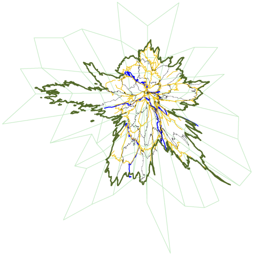
# Accessibilité routière des communes françaises

## Séminaire de méthodologie statistique



Sylvain Daubrée
Vincent Loonis

20 janvier 2015

# Anamorphose : principe

Soit $M_0$ le point le plus accessible de France continentale (situé dans l'Essonne), l'accessibilité d'un point quelconque $M$ est alors

$$A_M = A_{M_0} + \overbrace{A_M - A_{M_0}}^{z_M \geq 0}$$

On définit alors l'image $M'$ d'un point $M$ par l'anamorphose de telle sorte que

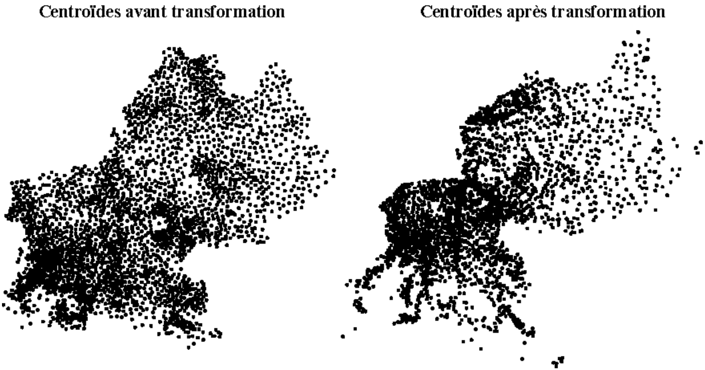
$$\overrightarrow{M_0 M'} = (A_M - A_{M_0}) \frac{\overrightarrow{M_0 M}}{||\overrightarrow{M_0 M}||} \Leftrightarrow M' = f_{z_M}(M)$$

# Anamorphose

**Centroïdes avant transformation**

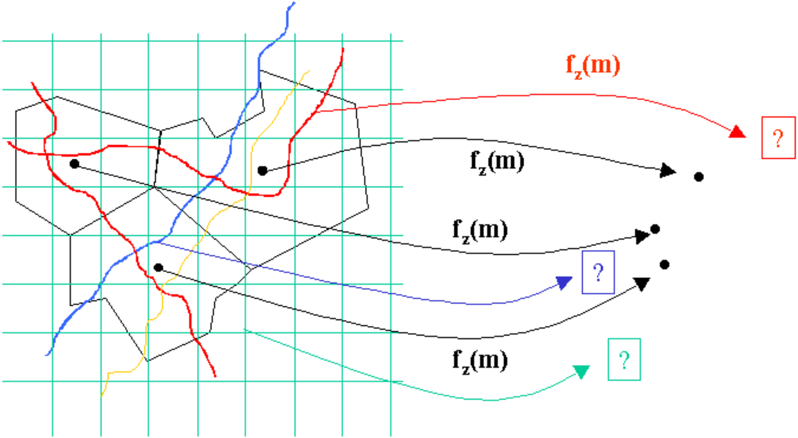
**Centroïdes après transformation**

# Anamorphose




Si on connaît $Z_M$ et donc $A_M$, on déduit facilement l'image de $M$. Le problème est que l'on ne connaît pas $Z_M$ pour tout $M$ du territoire...

# Un problème d'interpolation



# Problématique

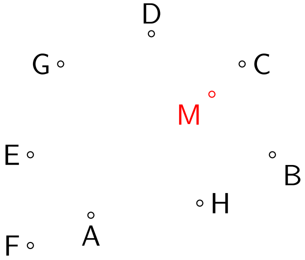
Pour un point quelconque du territoire, on cherche à connaître $\hat{A}_M$, et donc $\hat{z}_M = \hat{A}_M - A_{M_0}$

Plusieurs types de méthodes :

- géométrique : triangulation de Delaunay
- économétriques : régression locale et krigeage

# Interpolation basée sur une triangulation de Delaunay

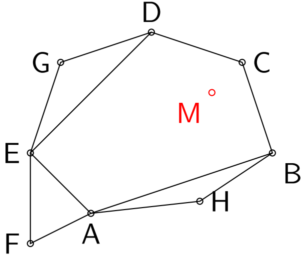
principe : interpolation géométrique




On connaît $z_A$, $z_B$, $z_C$, $z_D$, $z_E$, $z_F$, $z_G$, et $z_H$, que vaut $z_M$ ?

# Interpolation basée sur une triangulation de Delaunay

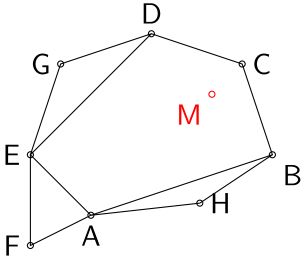
principe : interpolation géométrique




On connaît $z_A$, $z_B$, $z_C$, $z_D$, $z_E$, $z_F$, $z_G$, et $z_H$, que vaut $z_M$ ?

# Interpolation basée sur une triangulation de Delaunay

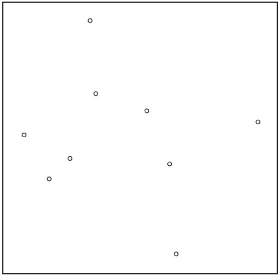
principe : interpolation géométrique




On connaît $z_A$, $z_B$, $z_C$, $z_D$, $z_E$, $z_F$, $z_G$, et $z_H$, que vaut $z_M$ ?

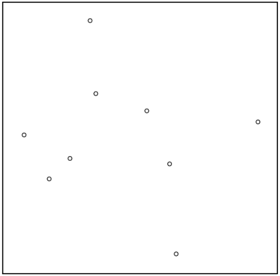
- partition de l'espace
- interpolation

# Triangulation de Delaunay

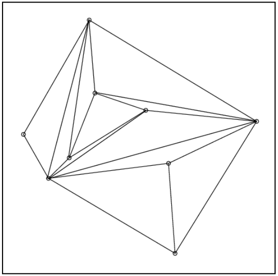


Sommets à trianguler

# Triangulation de Delaunay

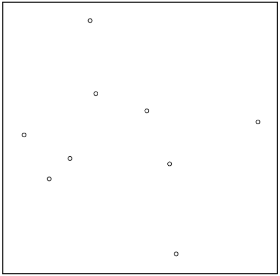


Sommets à trianguler

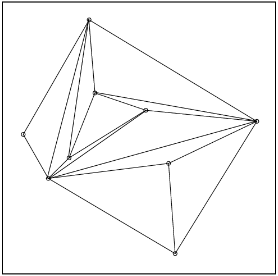


Triangulation quelconque

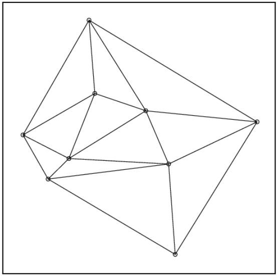
# Triangulation de Delaunay



Sommets à trianguler

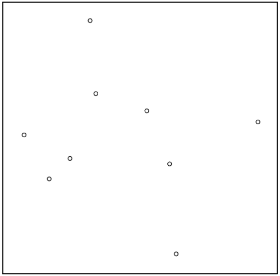


Triangulation quelconque

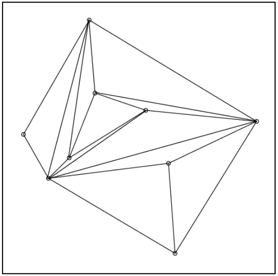


Triangulation de Delaunay

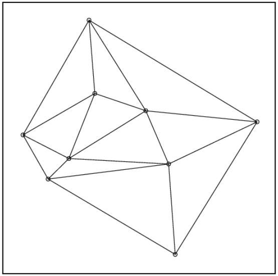
# Triangulation de Delaunay



Sommets à trianguler



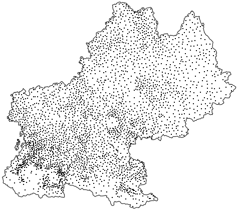
Triangulation quelconque

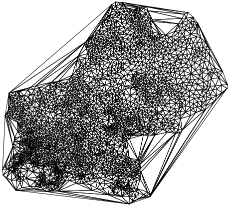


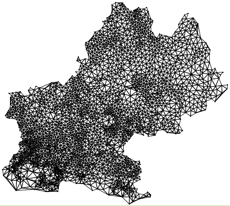
Triangulation de Delaunay

La triangulation de Delaunay maximise l'angle minimal

# Application aux données : triangulation

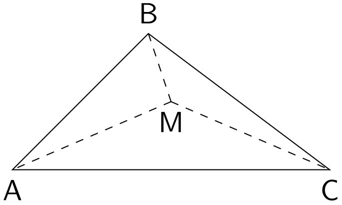






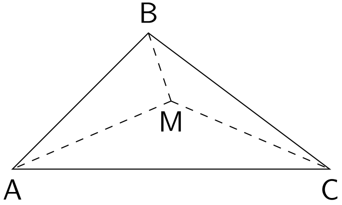
# Interpolation

- L'accessibilité interpolée d'un point est une moyenne pondérée des accessibilités des sommets du triangle auquel il appartient
- il faut déterminer les poids des sommets




$$z_M = \omega_A \cdot z_A + \omega_B \cdot z_B + \omega_C \cdot z_C$$

# Interpolation



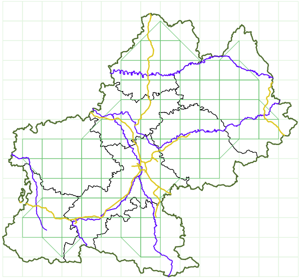

$$z_M = \omega_A \cdot z_A + \omega_B \cdot z_B + \omega_C \cdot z_C$$

On pose $\omega_A = \frac{\mathcal{A}_{BMC}}{\mathcal{A}_{ABC}},\ \omega_B = \frac{\mathcal{A}_{AMC}}{\mathcal{A}_{ABC}},\ \omega_C = \frac{\mathcal{A}_{BMA}}{\mathcal{A}_{ABC}}$, avec $\mathcal{A}_{ABC}$ l'aire du triangle $ABC$

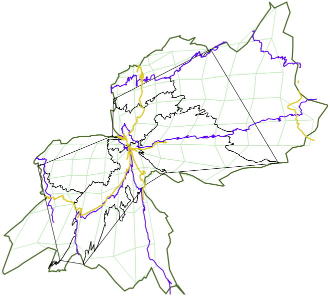
Ainsi :

- le poids d'un sommet est d'autant plus important que le point à interpoler en est proche
- la somme des poids est égal à 1
- la valeur interpolée d'un sommet est la vraie valeur

# Application aux données : interpolation



Avant



Après

# Conclusion sur l'usage de la triangulation de Delaunay

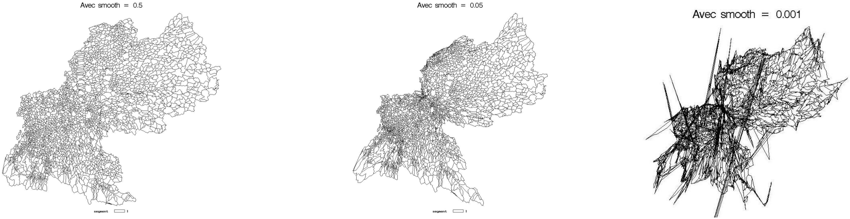
- méthode simple dans son principe et intuitive
- l'interpolation de l'accessibilité d'un point de l'échantillon donne la valeur réelle pour ce point (interpolation exacte)
- implémentation pouvant être complexe, calculs pouvant être longs
- impossibilité de calculer l'image de points situés en dehors de la triangulation

# Régression locale

- principe : approximer une variable aléatoire par un modèle fonction des données qui lui sont proches
- paramètres d'ajustement : **voisinage**, fonction de poids, famille de fonctions utilisées pour les régressions

$$\hat{z}_M = \sum_{N \in V_M} W_M(N) \cdot f(N)$$

# Application avec différents *smoothing parameter*




Il est possible de mettre en œuvre des algorithmes pour déterminer automatiquement le *smoothing parameter* (notamment en adaptant l'*Akaike information criterion*).

# Conclusion sur l'usage de la régression locale

- Méthode simple à mettre en œuvre
- besoin d'ajuster le *smoothing parameter* : à la fois avantage et inconvénient
- implémentation SAS insatisfaisante pour le contour extérieur
- méthode non exacte

# Le Krigeage

- on considère la variable d'intérêt comme une fonction aléatoire définie sur le plan
- la spécificité de la méthode est qu'on ne considère pas que les résidus sont spatialement indépendants, on va modéliser cette dépendance pour l'intégrer dans le modèle

Le modèle peut s'écrire :

$$Z(s) = m(s) + \delta(s)$$

avec $s$ un point du plan et $\delta$ fonction aléatoire de moyenne nulle et de structure de dépendance connue.

# Présentation de la méthode

On distingue essentiellement trois types de krigeage, selon que $m$ est

- stationnaire de moyenne connue (krigeage simple),
- stationnaire de moyenne inconnue (krigeage ordinaire),
- non stationnaire (krigeage universel). La tendance est alors modélisée à l'aide d'une régression linéaire.

Sous ces hypothèses, on définit l'estimateur de krigeage $\hat{Z}$ comme le meilleur estimateur linéaire sans biais (BLUE).

# Présentation de la méthode

La définition du kriegage définit des contraintes sur $\hat{Z}$:

- linéarité : on doit pouvoir écrire $\hat{Z} = a + \sum \lambda_i Z(s_i)$ avec les $s_i$ l'ensemble des sites pour lesquels on connait Z
- autorisation : l'espérance et la variance de l'erreur d'estimation doivent être définies
- absence de biais : $E[\hat{Z}(s_0) - Z(s_0)] = 0$ en tout site $s_0$
- optimalité : on doit minimiser $Var(\hat{Z}(s_0) - Z(s_0))$

# Krigeage

Résolution des équations de krigeage :

$$\hat{Z}(s_0) = (\gamma_0 + X(X^t\Gamma^{-1}X)^{-1}(x_0 - X^t\Gamma^{-1}\gamma_0))^t\Gamma^{-1}Z$$

avec

- $X$ et $x_0$ qui modélisent la tendance, la variation d'espérance sur le plan
- $\Gamma$ et $\gamma_0$ qui modélisent la dépendance spatiale des résidus : $\Gamma_{i,j}$ quantifie la dépendance spatiale entre les sites $i$ et $j$

# Présentation de la méthode

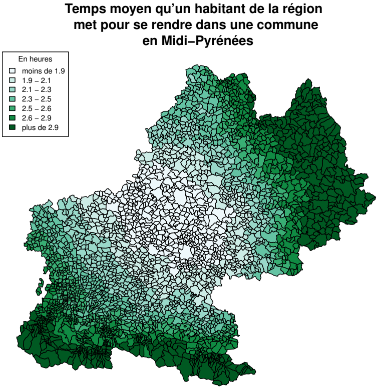
Les étapes du krigeage sont les suivantes :

- analyse exploratoire (forme de la tendance) : détermination de $X$ et $x_0$
- analyse variographique (modélisation de la structure de dépendance des résidus) : détermination de $\Gamma$ et $\gamma_0$
- krigeage (interpolation)

# Analyse exploratoire

- familiarisation avec les données
- forme de la tendance
  - cartes en 3D
  - cartes choroplèthes

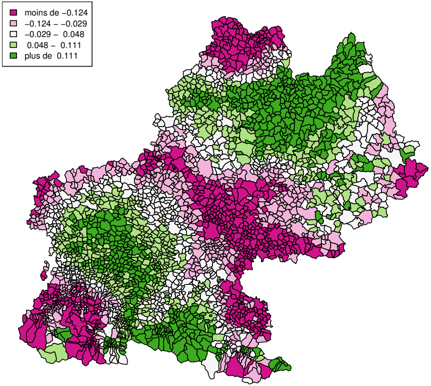
# Analyse exploratoire




$x$ et $y$ étant les coordonnées géographiques, le modèle devient :

$$Z = \underbrace{\alpha_1 x + \alpha_2 y + \alpha_3 xy + \alpha_4 x^2 + \alpha_5 y^2}_{\text{tendance}=m} + \delta$$

# Analyse exploratoire des résidus $\hat{\delta}$



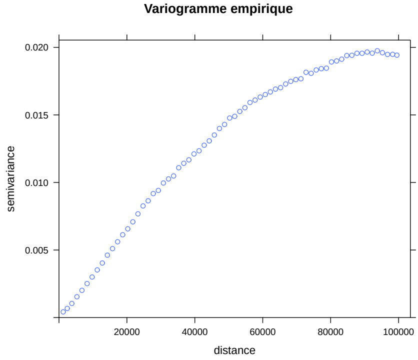
# Analyse variographique sur les résidus

- définition du semi-variogramme : $\gamma(h) = \frac{Var(\delta(s+h)-\delta(s))}{2}, h \in \mathbb{R}^2$
- on se place dans le cas isotropique, $\gamma$ ne dépend que de $||h||$
- estimation de cette fonction par un semi-variogramme expérimental par cet estimateur :

$$\hat{\gamma}(r) = \frac{1}{2n(r)} \sum_{\substack{s_1, s_2 / \\ ||s_2 - s_1|| \in [r-\epsilon, r+\epsilon]}} (z(s_1) - z(s_2))^2, r \in \mathbb{R}$$

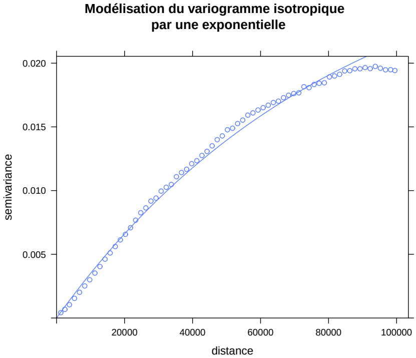
# Analyse variographique

Ajustement sur un semi-variogramme théorique valide à partir de ce semi-variogramme expérimental en utilisant les moindres carrés pondérés



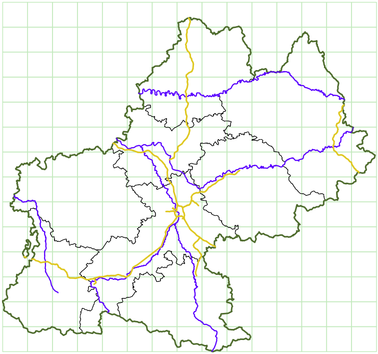
# Analyse variographique

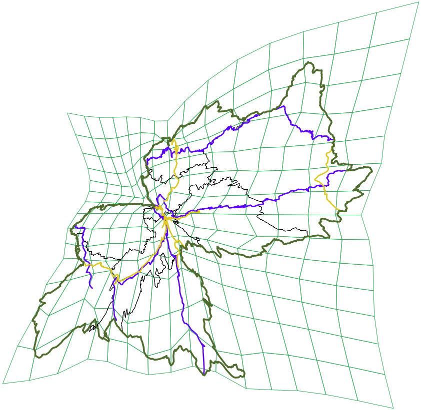
Ajustement sur un semi-variogramme théorique valide à partir de ce semi-variogramme expérimental en utilisant les moindres carrés pondérés




$$\gamma(h) = C(1 - \exp^{-\frac{h}{a}})$$

# Krigeage





# Conclusion sur l'usage du krigeage

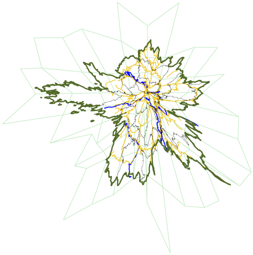
- Résultats satisfaisants, y compris en dehors de l'enveloppe des points échantillons
- calculs plutôt rapides
- méthode complexe, difficile à automatiser de bout en bout
- méthode exacte

# Conclusion

| | Interpolation géométrique | Régression locale | Krigeage |
|---|---|---|---|
| Mise en œuvre | Simple | Moyennement complexe | Complexe |
| Effets de bord | Pas d'interpolation du contour | Problème l'implémentation de SAS | Pas d'effet de bord |
| Temps de calcul | 30 minutes | Inférieur à 10 secondes | Inférieur à cinq secondes |
| Paramétrabilité | Inexistante | le *smoothing parameter* | Le voisinage et, dans une moindre mesure, le variogramme |

# Accessibilité routière des communes françaises

## Séminaire de méthodologie statistique



Sylvain Daubrée
Vincent Loonis

20 janvier 2015