

COURS 7 - statistiques spatiales, géostatistique

Statistiques spatiales et SIG

Arlette Antoni

Université de Bretagne Sud
Université Bretagne Pays de Loire

Année Universitaire 2021 -2022

Approche de la Géostatistique

En géostatistique, il y a deux tâches principales:

- découvrir les règles de dépendance
- faire des prédictions

Notations

- **p variables Z**
dans R^p
- **Implantation s**
 s est dans R^2

$$Z(s), s \in R^2$$

Quelles données sur quel support ? Observation d'une variable Z sur un support s

- **Nature de Z**

- binaire
- qualitative
- discrètes
- continu

- **type d'implantation de s**

- continu
- discret
- en réseaux

- **Semis de points (déjà vu)**

la variable d'intérêt est la position,
les positions s_i sont des variables aléatoires.

- **Données géostatistiques**

les positions s_i sont choisies
 Z est partout défini mais observé en stations
parfois irrégulièrement espacés

- **Réseau régulier ou lattice ou GRID**

réseau discret et régulier formant un maillage

- **Données agrégées sur maillage**

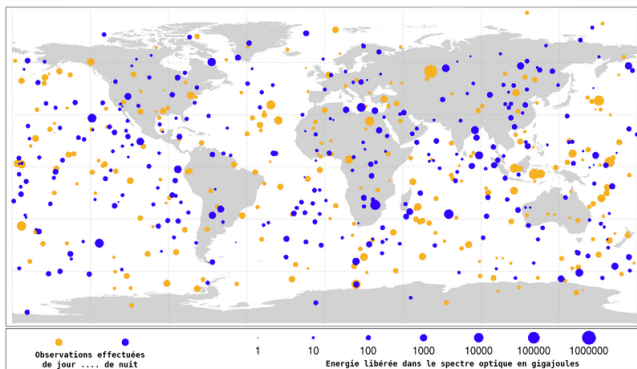
s est un sous ensemble (surface)
 Z est synthétique comme une moyenne..

- **Flux**

bilocalisés s_i , s_j et $Z_{i,j}$

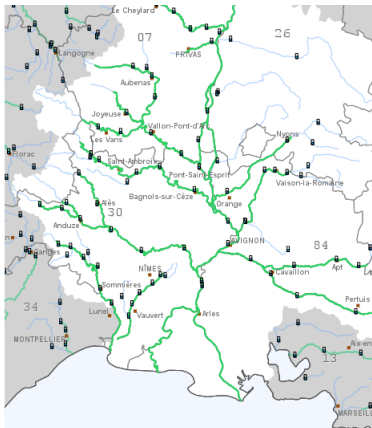
Semis de points

Dans cet exemple de semis de points qui concerne les aperçus d'entrée d'astéroïdes dans l'atmosphère, s'ajoute une variable mesurée en chaque point.

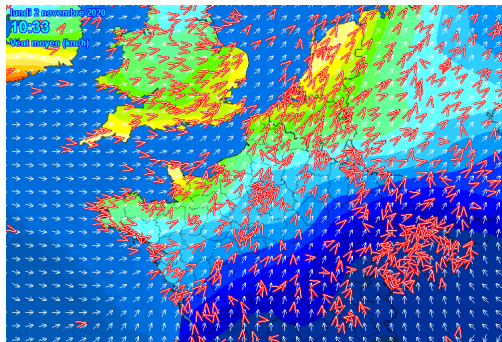


Données géostatistiques

L'image montre la position des capteurs (ou stations) fluviaux:
leur position est connue, mais pas forcément régulièrement espacées

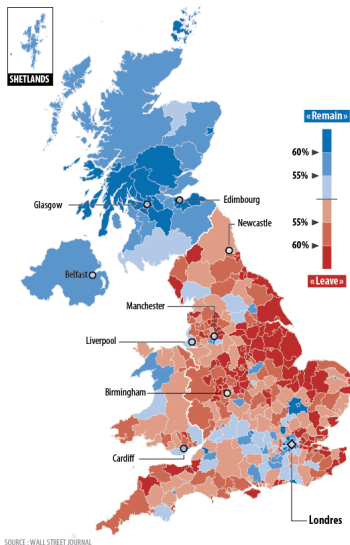


la direction du vent est calculée par déduction d'après image satellitaire



Données agrégées

Pour toutes les circonscriptions du royaume uni, calculs du pourcentage de OUI



Stationnaire

On dit qu'un processus est stationnaire s'il est invariant par translation, c'est à dire si ses propriétés ne varient pas d'un point à l'autre de l'espace.

Propriété essentielle dans la construction des estimateurs de covariance.

Isotropie

On dit qu'un processus est isotrope s'il est invariant par rotation, c'est à dire si ses propriétés ne varient pas en fonction de l'orientation dans l'espace.

Ne dépend que de la distance entre les points pas de la direction entre 2 points

Les propriétés de stationnarité et d'isotropie suffisent pour estimer des statistiques à l'ordre 1 et 2.

Donc on peut construire des estimateurs de covariance.

Quel est l'effet de la structure spatiale sur la variabilité des mesures ?

- La variabilité est-elle plus faible au sein des observations voisines qu'en général ?
- La liaison entre deux observations induit elle des différences plus importantes qu'en général ?

Variance et covariance temporelles

Si le processus est "stationnaire"

- On veut l'influence d'une observation y_t à un temps donné t sur une autre y_{t+h}
- Les 2 observations sont séparées par un temps h , seule information importante
- écriture d'un modèle qui exprime la prédiction de y_{t+h} en fonction de
 - y_t
 - et de l'erreur

Construction des statistiques temporelles

On couple les observations y_t et y_{t+h}

Ainsi si l'on a Y

112 115 145 171 196 204 242 284 315 340 360 417

h=0	h=1	h=2	h=3
y_t	y_{t+1}	y_{t+2}	y_{t+3}
112	115	145	171
115	145	171	196
145	171	196	204
171	196	204	242
196	204	242	284
...			
315	340	360	417
340	360	417	
360	417		
417			

Plus h est grand, moins il y a de couples de données.

- **covariance simple entre 2 variables**

$$\sigma_{xy}^{\hat{}} = \gamma(\hat{0}) = \frac{1}{n} \sum_{r=1}^n (x_r - \bar{x})(y_r - \bar{y})$$

- **auto covariance d'ordre 1, 2 .. h**

Si les observations sont régulièrement espacées dans le temps (souvent le cas)

- $\gamma(\hat{1}) = \frac{1}{n-1} \sum_{t=1}^{n-1} (y_t - \bar{y})(y_{t+1} - \bar{y})$

- $\gamma(\hat{2}) = \frac{1}{n-2} \sum_{t=1}^{n-2} (y_t - \bar{y})(y_{t+2} - \bar{y})$

- $\gamma(\hat{h}) = \frac{1}{n-h} \sum_{t=1}^{n-h} (y_t - \bar{y})(y_{t+h} - \bar{y})$

Comment généraliser la distance h dans le domaine spatial ?

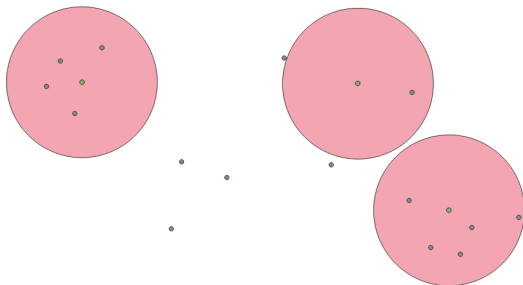
- **corrélations** $\rho_{xy}^{\hat{}} = \rho(\hat{0}) = \frac{\sigma_{xy}^{\hat{}}}{\sqrt{\sigma_x^{\hat{}} \sigma_y^{\hat{}}}}$
- **auto corrélations d'ordre 1, 2 .. h**
 - $\rho(\hat{1}) = \frac{\gamma(\hat{1})}{\gamma(\hat{0})}$
 - $\rho(\hat{2}) = \frac{\gamma(\hat{2})}{\gamma(\hat{0})}$
 - $\rho(\hat{h}) = \frac{\gamma(\hat{h})}{\gamma(\hat{0})}$

Variance et covariance spatiales

Voisinage d'une observation

spatial

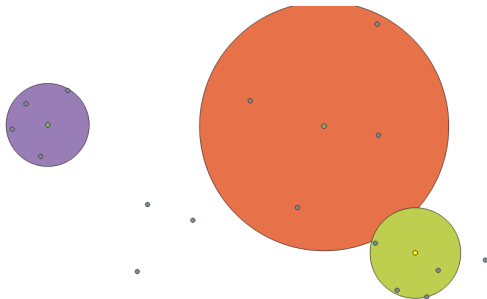
à partir d'une distance



Voisinage d'une observation

spatial

à partir d'un nombre d'observations



les k plus proches voisins

Indices de corrélations spatiales

revus par Cliff et Ord (1969 et 1973)

- **Moran**

$$\left(\frac{n}{m}\right) \frac{\sum_{r=1}^n \sum_{v=1}^{n_r} w_{rv} (z_r - \bar{z})(z_v - \bar{z})}{\sum_{r=1}^n (z_r - \bar{z})^2}$$

- **Geary**

$$\left(\frac{n-1}{2m}\right) \frac{\sum_{r=1}^n \sum_{v=1}^{n_r} w_{rv} (z_r - z_v)^2}{\sum_{r=1}^n (z_r - \bar{z})^2}$$

Indices de corrélations spatiales

Notations

- z_r valeur de la variable pour l'entité géographique "r"
- v désigne un voisin de l'entité "r"
- n_r le nombre de voisins de "r"
- W =matrice de pondération nxn
- $m = \sum_{r=1}^n \sum_{v=1}^{n_r} w_{rv}$
- par exemple, si $w_{rv} = 1$ pour les "r,v" voisins et "0" autrement alors m = nombre total de paires de voisins
- contraintes supplémentaires
 $\sum_{v=1}^{n_r} w_{rv} = 1$ standardisation en ligne

Indices de corrélations spatiales

Ecriture matricielle de Saporta

D une matrice de poids

Les données Y sont les données Z centrées

- **Moran**

$$I = \frac{Y^t W Y}{Y^t N Y}$$

- **Geary**

$$\frac{Y^t (N - W) Y}{Y^t N Y}$$

Les formules conduisent à un indice élevé si

- les écarts à la moyenne sont élevés
- les valeurs entre voisins sont proches (indice de Moran proche de 1)
- les valeurs entre voisins sont à l'opposé (indice de Moran proche de -1)
- **Valeur globale** correspondant à plusieurs configurations possibles ne caractérise pas la configuration mais seulement l'intensité du lien
- **comparaison à la moyenne globale** insensible à des valeurs de voisinage proches mais près de la moyenne

Indices de corrélations spatiales

Espérance de l'indice de Moran

Si l'on centre et l'on réduit la variable z : $z' = \frac{z - \bar{z}}{sd(z)}$

avec $sd(z) = \sqrt{\frac{\sum_{r=1}^n (z_r - \bar{z})^2}{n}}$

L'indice s'écrit très simplement :

$$I = \left(\frac{1}{m}\right) \sum_{r=1}^n \sum_{v=1}^n w_{rv} z'_r z'_v$$

et

$$E(I) = \frac{-1}{n-1}$$

Le score $\frac{I - E(I)}{\sqrt{Var(I)}}$ suit une normale centrée réduite et permet de tester

Test basé sur les permutations

Si il n'y a pas de structure spatiale alors n'importe quelle valeur de z peut être affectée à n'importe quel support s .

On permute 500 à 1000 fois, pour chaque permutation on a une valeur de l'indice de Moran.

On récupère ainsi une approximation de la densité sous l'hypothèse H_0 d'indépendance spatiale.

On compare la valeur trouvée initialement aux quantiles d'après la densité estimée.

LISA (Indicateurs Locaux d'Association Spatiale) de Luc Anselin (1995)

- **Contribution de chaque entité r à l'indice de Moran**

$$\frac{n}{m} \frac{\sum_{r=1}^n \sum_{v=1}^{n_r} w_{rv} (z_r - \bar{z})(z_v - \bar{z})}{\sum_{r=1}^n (z_r - \bar{z})^2}$$



$$\frac{n}{m} \sum_{r=1}^n C_r$$

- avec $C_r = \frac{\sum_{v=1}^{n_r} w_{rv} (z_r - \bar{z})(z_v - \bar{z})}{\sum_{r=1}^n (z_r - \bar{z})^2}$

- **Illustration des C_r**