P_i K\left(\frac{u-p_i}{r}\right),$$

où $u = (u_x, u_y)$ est le point d'estimation, $p_i = (x_i, y_i)$ les points analysés et K un noyau (une fonction positive et d'espérance égale à 1), P_i la pondération et r le rayon de lissage.

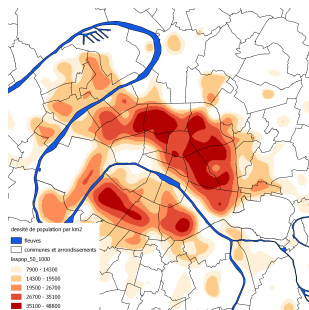
Le rayon de lissage joue le rôle de filtre



$r=200m$



$r=400m$



$r=1000m$

- un rayon de lissage important conserve uniquement les structures globales et masque des spécificités locales;
- un rayon de lissage faible conserve les détails locaux mais également les potentielles erreurs contenues dans la source étudiée.

Le traitement des effets de bords

Par rapport aux estimations non paramétriques auxquelles il s'apparente, le lissage spatial introduit un concept supplémentaire : **la fenêtre d'analyse** ("le territoire englobant").

- Si la fenêtre contraint le phénomène, par exemple si la fenêtre dessine une frontière maritime, on privilégie un traitement conservatif des effets de bords (Diggle)
- Si la fenêtre est uniquement induite par la capacité de l'observateur à décrire son phénomène, on choisit au contraire une correction uniforme qui tend à introduire une continuité entre l'intérieur et l'extérieure de la fenêtre.

Le lissage : À retenir

Le noyau

Comme en statistique non paramétrique, le choix du noyau n'est pas crucial.

Les temps de calcul

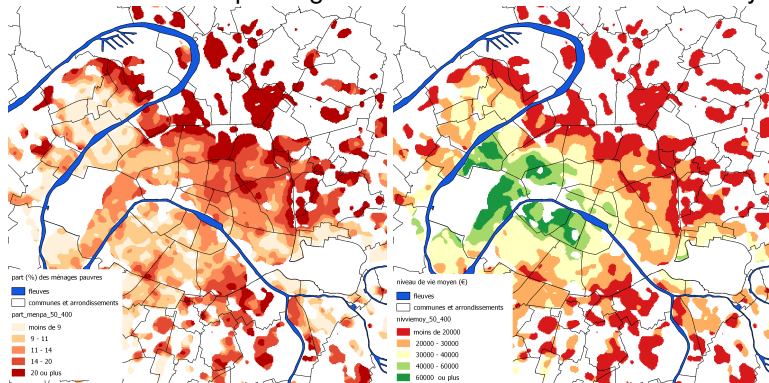
Le choix du traitement des effets de bords peut induire des différences de temps de calcul considérables.

Le rayon de lissage

Le choix du rayon de lissage, soit de l'échelle de restitution de l'information conditionne fortement l'aspect des cartes produites.

Les cartes de ratio et de moyenne

On peut lisser une population et une sous population pour produire des cartes de taux. On peut également calculer des indicateurs moyens.

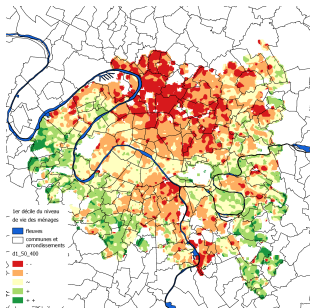


Attention

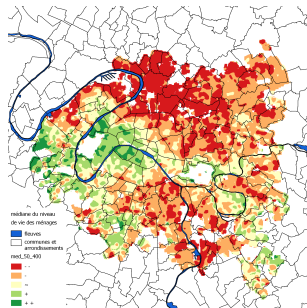
On ne lisse pas un taux, mais le numérateur et le dénominateur séparément puis on calcule pour chaque carreau le ratio.

Médiane géographiquement pondérée

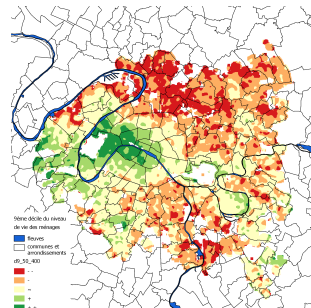
Une façon d'étendre le lissage est de choisir un opérateur d'agrégation des données non linéaires tel que la médiane, voir des quantiles..



Premier décile



Médiane



Neuvième décile

Choix du rayon de lissage

- Jusqu'à présent nous avons éludé la question du choix du rayon de lissage en expliquant qu'il est un paramètre de restitution de l'information.
- Des méthodes de choix de rayon optimal existent mais nous montrons dans le manuel l'hétérogénéité de leur résultat.
- De toute façon, même en choisissant un rayon optimal, on ne fait que déplacer le problème. La détermination de la fenêtre impacte les calculs d'optimisation.
- Des méthodes de lissages adaptatifs existent, mais sont-elles pertinentes pour des données sociales?
En effet, le rayon de lissage caractérise l'environnement des individus, on peut lui attribuer un sens social.

Les packages R

Plusieurs packages sont susceptibles de réaliser des lissages

- Le package spatstat qui est la boîte à outils dédiée à l'analyse spatiale.
- Le package btb: Beyond the Border développé par le Pôle d'analyse Urbaine dont Arlindo Dos Santos assure la maintenance.