



Sciences Economiques & Sociales de la Santé
& Traitement de l'Information Médicale

www.sesstim-orspaca.org

Dr. Aline GUTTMANN

Clermont Université d'Auvergne

ISIT UMR 6284 CNRS UDA

Méthodes statistique en épidémiologie spatiale

octobre 2014



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



Méthodes Statistiques en Épidémiologie Spatiale

Principes et Évaluation

Clermont Université, Université d'Auvergne, ISIT, BP 10448,
F-63000 Clermont-Ferrand

10 octobre 2014



ISIT

UMR 6284 CNRS UDA



Introduction



Définition

“Spatial epidemiology is a rapidly growing field of research concerned with the analysis of the geographical distribution of disease. This principally involves mapping of the location of disease cases and the analysis of the mapped data using spatial statistical methods.”

Andrew B. Lawson

Lawson AB: *Statistical Methods in Spatial Epidemiology*. 2nd edition. Wiley; 2006.



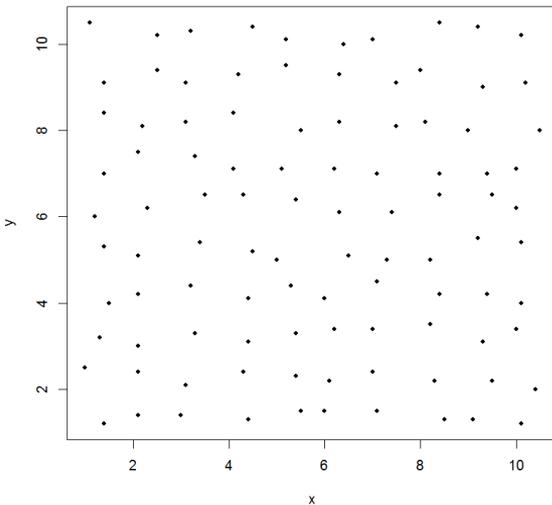
Répartition Spatiale des cas

Processus spatial ponctuel – spatial point process

Répartition spatiale

régulière

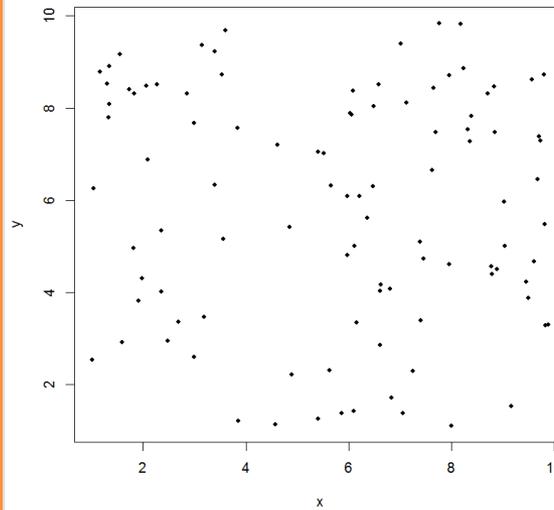
Effet de voisinage :
répulsion



complètement aléatoire

Aucun effet de
voisinage

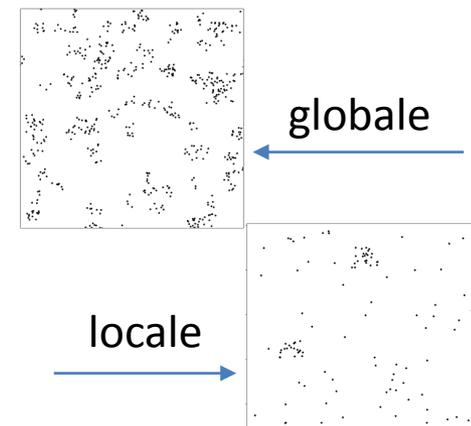
PPH



agrégée

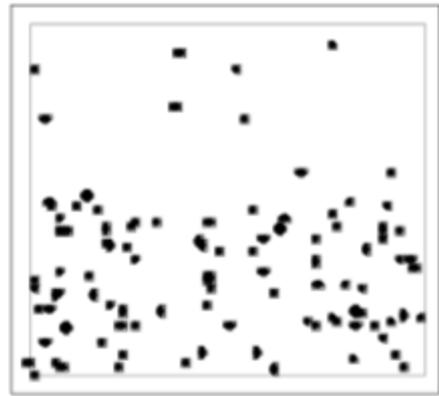
Effet de voisinage :
attraction

PPI

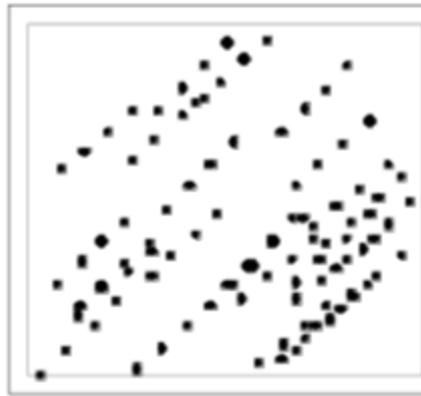


Un processus est dit homogène (stationnaire) s'il est invariant par translation, c'est-à-dire si ses propriétés ne varient pas d'une position à l'autre de l'espace.

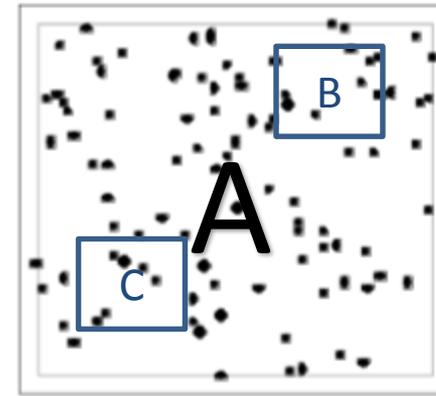
Un processus est dit isotrope s'il est invariant par rotation, c'est-à-dire si ses propriétés ne varient pas avec l'orientation de l'espace.



Non stationnaire



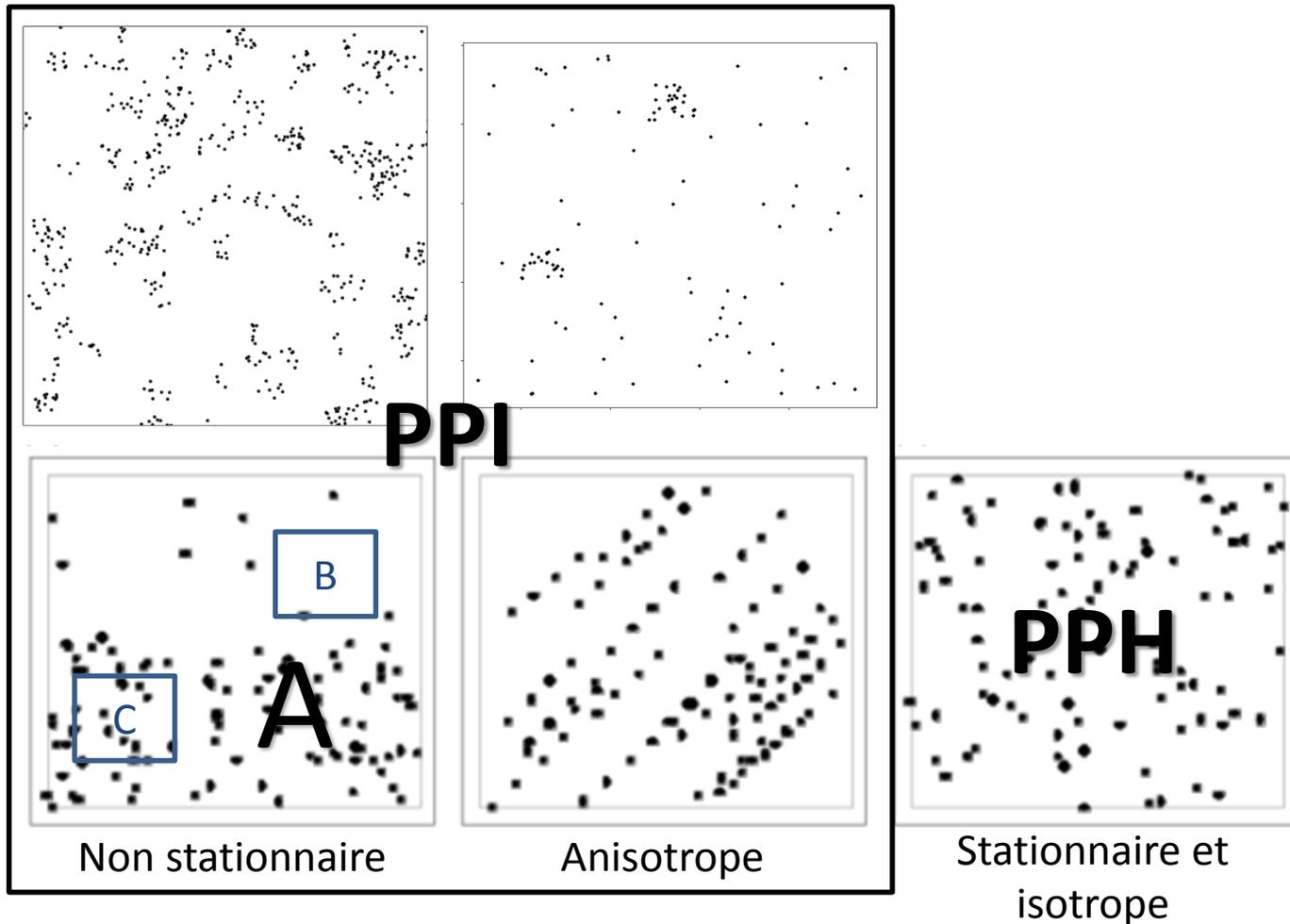
Anisotrope



Stationnaire et
isotrope

Processus de Poisson Homogène d'intensité λ

- 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



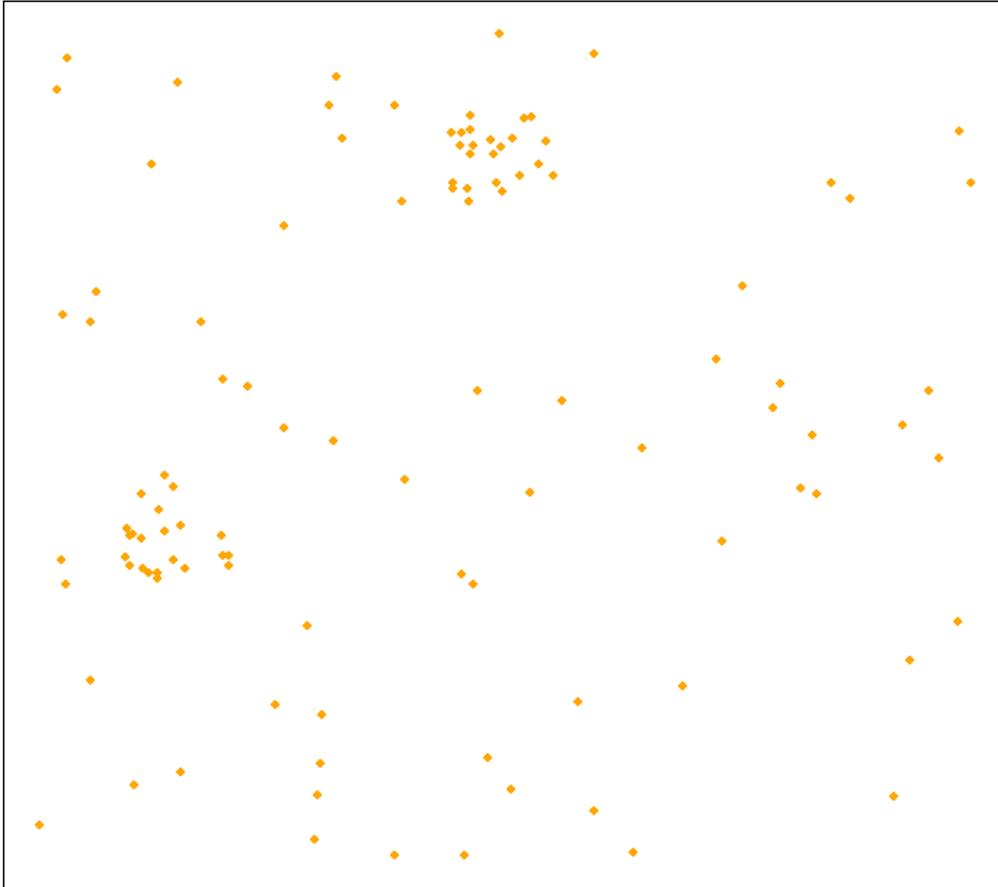
Processus de Poisson Inhomogène d'intensité $\lambda(\cdot)$

- 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



SIG

Cartographie et Données Sources

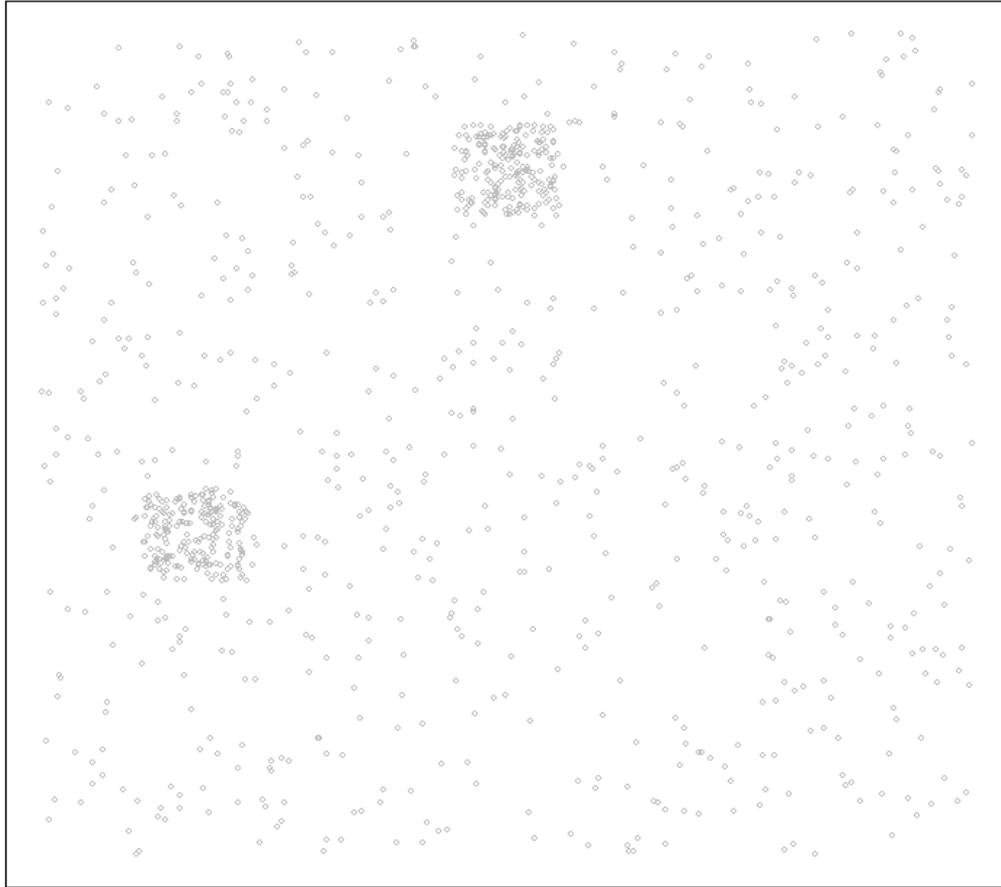


Agrégation ?



SIG

Cartographie et Données Sources



Agrégation ?

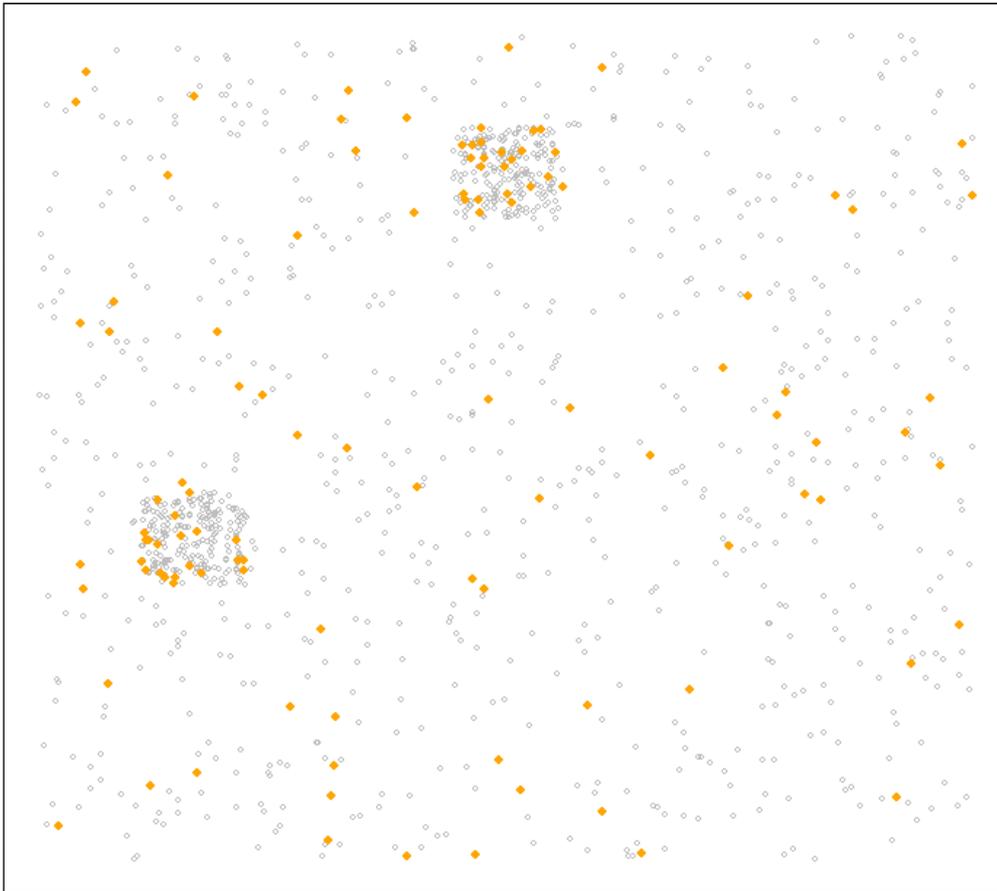


Population à
risque



SIG

Cartographie et Données Sources



Agrégation ?

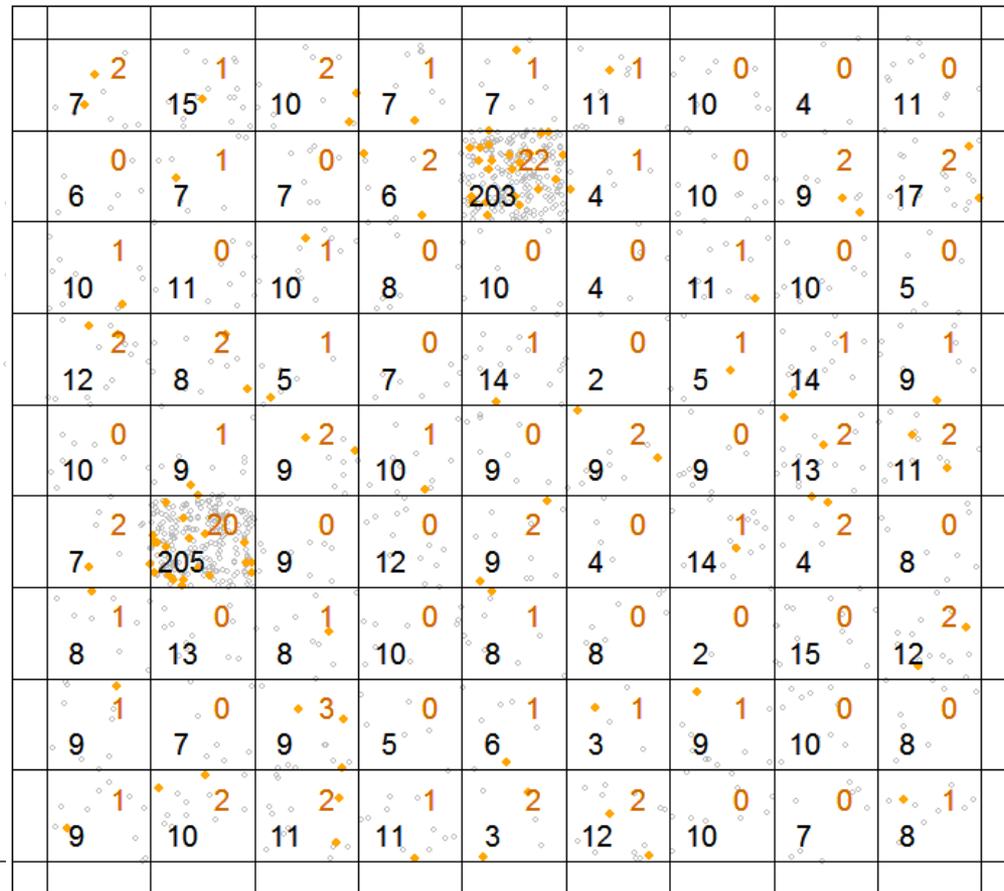
Population à
risque

**Données
ponctuelles**
toute la population
ou
un échantillon (cas-
contrôle)



SIG

Cartographie et Données Sources



Agrégation ?

Population à
risque

**Données
ponctuelles**
toute la population
ou
un échantillon (cas-
contrôle)

**Données
groupées**



Statistiques Spatiales

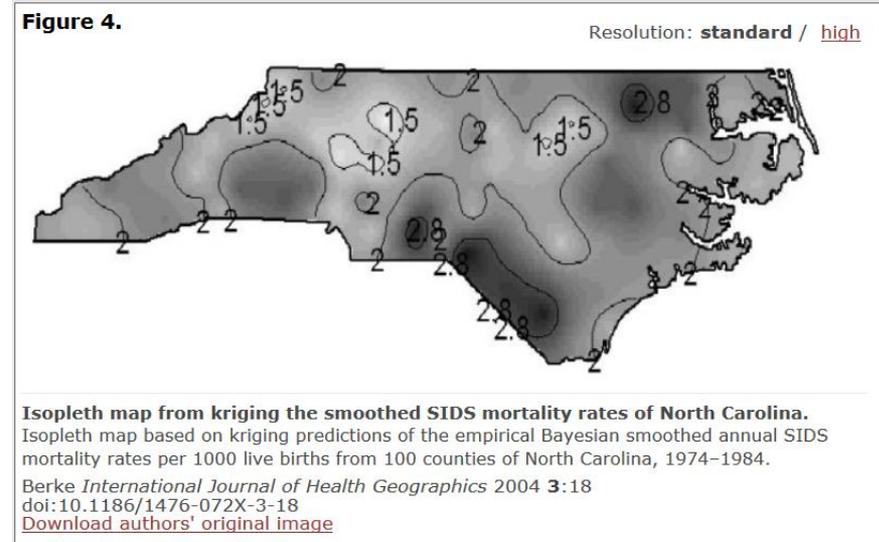
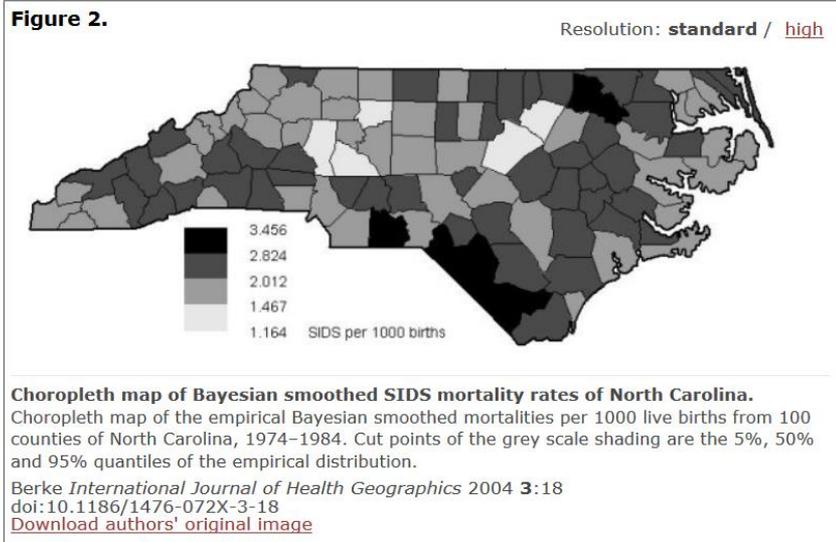
Analyse de l'agrégation sur données groupées



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.



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



Statistiques descriptives

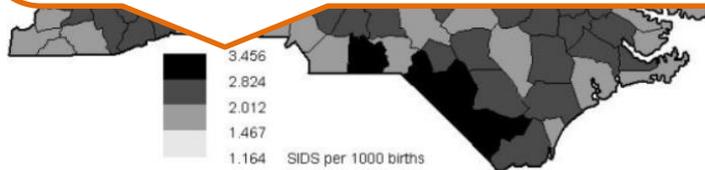
Disease mapping

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



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

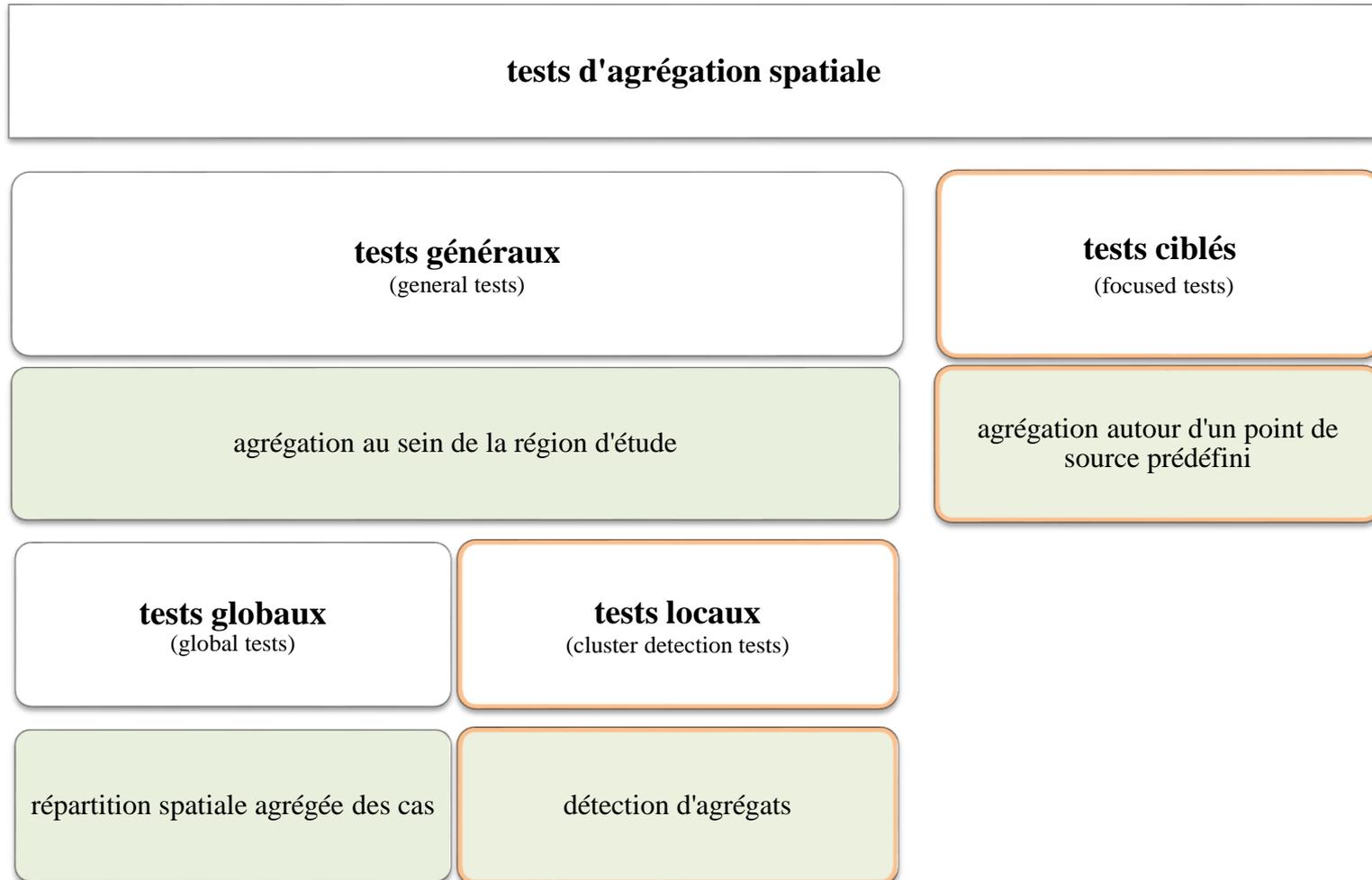
[Download authors' original image](#)

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



Statistiques Inférentielles

Classification





Statistiques Inférentielles

Tests globaux

Autocorrélation spatiale

Moran's I test

$$I = \frac{N \times \sum_{ij} w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{2(\sum_{ij} w_{ij}) * \sum_{i=1}^N (Y_i - \bar{Y})^2}$$

Où $Y_i = \frac{c_i}{p_i}$ est la proportion de cas observés

dans l'US i , $\bar{Y} = \frac{\sum_{i=1}^N c_i}{N p_i}$ est la *moyenne des*

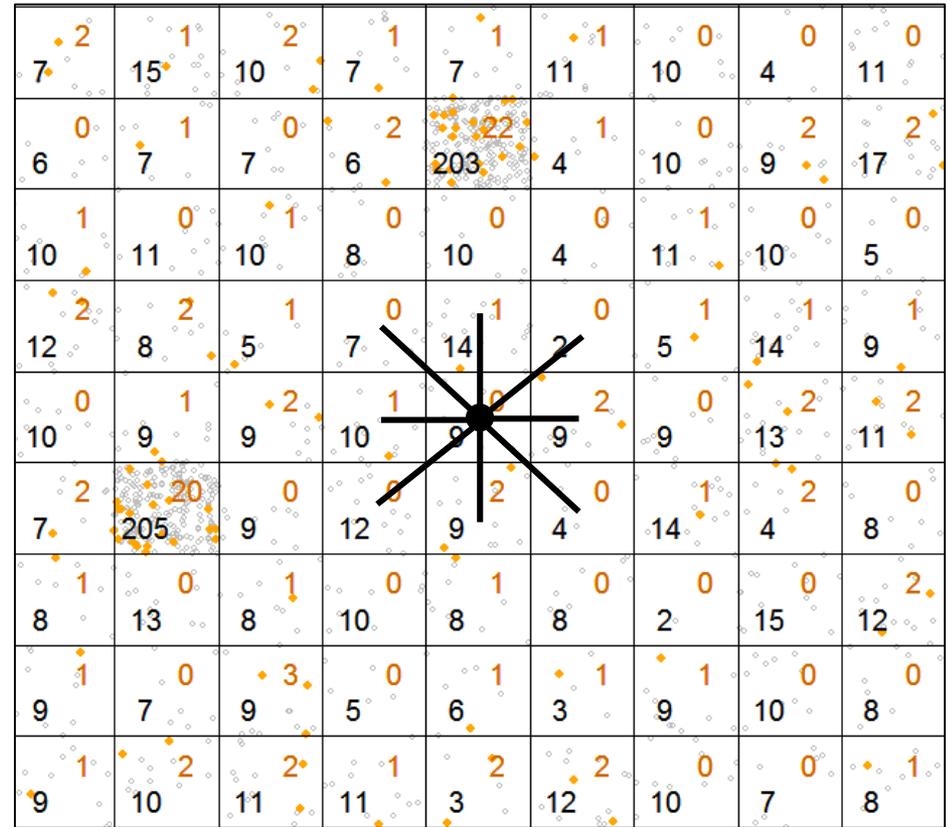
proportions de cas sur l'ensemble de la région

et w_{ij} est l'élément de la matrice de proximité

correspondant au couple d'US i et j .

Inférence par méthode de Monte Carlo

Moran PAP (1948) *The interpretation of statistical maps*. J Roy Stat Soc, Series B 10:243-251



Relations spatiales (voisinage) entre US : W_{ij}

- Binaire (0/1)
- quantitative (distance $W_{ij} = e^{-\left(\frac{d_{ij}}{r}\right)}$)
- Intermédiaire (distance maximum)



Statistiques Inférentielles

Focused tests – tests locaux avec point de source prédéfini

Test de Stone

Les US sont triées de l'US₁ (source) à l'US_n (US la plus éloignées de la source)

$$H_0: \theta_1 = \theta_2 = \dots = \theta_n = \lambda$$

$$H_1: \theta_1 \geq \theta_2 \geq \dots \geq \theta_n$$

$$T = \max_{1 \leq j \leq n} \frac{\sum_{i=1}^j c_i}{\sum_{i=1}^j e_i}$$

- Attention : utiliser « en aveugle » = une source prédéfinie \neq zone à risque accru



Statistiques Inférentielles

Cluster Detection Tests – Test locaux de détection d'agrégats

- Statistiques locales d'autocorrélation spatiale
 - Version locale des statistiques globales (décomposition)
 - 1 statistique par US → une inférence par US
 - Inférence de Monte Carlo - Ajustement des p-values
- 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 :

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

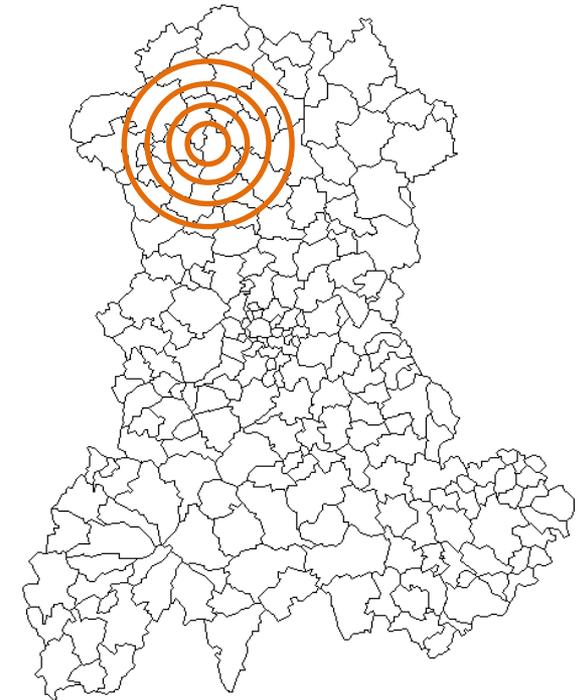
$I(x \in Z_i)$ et $I(x \in \bar{Z}_i)$: indicatrices = 1 si $x \in Z_i$ (respectivement $x \in \bar{Z}_i$) et 0 si non.

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





Principes et enjeux

- Quel crédit accorder à la conclusion du test ?
 - Lorsqu'aucun agrégat n'est détecté : manque de puissance ?
 - Lorsque un agrégat est détecté :
 - fausse alerte ?
 - l'agrégat détecté = agrégat réel?

Évaluation = étude de simulation

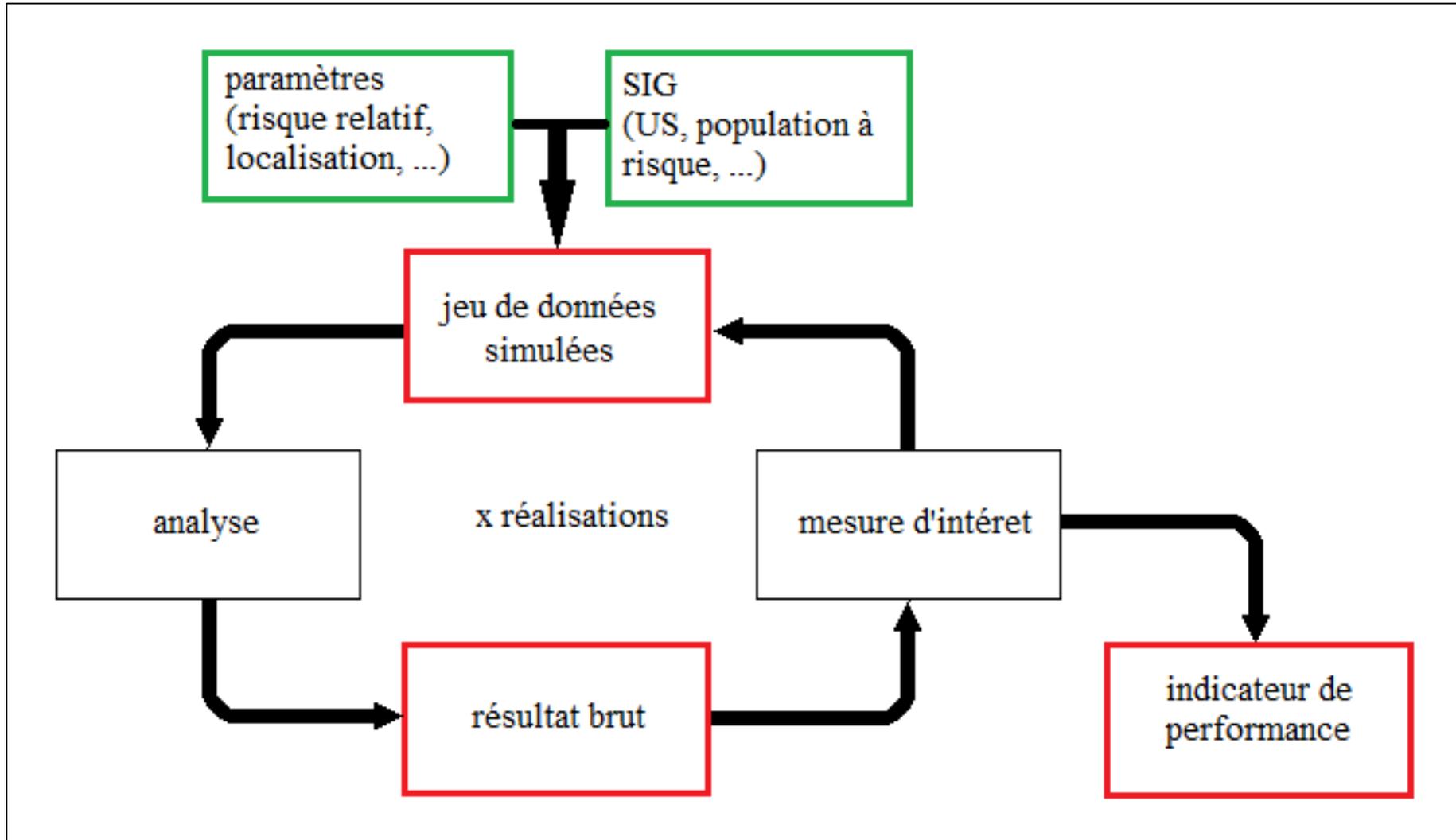
Problématique

- Pas de consensus sur la méthodologie
 - Protocoles de simulation
 - Indicateurs de performance
- Comparabilité des études
- Multiplicité des indicateurs
- Interprétation / généralisation des résultats



Études de simulation

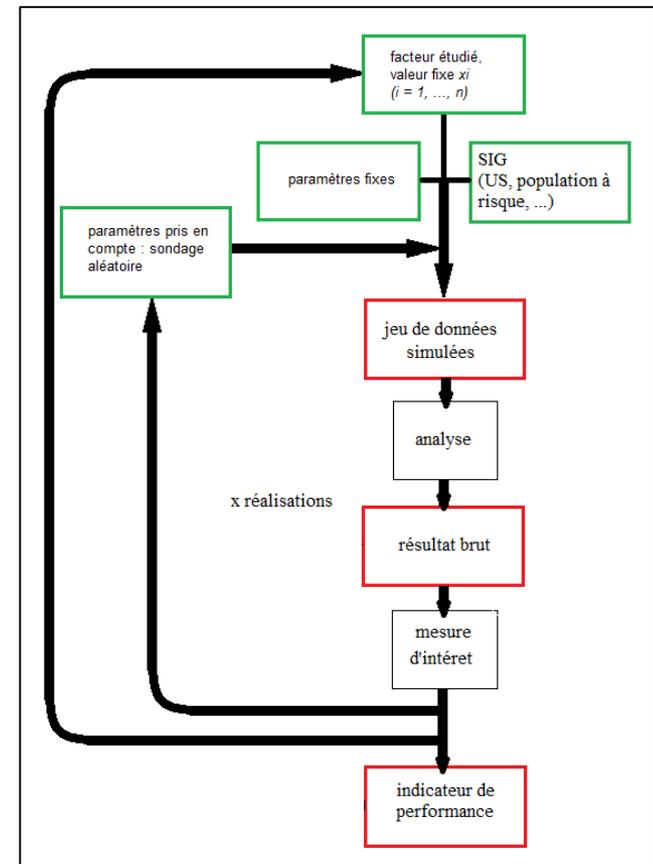
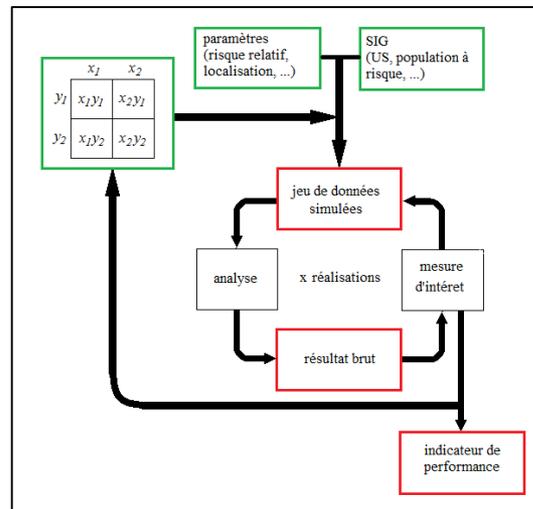
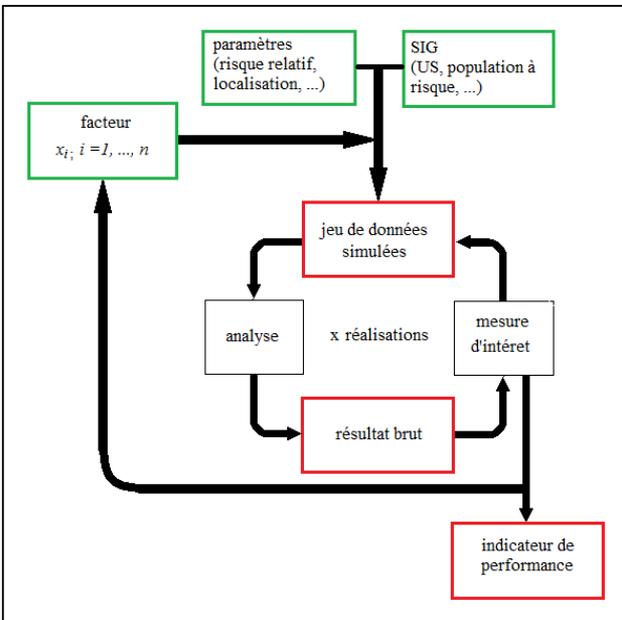
Schéma classique





Études de simulation

Analyse d'un ou plusieurs facteurs

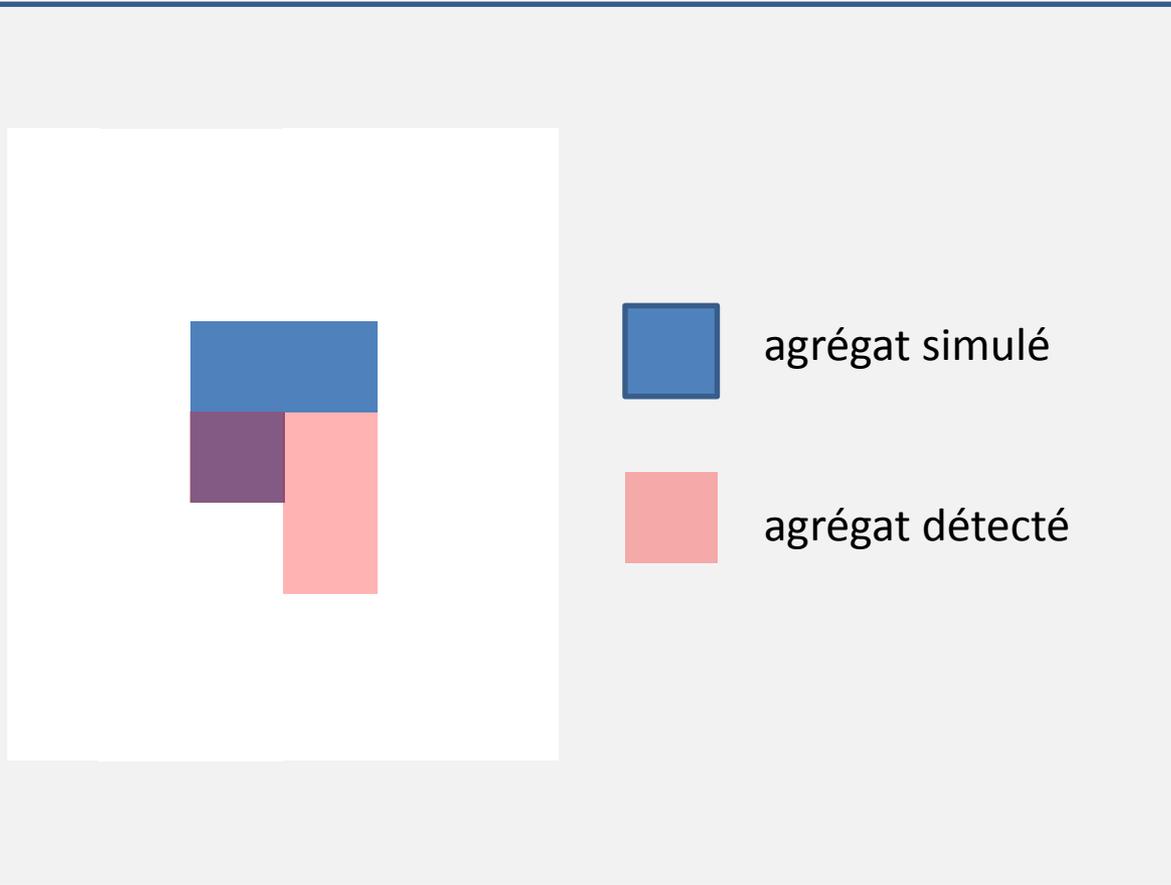




Études de simulation

Indicateurs de performance

...Les agrégats



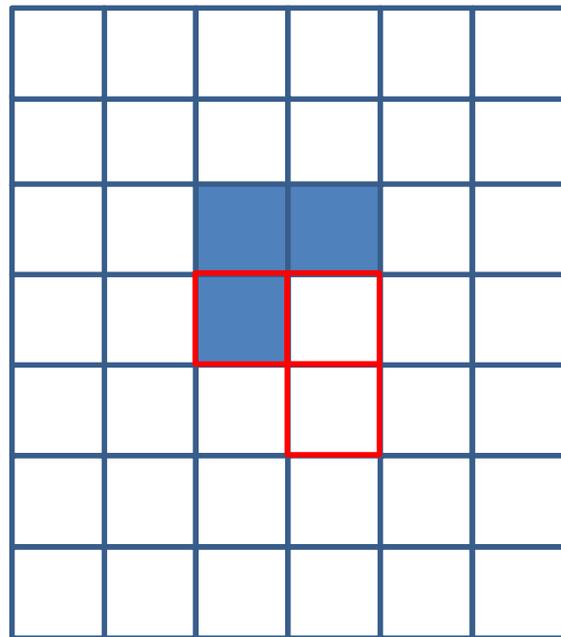
- Puissance usuelle
 - Agrégat détecté : 0/1
 - Somme(agrégats détectés)/nombre de réalisation
- Puissance conditionnelle
 - Agrégat simulé = agrégat détecté : 0/1
 - Agrégat simulé \in agrégat détecté : 0/1
 - Centroïde de l'agrégat simulé \in agrégat détecté : 0/1
 - ...



Études de simulation

Indicateurs de performance

...Les US



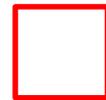
Unité Spatiale (US)



US incluse dans l'agrégat simulé



US hors de l'agrégat simulé



US incluse dans l'agrégat détecté

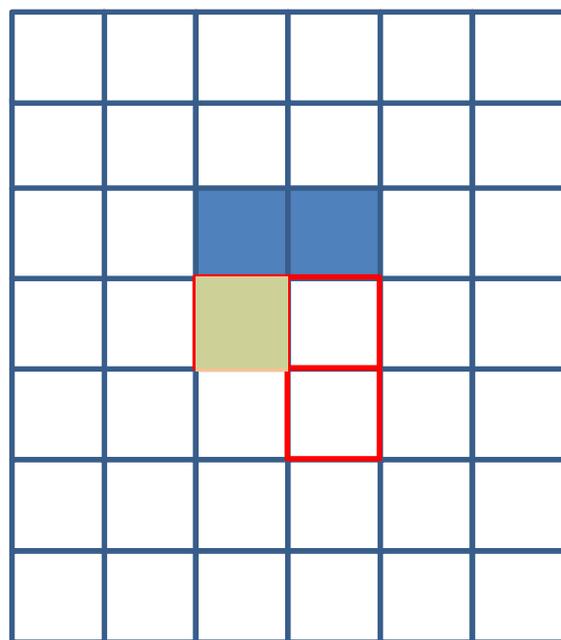
	US dans l'agrégat simulé	US hors agrégat simulé
US dans l'agrégat détecté	VP	FP
US hors agrégat détecté	FN	VN



Études de simulation

Indicateurs de performance

...Les US



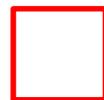
Unité Spatiale (US)



US incluse dans l'agrégat simulé



US hors de l'agrégat simulé



US incluse dans l'agrégat détecté

$$Se = \frac{VP}{VP + FN}$$

Probabilité qu'une US appartienne à l'agrégat détecté si elle fait partie de l'agrégat simulé.

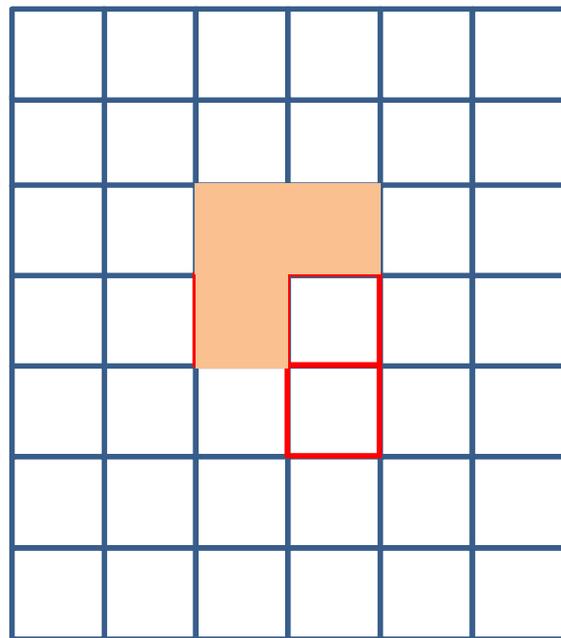
	US dans l'agrégat simulé	US hors agrégat simulé
US dans l'agrégat détecté	VP	FP
US hors agrégat détecté	FN	VN



Études de simulation

Indicateurs de performance

...Les US



Unité Spatiale (US)



US incluse dans l'agrégat simulé



US hors de l'agrégat simulé



US incluse dans l'agrégat détecté

$$Se = \frac{VP}{VP + FN}$$

Probabilité qu'une US appartienne à l'agrégat détecté si elle fait partie de l'agrégat simulé.

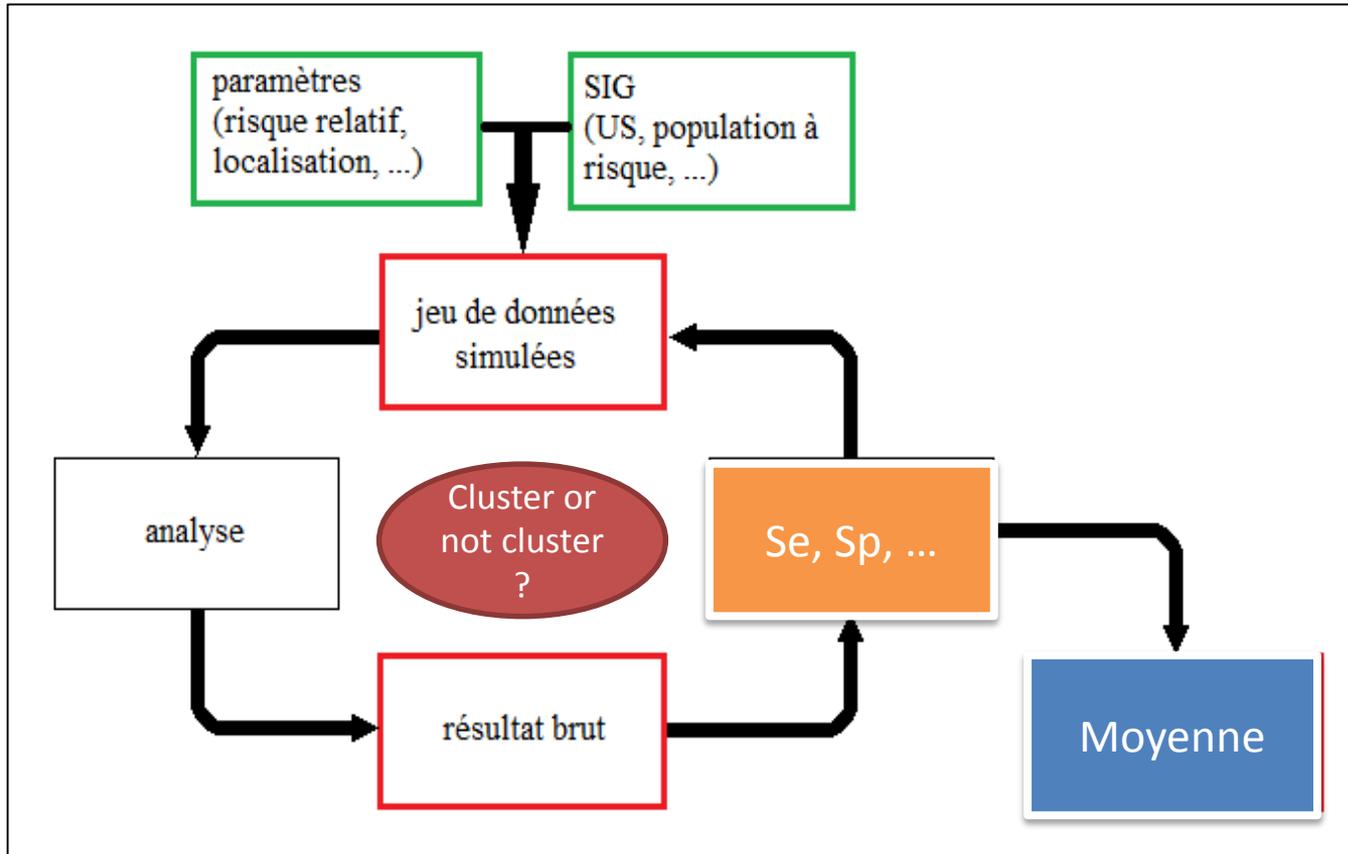
	US dans l'agrégat simulé	US hors agrégat simulé
US dans l'agrégat détecté	VP	FP
US hors agrégat détecté	FN	VN



Études de simulation

Indicateurs de performance

...Les US



➤ Moyenne (parfois IC 95%)

➤ Dénominateur :

➤ Seulement les réalisations aboutissant à la détection d'un agrégat

➤ Toutes les réalisations : quid des réalisations sans détection ?

➤ Non détecté = pas de positif : VP + FP = 0

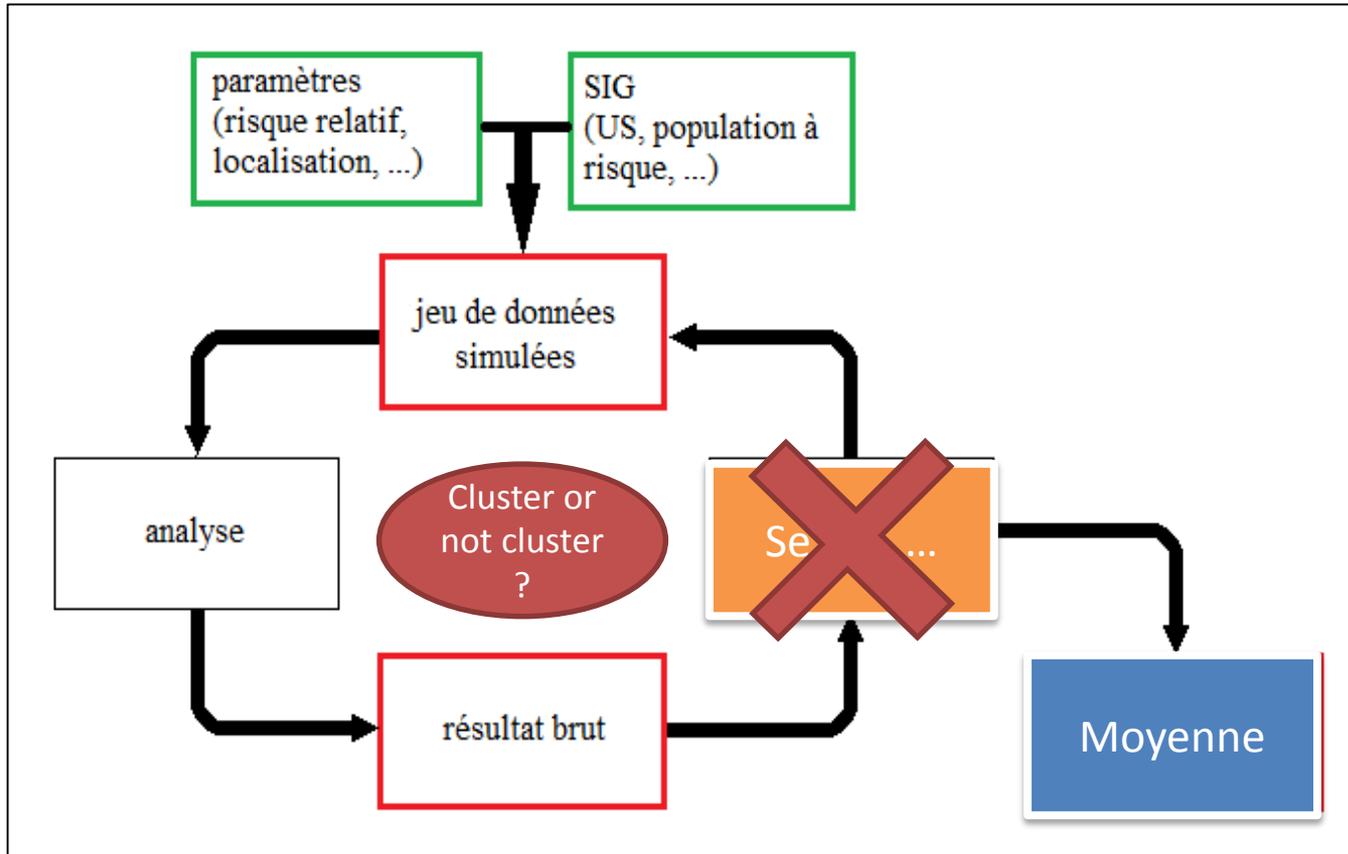
➤ Agrégat vraisemblable



Études de simulation

Indicateurs de performance

...Les US



➤ Moyenne (parfois IC 95%)

➤ Dénominateur :

➤ **Seulement les réalisations aboutissant à la détection d'un agrégat**

➤ Toutes les réalisations : quid des réalisations sans détection ?

➤ Non détecté = pas de positif : VP + FP = 0

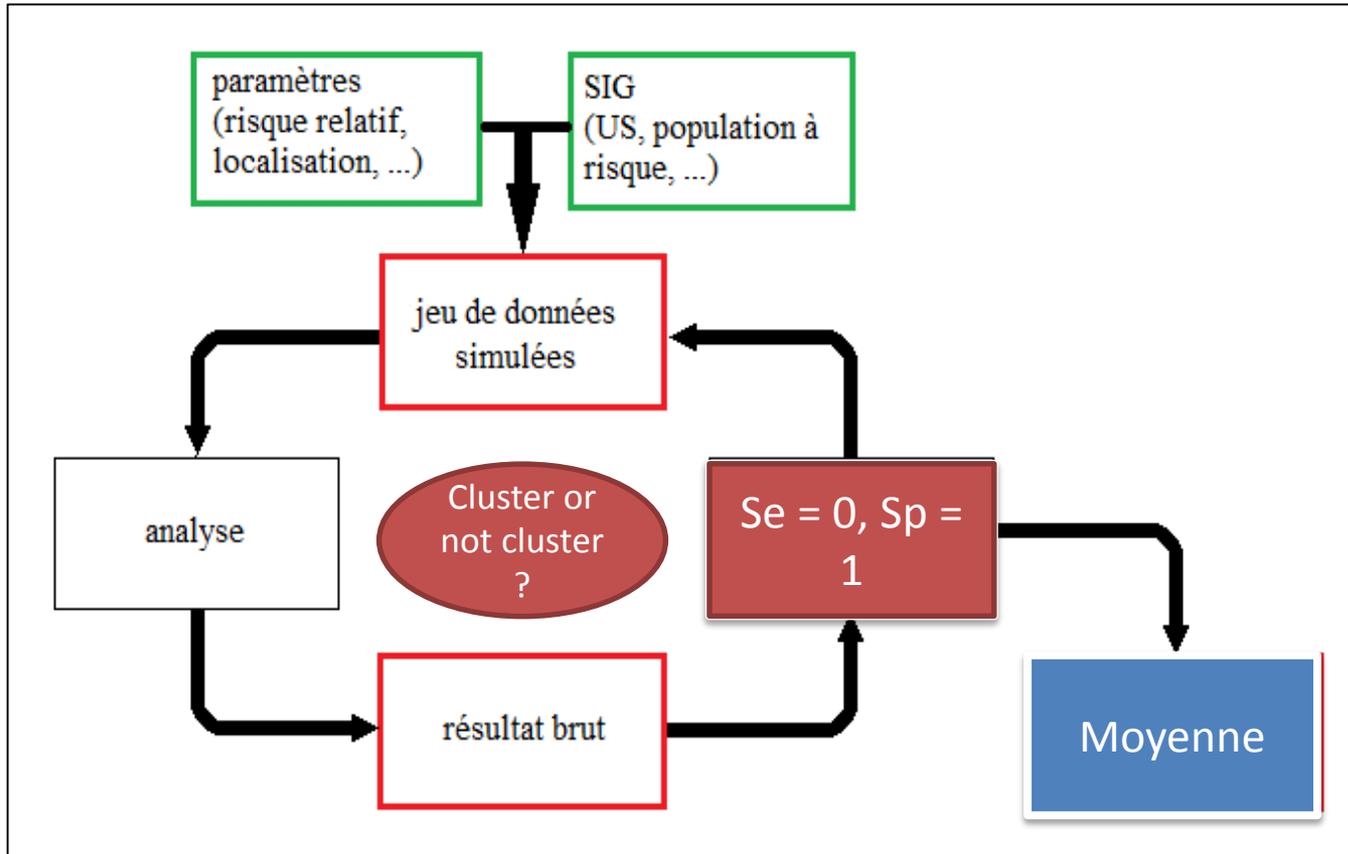
➤ Agrégat vraisemblable



Études de simulation

Indicateurs de performance

...Les US



➤ Moyenne (parfois IC 95%)

➤ Dénominateur :

➤ Seulement les réalisations aboutissant à la détection d'un agrégat

➤ Toutes les réalisations : quid des réalisations sans détection ?

➤ Non détecté = pas de positif : VP + FP = 0

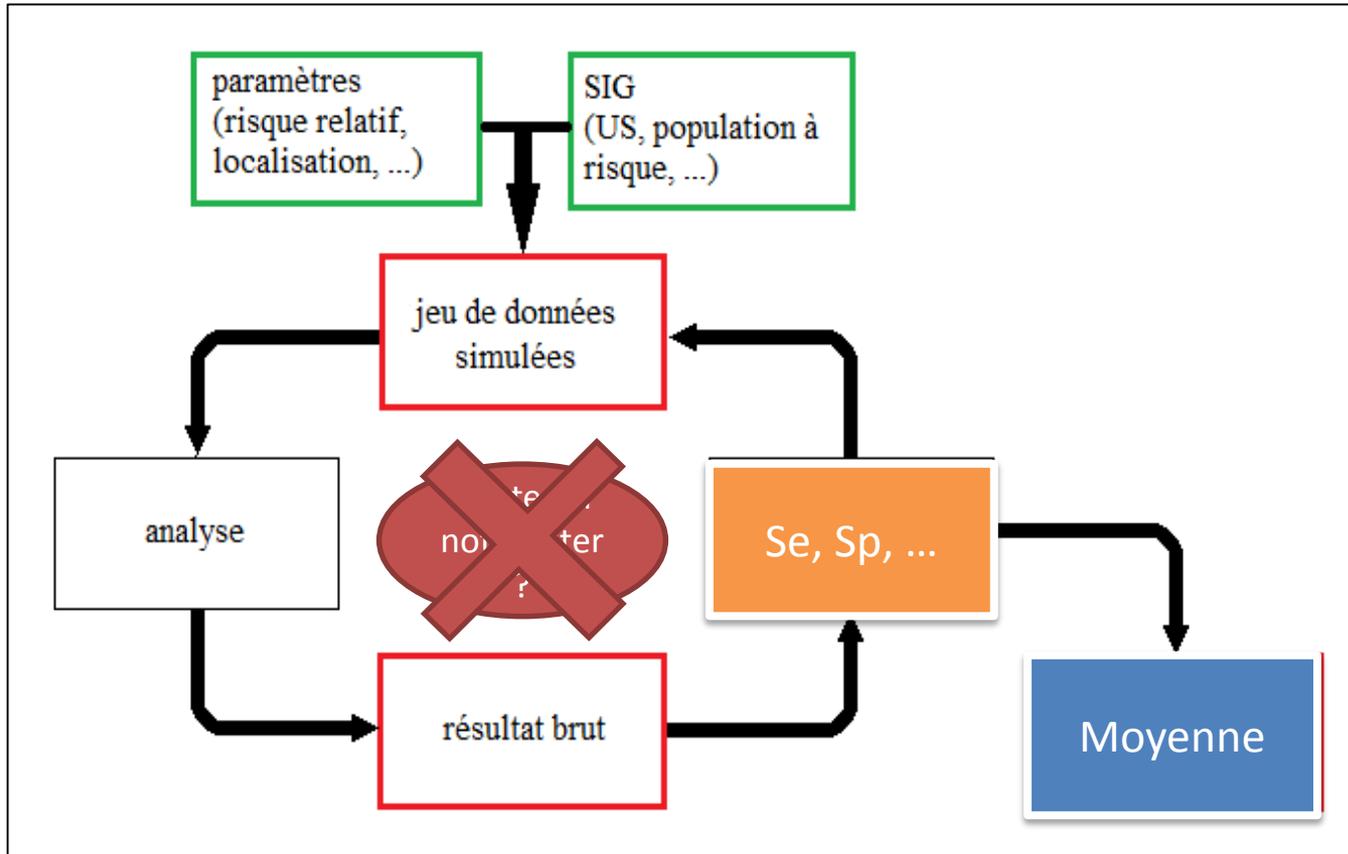
➤ Agrégat vraisemblable



Études de simulation

Indicateurs de performance

...Les US



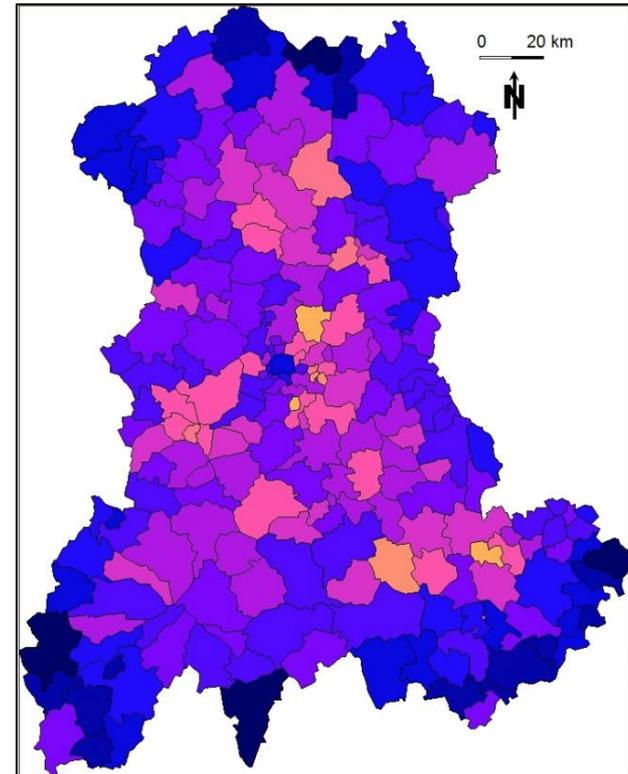
- Moyenne (parfois IC 95%)
- Dénominateur :
 - Seulement les réalisations aboutissant à la détection d'un agrégat
 - Toutes les réalisations : quid des réalisations sans détection ?
 - Non détecté = pas de positif : $VP + FP = 0$
 - Agrégat vraisemblable



Études de simulation

Erreur de Type I – détection à tort

- quantitative : proportion d'erreur de type I = α ?
- qualitative
 - méthodes locales
 - répartition homogène ?
 - effet de bord



Guttman A, Li X, Gaudart J, Gérard Y, Demongeot J, Boire J-Y, Ouchchane L: *Spatial heterogeneity of type I error for local cluster detection tests.* *Int J Health Geogr* 2014, 13:15.



Conclusion

- Méthodes statistiques en épidémiologie spatiale
 - développement rapide
 - objectifs variés
 - agrégation spatiale des cas – détection d'agrégats
 - analyses écologiques
 - modélisation spatiale des maladies infectieuses
- Méthodologie d'évaluation
 - outils classiques
 - études de puissances
 - évaluation des méthodes diagnostic (Se, Sp, ROC...)
 - outils propres aux méthodes locales





Méthodes Statistiques en Épidémiologie Spatiale

MERCI

Clermont Université, Université d'Auvergne, ISIT, BP 10448,
F-63000 Clermont-Ferrand

10 octobre 2014



ISIT

UMR 6284 CNRS UDA