

# Webinar



Sciences Economiques et  
Sociales de la Santé et Traitement  
de l'Information médicale  
[sesstim.univ-amu.fr](http://sesstim.univ-amu.fr)



[openhealth-institute.org](http://openhealth-institute.org)

## *Dr Aline GUTTMANN*

FMD, PhD, PHU

Institut Pascal UMR 6602 UCA/ CNRS/SIGMA - T.G.I. (Thérapies Guidées par l'image)

CHU Clermont-Ferrand -Pôle Santé Publique - U.B.I.M (Unité de Biostatistique et Informatique Médicale)

## **Modélisation des données spatiales hétérogènes. Applications en santé.**

mars 2017



**Cliquez ici pour voir l'intégralité des ressources associées à ce document**



# *Modélisation des données spatiales hétérogènes*

## ***Applications en santé***

Université Clermont Auvergne, Faculté de Médecine (R2, 5e étage),  
28 Place Henri Dunant - BP 38, 63001 CLERMONT-FERRAND Cedex 1

03 mars 2016

**INSTITUT PASCAL** - UMR 6602 UCA/CNRS/SIGMA - *T.G.I. (Thérapies Guidées par l'Image)*



# Statistique spatiale

« La statistique spatiale étudie des phénomènes dont l'observation est un processus aléatoire  $X = \{X_s, s \in S\}$  indexé par un ensemble spatial  $S$ ,  $X_s$  appartenant à un espace d'états  $E$ . »

*Gaetan, C. & Guyon, X. Modélisation et statistique spatiales Springer Berlin Heidelberg, 2008*

- Ensemble spatial  $S$ , sites d'observation  $s$ 
  - sites d'observation  $s$ 
    - fixe
    - aléatoire
  - ensemble spatial  $S$ 
    - unidimensionnel
    - bidimensionnel
    - tridimensionnel
    - ...
  - dimension temporelle
- Espace d'états  $E$ 
  - quantitatif continu
  - dénombrement
  - booléen
  - ...



# Données géostatistiques



# Définition

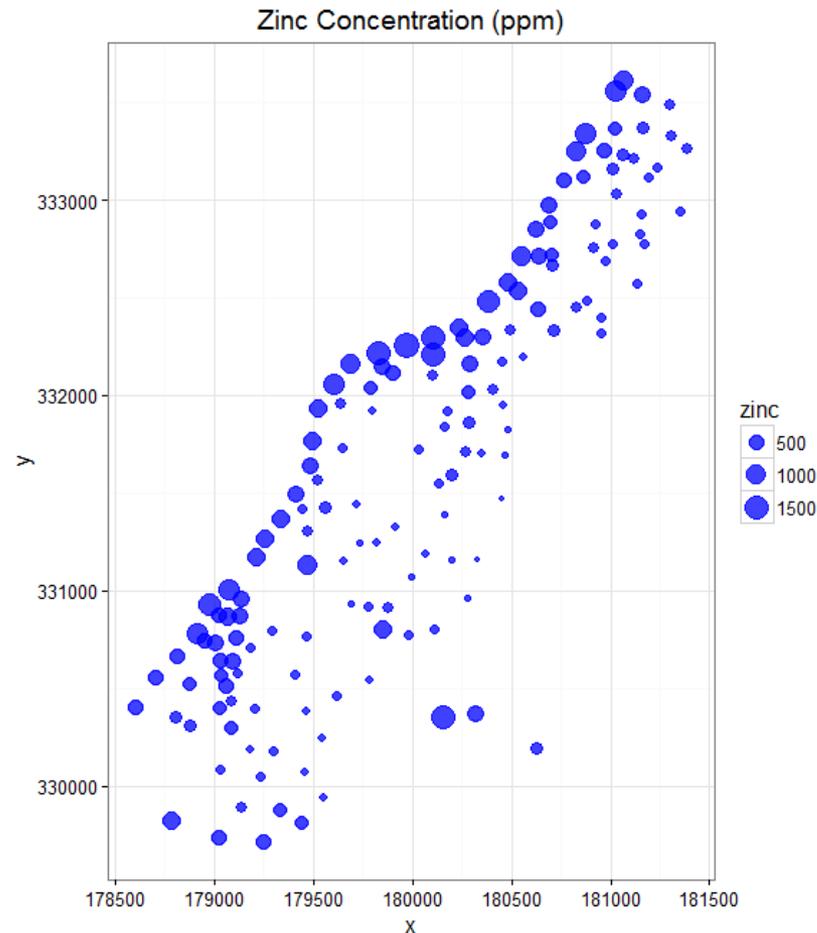
- les sites d'observations
  - fixés et déterministes
  - réguliers ou non
- $X_s$  est un réel
- Exemples
  - prospection minière
  - surveillance aérobiologique, météorologique et de la radioactivité
- Questions d'intérêt
  - modélisation des variations
  - prédiction et reconstruction



# Prédiction et reconstruction

## Exemple avec r – jeu de données « meuse »

- locations and topsoil heavy metal concentrations,
- along with a number of soil and landscape variables at the observation locations,
- collected in a flood plain of the river Meuse, near the village of Stein (NL).
- Heavy metal concentrations are from composite samples of an area of approximately 15 m x 15 m.



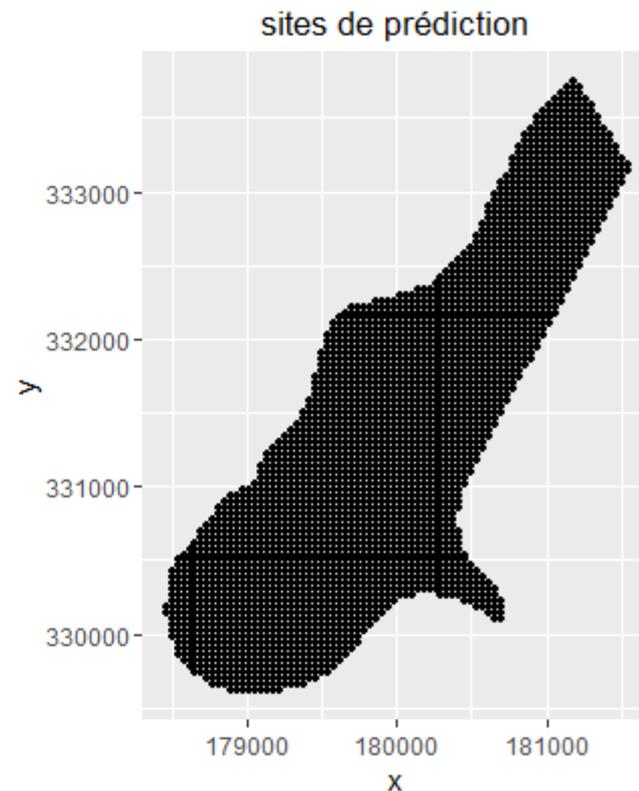
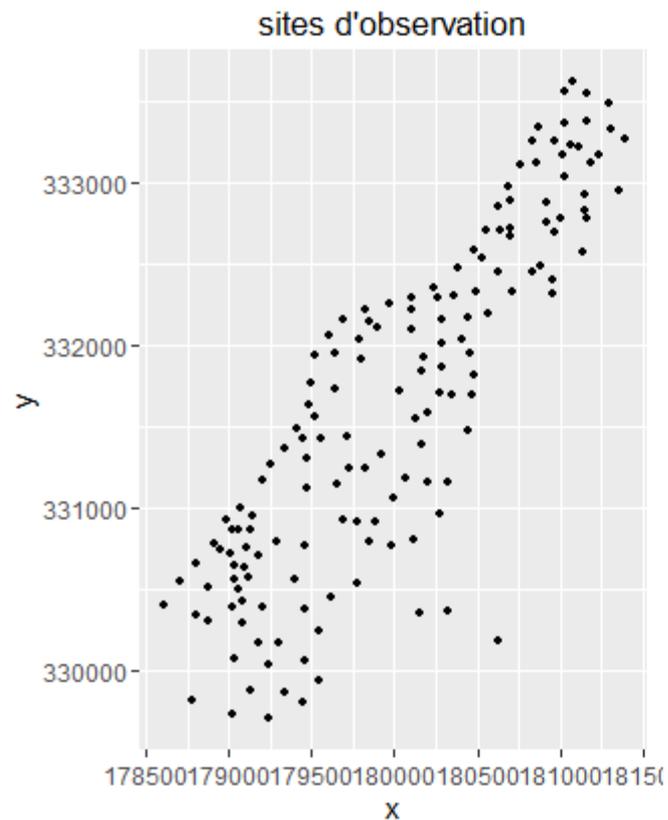


# Prédiction et reconstruction

Exemple avec r – jeu de données « meuse »

## Objectif

À partir des quelques valeurs observées, quelles sont les concentrations en zinc dans l'ensemble de la région d'étude?



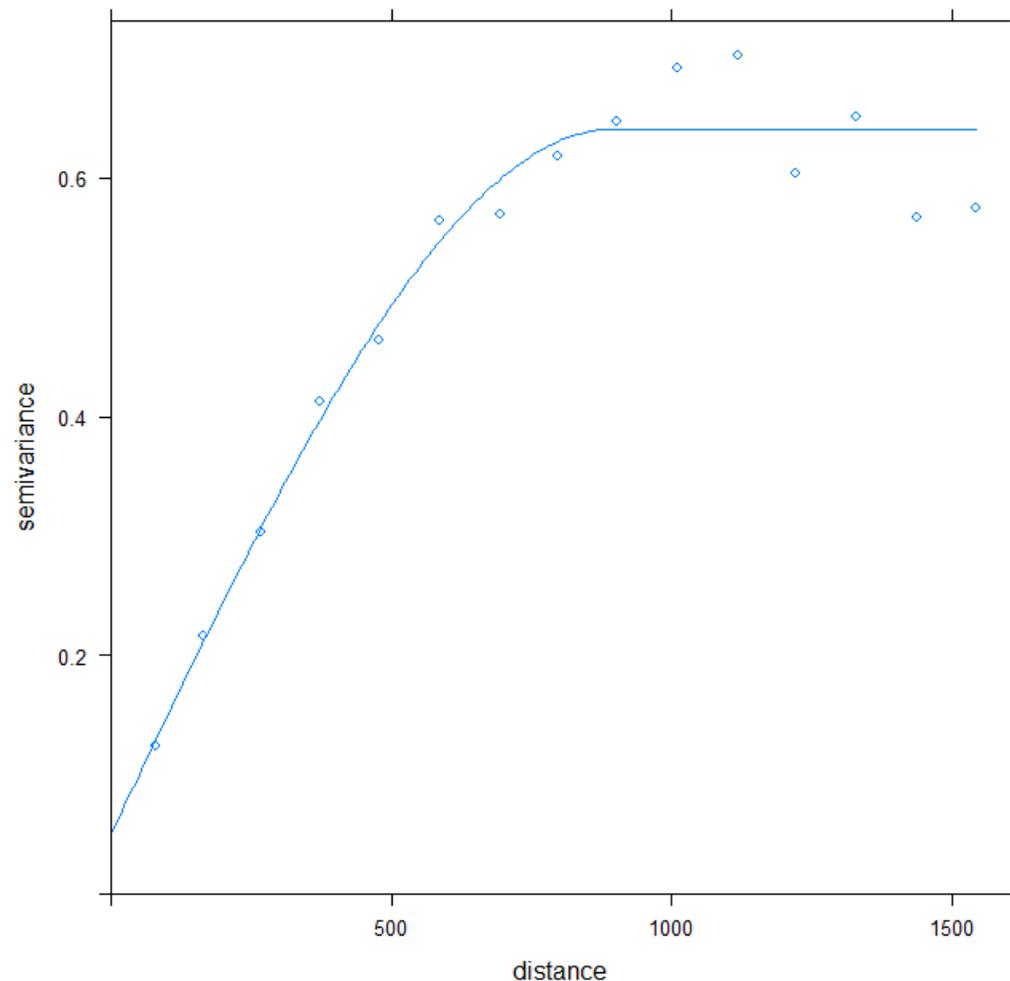


# Prédiction et reconstruction

Exemple avec r – jeu de données « meuse »

## Krigeage

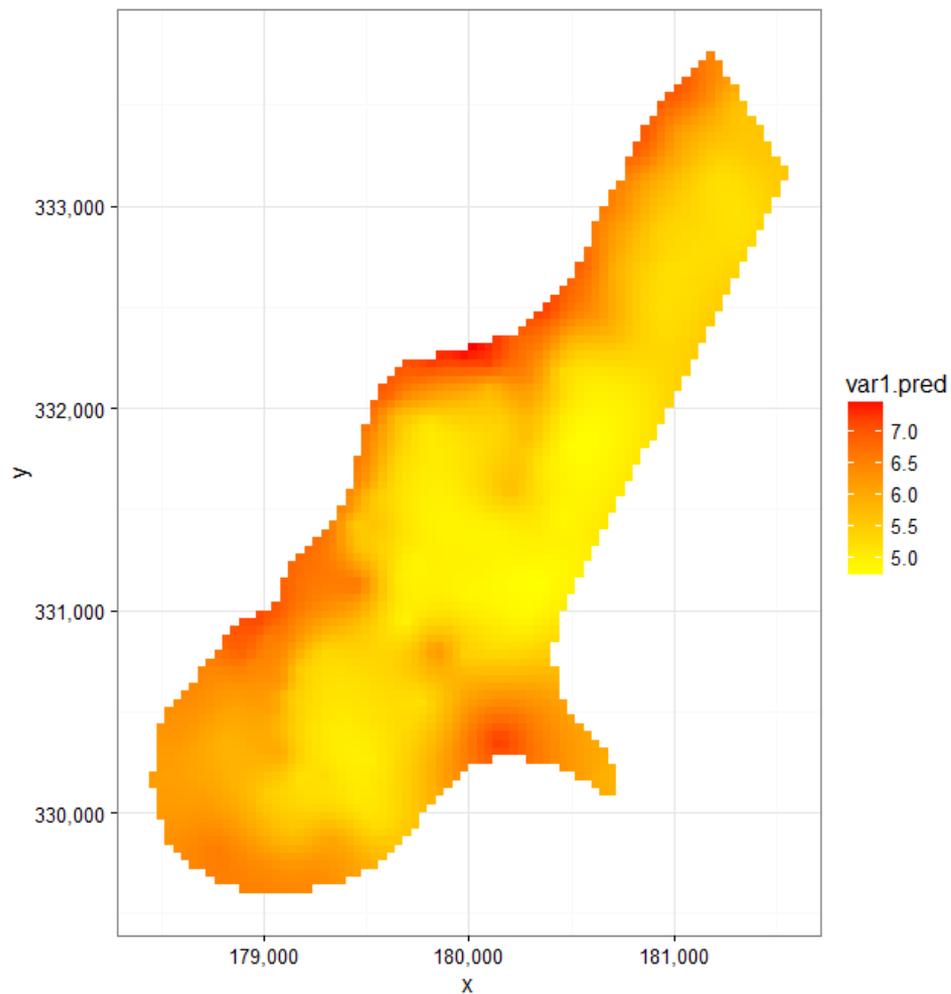
- Principe : interpolation ou estimation des valeurs non observées à partir des valeurs observées
  - Variogramme
- Les valeurs observées pour des points rapprochés sont, en moyenne, plus proches que les valeurs observées entre des points éloignés





# Prédiction et reconstruction

Exemple avec r – jeu de données « meuse »





# Données ponctuelles



# Définition

## Les données ponctuelles

- sites  $s$  d'observation
  - aléatoires (en nombre et localisation)
  - processus ponctuel spatial (PPS)
- PPS est dit marqué si, pour chaque  $s$ , on recueille une mesure d'intérêt
- Exemples
  - foresterie (disposition des arbres – essence)
  - géomarketing (localisation d'entreprises – type de services/biens)
  - sismologie (secousses sismiques – intensité)
  - épidémiologie (adresses de patients – caractéristiques démographiques ou état de santé)



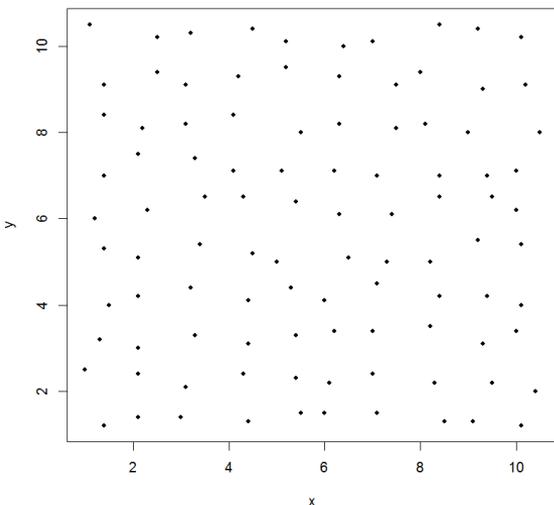
# Processus ponctuel spatial – spatial point process

## Question d'intérêt

### Répartition spatiale

régulière

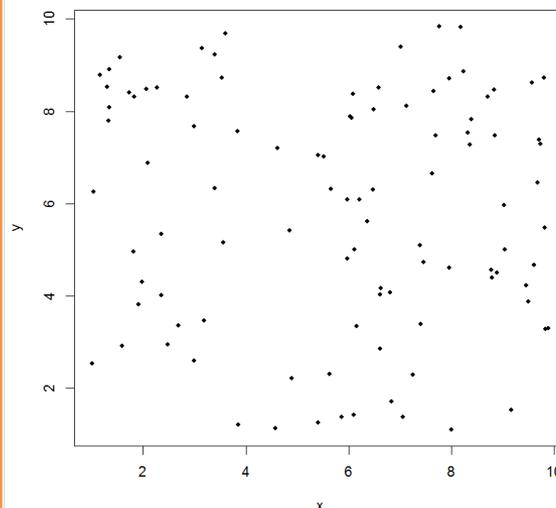
Effet de voisinage :  
répulsion



complètement aléatoire

Aucun effet de  
voisinage

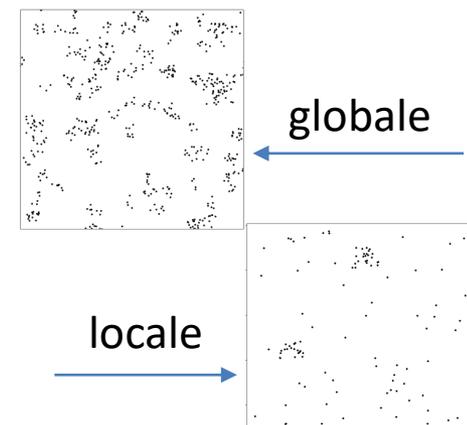
P. Poisson Homogène



agrégée

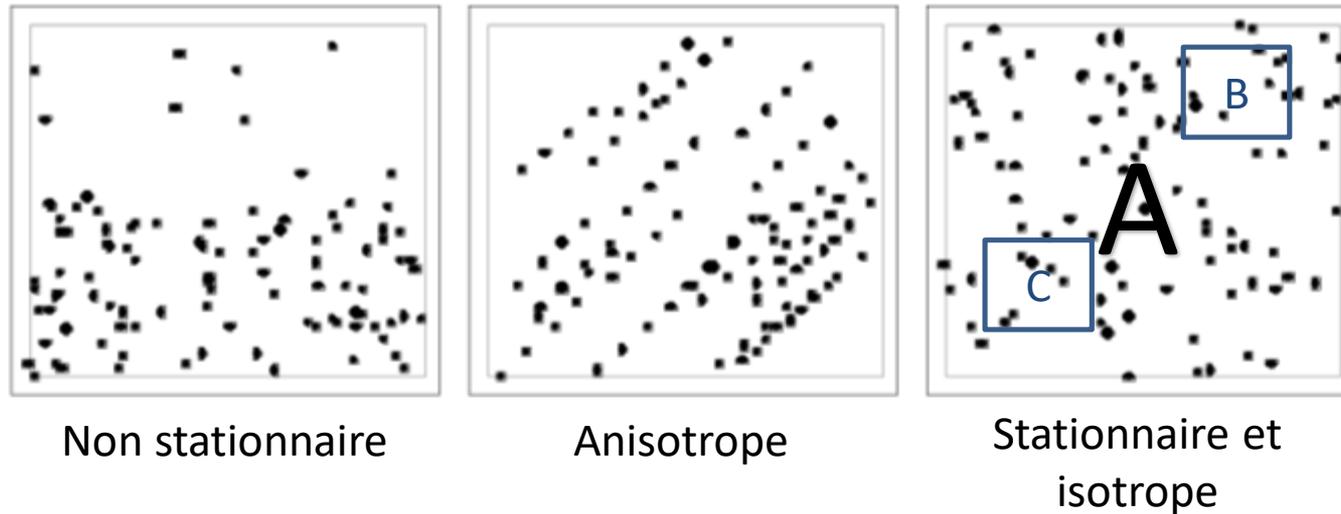
Effet de voisinage :  
attraction

P. Poisson Inhomogène



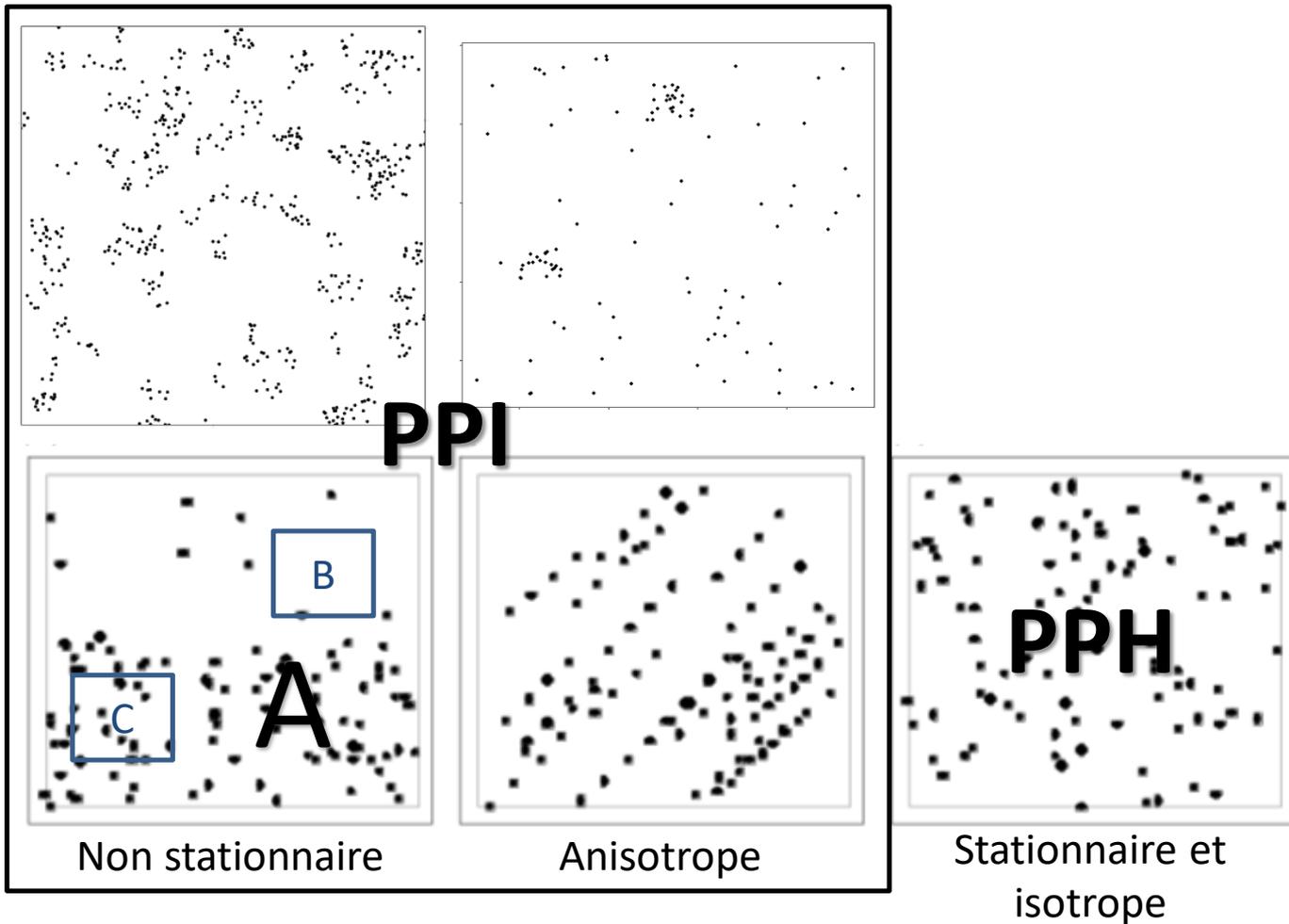
Un processus est dit homogène (stationnaire) s'il est invariant par translation, c'est-à-dire si ses propriétés ne varient pas d'une position à l'autre de l'espace.

Un processus est dit isotrope s'il est invariant par rotation, c'est-à-dire si ses propriétés ne varient pas avec l'orientation de l'espace.



### Processus de Poisson Homogène d'intensité $\lambda$

- Pour tout B dans A,  $N_B$  suit une loi de Poisson de paramètre  $\lambda/B$  où  $|B|$  = aire de B
- Si B et C sont disjoints,  $N_B$  et  $N_C$  sont deux variables indépendantes



Processus de Poisson Inhomogène d'intensité  $\lambda(\cdot)$

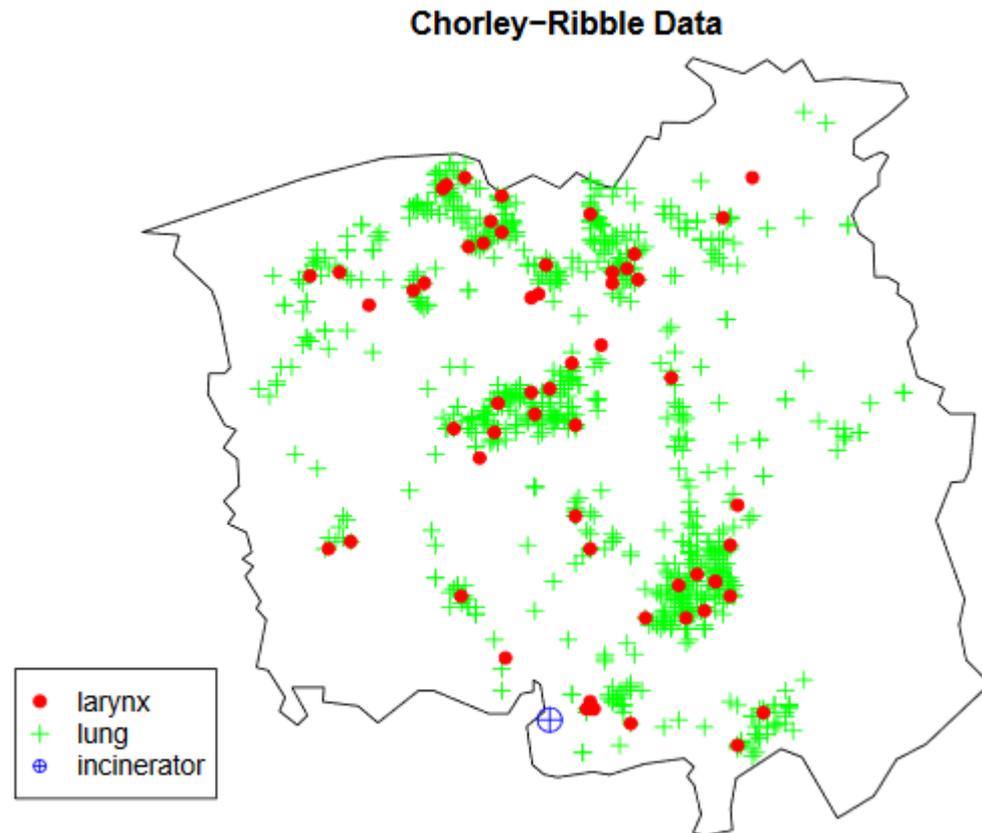
- Pour tout B dans A,  $N_B$  suit une loi de Poisson de paramètre  $\lambda(B)$
- Si B et C sont disjoints,  $N_B$  et  $N_C$  sont deux variables indépendantes



# Exemple

## Chorley-Ribble data

- Localisations de cas de cancer du larynx ( $n = 58$ ) et des poumons ( $n = 978$ ) survenus entre 1974 et 1983 et localisation d'un incinérateur désaffecté
- Question : Après la prise en compte de la variation spatiale de la population à risque (approché par les cancers du poumon), existe-t-il une augmentation de l'incidence des cancers du larynx près de l'incinérateur ?



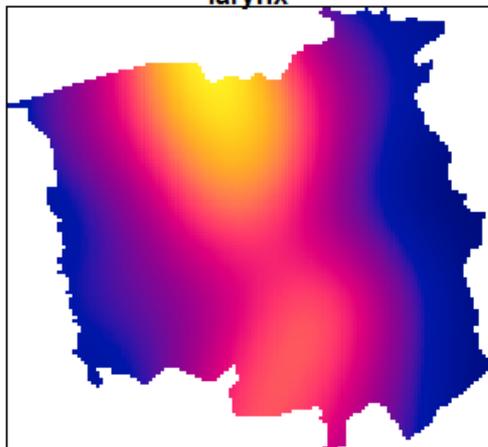


# Exemple

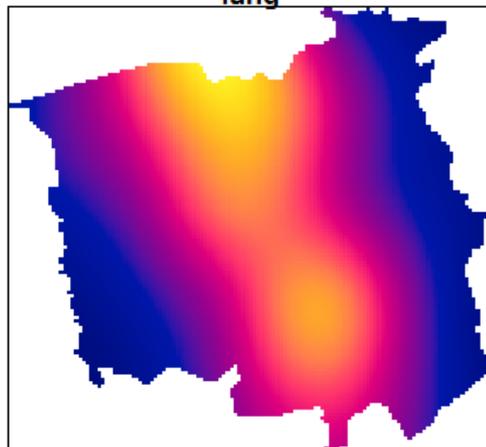
## Chorley-Ribble data

density(split(chorley))

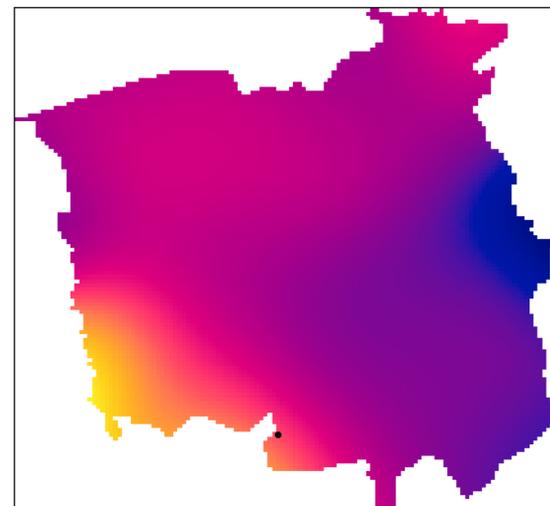
larynx



lung



Densité relative des  
cancers du larynx



Diggle, P. (1990) A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Soc. Series A* **153**, 349-362.



# Données latticielles



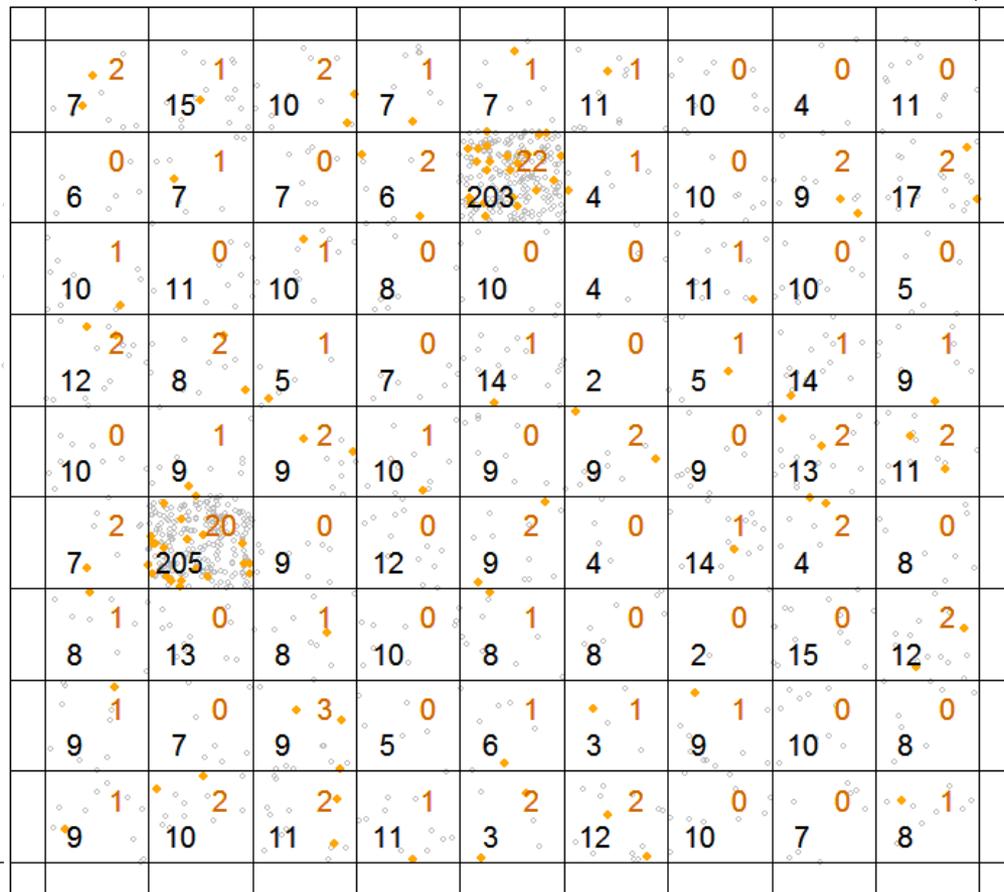
# Définition

- sites  $s$  d'observation
  - fixés et déterministes
  - unités spatiales régulières ou non
- $X_s$  est une mesure agrégée sur l'unité  $s$
- Exemples
  - imagerie
  - épidémiologie
- Questions d'intérêt
  - modèles explicatifs
  - corrélation spatiale
  - restauration d'images



# SIG en santé

## Cartographie et Données Sources



Agrégation ?

Population à  
risque

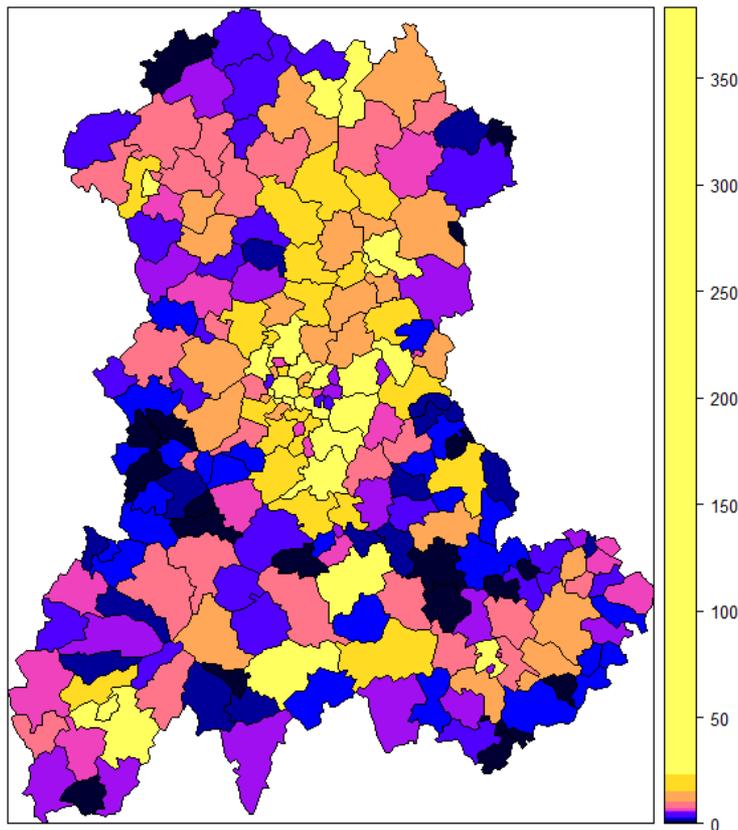
Données  
ponctuelles

Données  
groupées

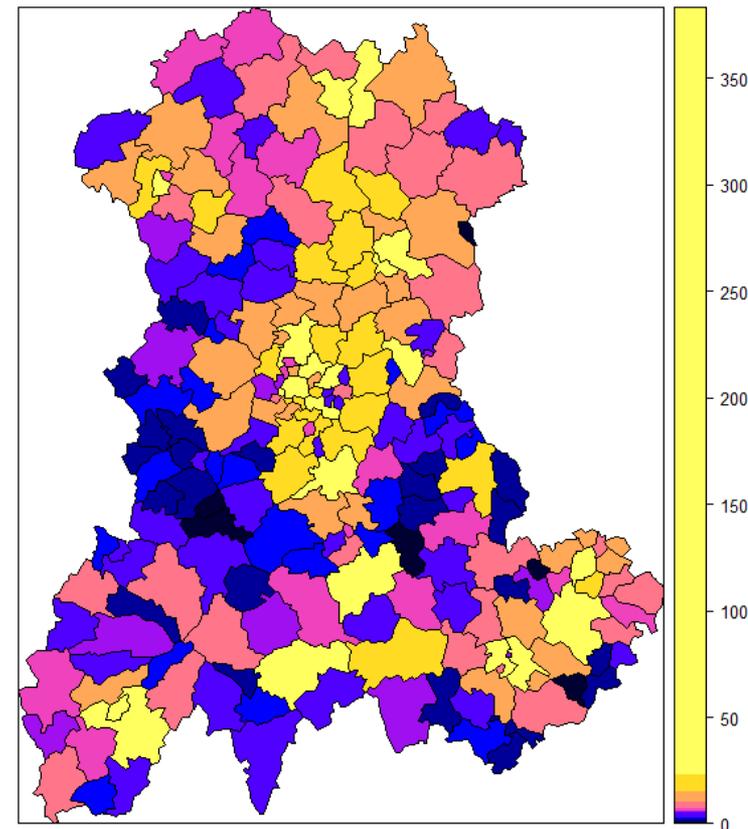


# Exemple

## Malformations congénitales en Auvergne



Cas observés



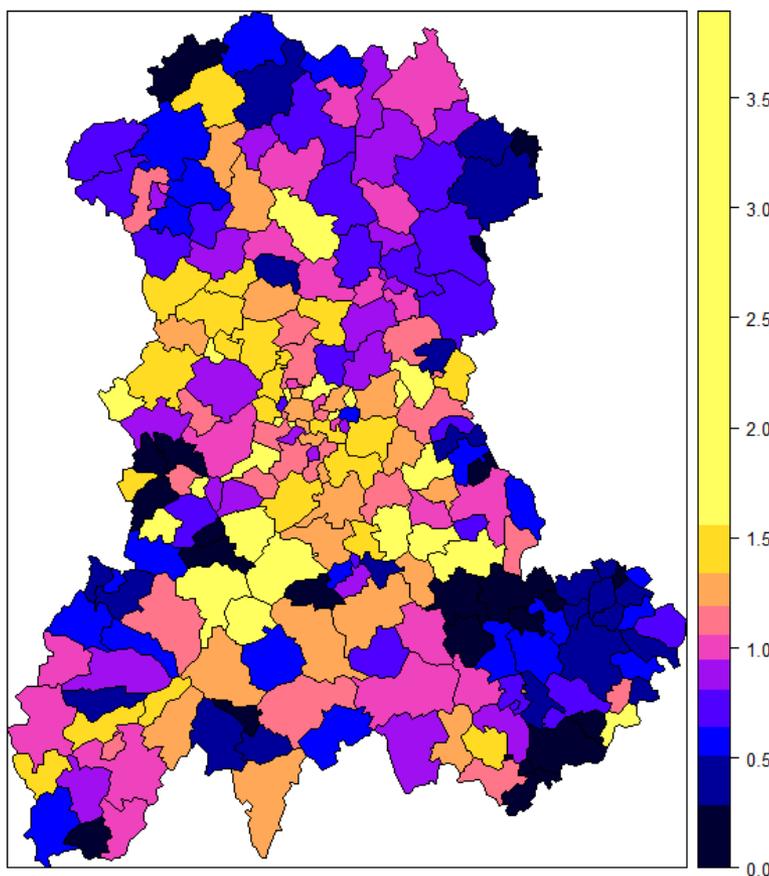
Cas attendus

Toutes malformations confondues, hors anomalies chromosomiques et syndromes pléiotropiques à caryotype normal

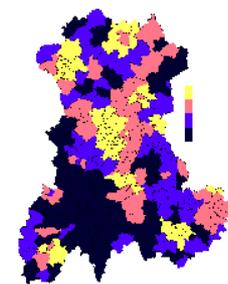


# Exemple

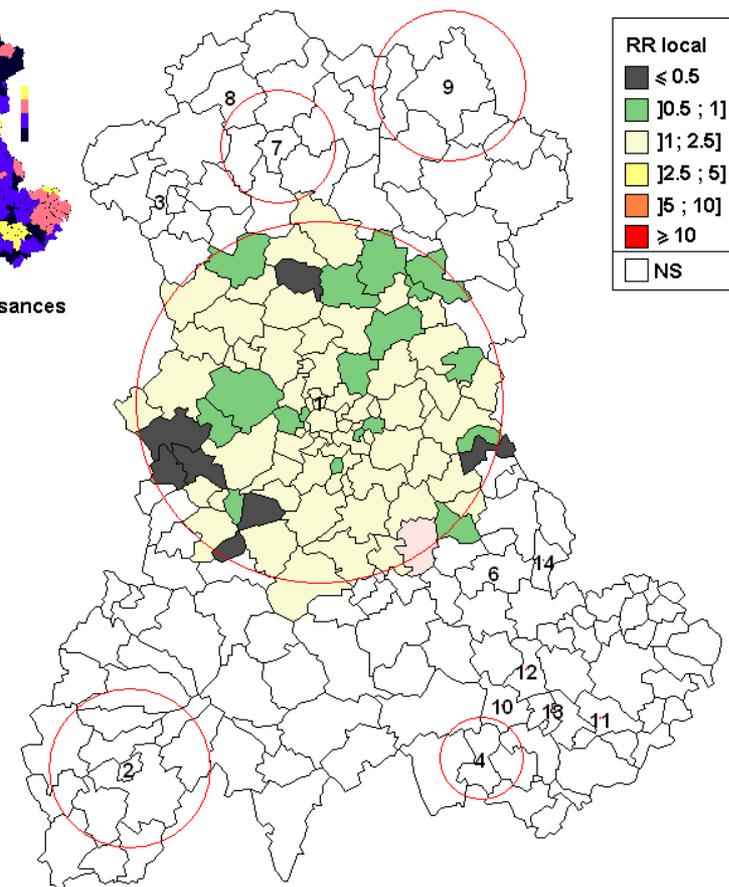
## Malformations congénitales en Auvergne



Cas observés/Cas attendus



densité de naissances



Scan Spatio temporel

1999 à 2006 : Toutes malformations confondues, hors anomalies chromosomiques et syndromes pléiotropiques à caryotype normal



# Statistiques descriptives

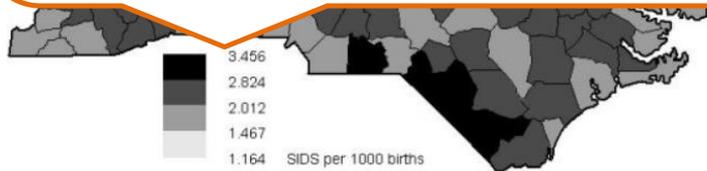
## Disease mapping

Berke O: **Exploratory disease mapping: kriging the spatial risk function from regional count data.** *Int J Health Geogr* 2004, **3**:18

Une **carte choroplèthe** (du grec χώρος : « zone/région » et πληθαίν : « multiple ») est une carte thématique où les régions sont colorées ou remplies d'un motif qui montre une mesure statistique

"choroplèthe." *Wikipédia, l'encyclopédie libre*. 14 avr 2014, 15:51 UTC. 7 oct 2014, 10:15

[http://fr.wikipedia.org/wiki/Carte\\_choropl%C3%A8the#cite\\_note-cheysson1990-3](http://fr.wikipedia.org/wiki/Carte_choropl%C3%A8the#cite_note-cheysson1990-3)



**Choropleth map of Bayesian smoothed SIDS mortality rates of North Carolina.**

Choropleth map of the empirical Bayesian smoothed mortalities per 1000 live births from 100 counties of North Carolina, 1974–1984. Cut points of the grey scale shading are the 5%, 50% and 95% quantiles of the empirical distribution.

Berke *International Journal of Health Geographics* 2004 **3**:18

doi:10.1186/1476-072X-3-18

[Download authors' original image](#)

Un isoplèthe, une isoligne ou isarithme, est une ligne joignant des points d'égale valeur sur une carte. Elle sépare des zones de faibles valeurs et des zones de valeurs plus élevées.

"Isoplèthe." *Wikipédia, l'encyclopédie libre*. 14 avr 2014, 15:51 UTC. 7 oct 2014, 10:15

<http://fr.wikipedia.org/w/index.php?title=Isopl%C3%A8the&oldid=102912965>



**Isopleth map from kriging the smoothed SIDS mortality rates of North Carolina.**

Isopleth map based on kriging predictions of the empirical Bayesian smoothed annual SIDS mortality rates per 1000 live births from 100 counties of North Carolina, 1974–1984.

Berke *International Journal of Health Geographics* 2004 **3**:18

doi:10.1186/1476-072X-3-18

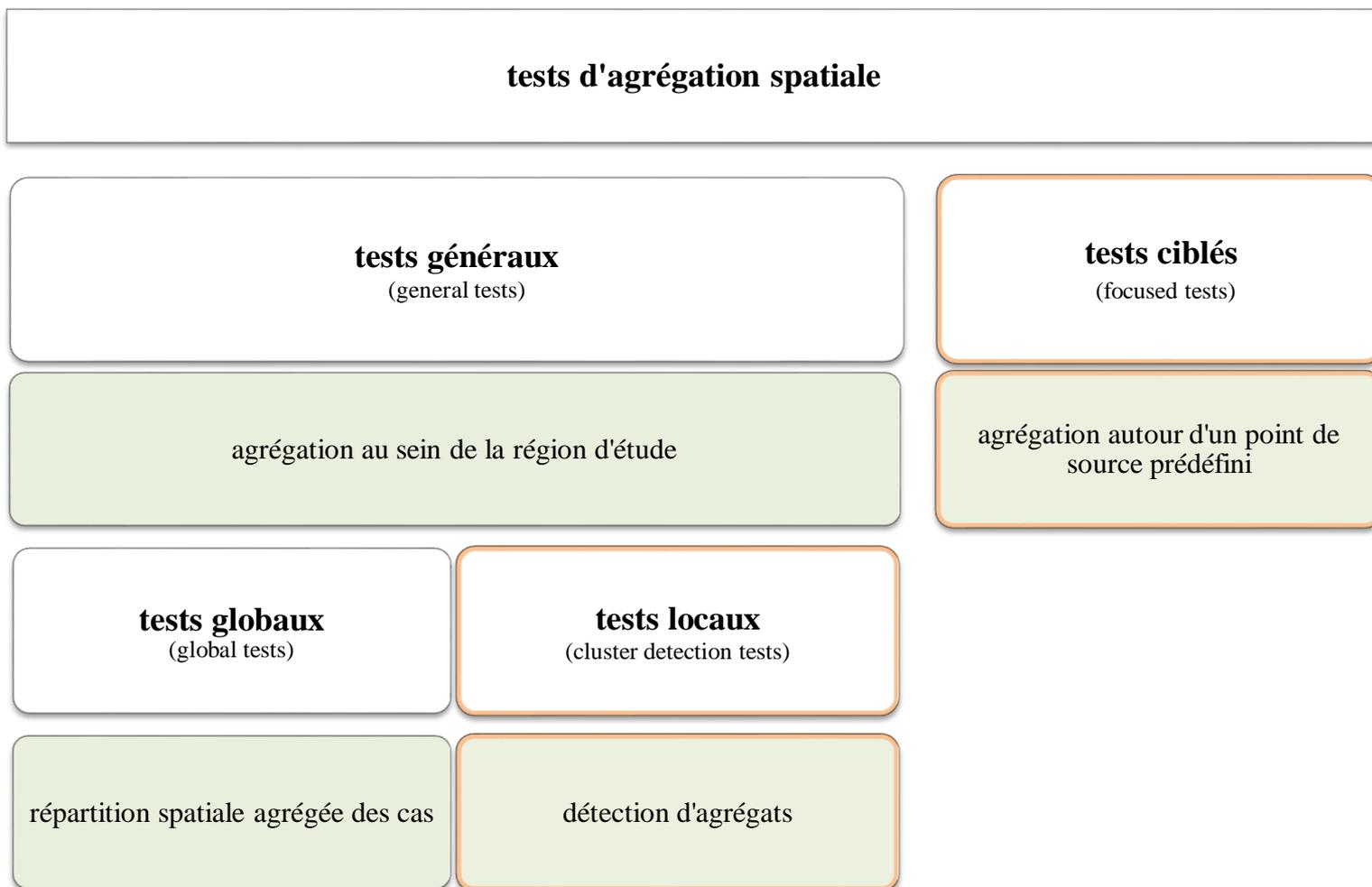
[Download authors' original image](#)

Tango T: *Statistical Methods for Disease Clustering*. New York, NY: Springer New York; 2010. [*Statistics for Biology and Health*]



# Statistiques Inférentielles

## Classification







# Statistiques Inférentielles

## Cluster Detection Tests – Test locaux de détection d'agrégats

- Statistiques de balayage
  - Openshaw S, Charlton M, Wymer C, Craft A: **A mark 1 geographical analysis machine for the automated analysis of point data sets.** *Int J Geogr Inf Syst* 1987, **1**:335–358.
  - Kulldorff M: **a spatial scan statistic.** *Commun Stat Theor M* 1997, **26**:1481–1496.
  - Un agrégat potentiel (le plus vraisemblable)
  - Une statistique = un test
  - Inférence de Monte Carlo



# Statistiques Inférentielles

## Scan Spatial de Kulldorff

### Fonction de vraisemblance pour toutes les zones $Z_i$ .

La probabilité d'observer  $c$  cas est telle que :

$$Prob(c|Z_i \subset G) = \frac{e^{-\left(\pi P_{Z_i} + \delta(P - P_{Z_i})\right)} \times \left(\pi P_{Z_i} + \delta(P - P_{Z_i})\right)^c}{c!}$$

Fonction  $f(x)$  de « densité de probabilité » des cas dans  $G$  sachant  $Z_i$  telle que :

$$f(x) = \frac{\pi P_x}{\pi P_{Z_i} + \delta(P - P_{Z_i})} I(x \in Z_i) + \frac{\delta P_x}{\pi P_{Z_i} + \delta(P - P_{Z_i})} I(x \in \bar{Z}_i)$$

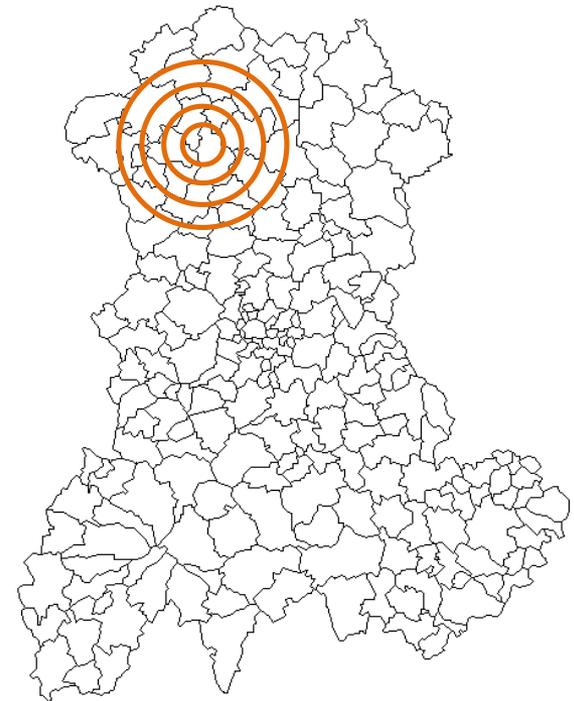
$I(x \in Z_i)$  et  $I(x \in \bar{Z}_i)$  : indicatrices = 1 si  $x \in Z_i$  (respectivement  $x \in \bar{Z}_i$ ) et 0 si non.

La fonction de vraisemblance pour  $Z_i$  est égale à :

$$L(Z_i, \pi, \delta) = Prob(c) \times \prod_{j=1}^c f(x_j)$$

$c$  : dénombrement des cas

$P$  : taille de la population à risque





# Conclusion

## La statistique spatiale

- développement rapide
- nombreux domaines d'application en santé
  - diagnostic, thérapeutique
  - médecine légale
  - veille sanitaire
  - ...
- données à structure complexe, constituant des classes hétérogènes
- objectifs variés
  - agrégation spatiale des cas – détection d'agrégats
  - analyses écologiques
  - modélisation spatiale des maladies infectieuses
  - ...



# MERCI

Université Clermont Auvergne, Faculté de Médecine (R2, 5e étage),  
28 Place Henri Dunant - BP 38, 63001 CLERMONT-FERRAND Cedex 1

03 mars 2017

**INSTITUT PASCAL** - UMR 6602 UCA/CNRS/SIGMA - *T.G.I. (Thérapies Guidées par l'Image)*