

Géostatistique

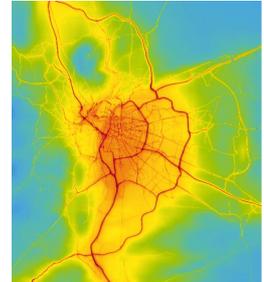
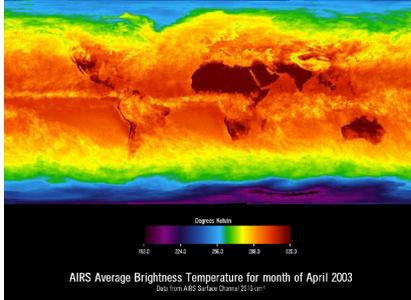
M. Ribatet

Master Biostat — Université de Montpellier

1. Processus stochastiques	6
Quelques rappels	7
Processus gaussiens	9
Processus d'ordre 2	12
Stationnarité stricte	13
Stationnarité faible	14
Isotropie	18
Continuité de $K(\cdot)$	21
Continuité de $Z(\cdot)$	22
Dérivabilité de $Z(\cdot)$	24
Corrélation paramétrique	25
Stationnarité intrinsèque	27
Variogramme	28
Propriétés de $\gamma(\cdot)$	29
Variogrammes paramétriques	31
2. Modélisation statistique	33
Données de calcium	34
Variogramme empirique	39
Ajustement d'un variogramme	47
Krigage	50
Modélisation par processus gaussiens	61
DM	63
3. Simulations de processus gaussiens	67
Motivations	68
Simulations non conditionnelles	69
Simulations conditionnelles	70

Qu'est ce que la géostatistique ?

- La **géostatistique** est une branche des statistiques visant à donner une **description** de quantités distribuées **spatialement** ou encore **spatio-temporellement**.
- Voici quelques quantités d'intérêt :
 - la température, précipitation, neige, ...
 - concentration d'un polluant, ozone, ...
 - teneur d'un minerai d'un gisement, ...



Géostatistique

Mathieu Ribatet – 2 / 70

Objectifs

- L'objectif d'une analyse géostatistique sera bien souvent la **prédiction**, e.g.,
 - Température à Carnon sachant celles de Palavas, du Grau du roi et de la Grande Motte ?
 - Quel est l'endroit ayant la plus forte concentration d'un polluant ?
- Toutefois elle permettra bien également de répondre à d'autres questions :
 - les températures à Montpellier et Nîmes sont elles fortement liées ?
 - Quelles variables semblent influencer la pollution ?

Géostatistique

Mathieu Ribatet – 3 / 70

Quelques références

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics: Modelling Spatial Uncertainty*. Wiley, New York.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. John Wiley & Sons inc., New York.

Diggle, P., Ribeiro, P., and Justiniano, P. (2007). *Model-based Geostatistics*. Springer Series in Statistics. Springer.

Stein, M. L. (1999). *Interpolation for spatial data: Some theory for kriging*. Springer, New York.

Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications*. Springer, New York, third edition edition.

Ce que nous ne verrons pas :-)

- Les données sur réseaux // grille, e.g.,
 - analyse d'une image via un champ de Markov ;
 - nombre de malades de la grippe sur l'agglomération Montpelliéraine.
- Théorie ergodique ;
- Cadre multivarié, e.g., modélisation **jointe** de la température et l'humidité.
- Théorie spectrale des processus (Bochner, Stein, Schoenberg, ...)

Quelques rappels

Définition 1 (Rappel). Un processus stochastique défini sur \mathcal{X} et à valeur dans E est une famille de variables aléatoires indexées par \mathcal{X} et à valeur dans E toutes définies sur un même espace de probabilité (Ω, \mathcal{F}, P) .

Exemple 1 (Marche aléatoire). Soient $\varepsilon_1, \varepsilon_2, \dots \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ et posons

$$Z_0 = \varepsilon_0, \quad Z_{t+1} = Z_t + \varepsilon_{t+1}, \quad t \geq 0.$$

Ceci définit un processus stochastique sur $\mathcal{X} = \mathbb{N}$ et à valeur dans $E = \mathbb{R}$.

Lois fini-dimensionnelles

Proposition 1. Un processus stochastique est totalement caractérisé[†] par ses lois fini-dimensionnelles, i.e., pour tout $n \geq 1$ et $(s_1, \dots, s_n) \in \mathcal{X}^n$,

$$\Pr\{Z(s_1) \in B_1, \dots, Z(s_n) \in B_n\}, \quad B_1, \dots, B_n \text{ boréliens de } E.$$

Ces dernières doivent cependant satisfaire les hypothèses du théorème d'extension de Kolmogorov, i.e., pour toute permutation σ de $\{1, \dots, n\}$,

$$\nu_{s_1, \dots, s_n}(B_1, \dots, B_n) = \nu_{\sigma(s_1), \dots, \sigma(s_n)}\{B_{\sigma(1)}, \dots, B_{\sigma(n)}\}$$

et

$$\nu_{s_1, \dots, s_{n-1}}(B_1, \dots, B_{n-1}) = \nu_{s_1, \dots, s_n}(B_1, \dots, B_{n-1}, E).$$

Remarque. Le théorème d'extension de Kolmogorov permet d'assurer l'existence d'un processus stochastique ayant des lois fini-dimensionnelles données ν .

Processus gaussiens

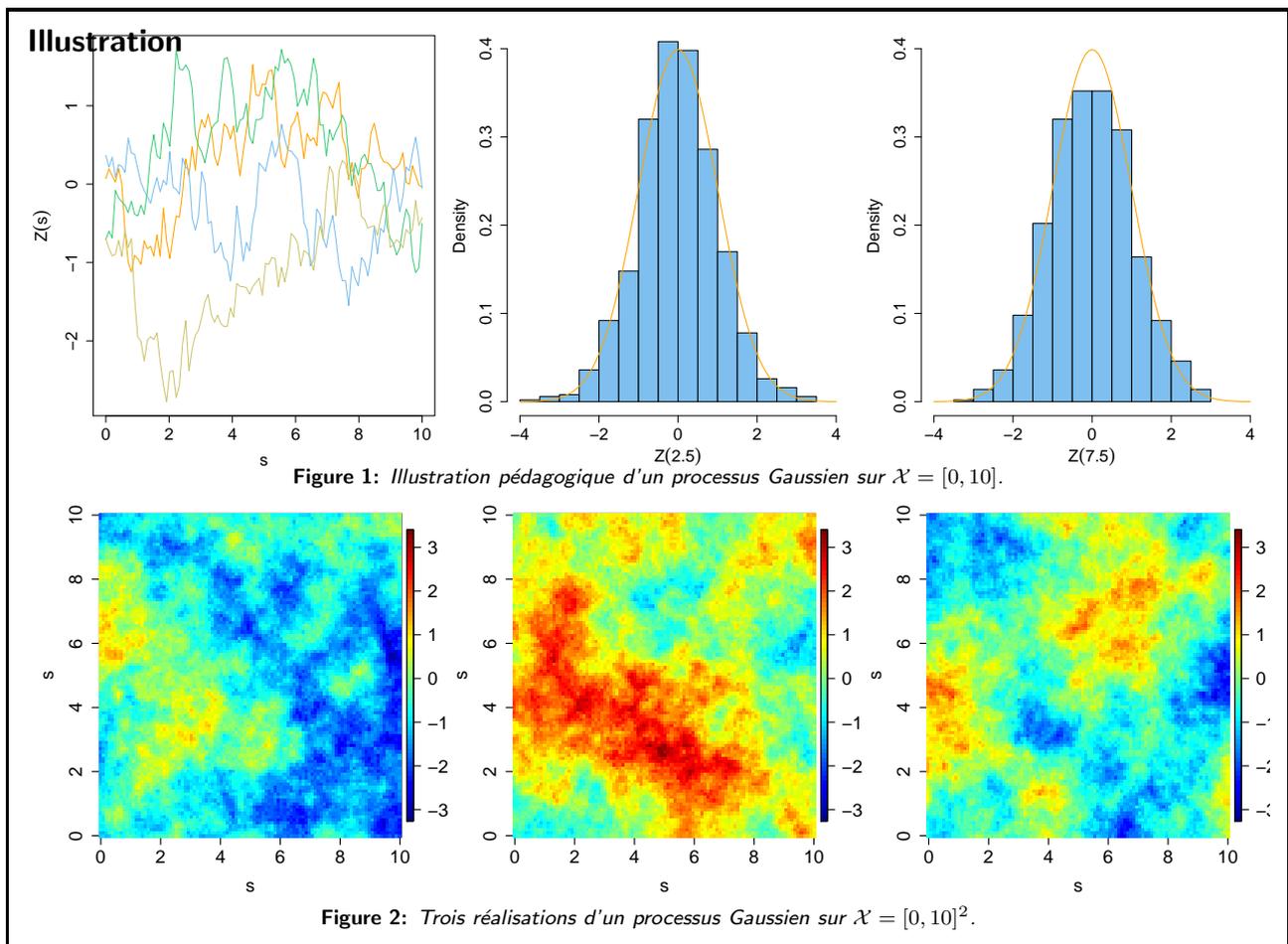
Définition 2. Un processus $\{Z(s) : s \in \mathcal{X}\}$ est un **processus gaussien** si ses lois fini-dimensionnelles sont Gaussiennes.

Exemple 2. L'exemple précédent de la marche aléatoire

$$Z_{t+1} = Z_t + \varepsilon_{t+1}, \quad \varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2),$$

définit un processus Gaussien sur \mathbb{N} .

Remarque. Certaines personnes utilisent le terme “processus Gaussien” lorsque $\mathcal{X} = \mathbb{R}$ ou \mathbb{N} et “champs Gaussien” lorsque $\mathcal{X} = \mathbb{R}^d$, $d \geq 2$. Personnellement je ne fais aucune distinction et parlerai toujours de processus stochastique (Gaussien ou non d'ailleurs).



Cadre géostatistique

- Puisque la géostatistique vise, par exemple, à donner une description d'une quantité distribuée spatialement ou spatio-temporellement, nous nous restreindrons au cas où $\mathcal{X} = \mathbb{R}^2$ voire $\mathcal{X} = \mathbb{R}^3$.
- Par exemple pour la modélisation de la pluie, nous définirons donc un processus stochastique défini sur \mathbb{R}^2 et à valeur dans \mathbb{R}_+ .
- Dans la suite, nous noterons indifféremment le processus stochastique par $\{Z(s) : s \in \mathcal{X}\}$, $Z(s)$, $Z(\cdot)$ ou tout simplement Z .

Géostatistique

Mathieu Ribatet – 11 / 70

Processus d'ordre 2

Définition 3. Un processus $\{Z(s) : s \in \mathcal{X}\}$ est dit **d'ordre 2** si pour tout $s \in \mathcal{X}$

$$\text{Var}\{Z(s)\} < \infty \quad (\text{et donc a fortiori } \mathbb{E}\{Z(s)\} < \infty)$$

- Ceci nous permet alors de considérer la **tendance** (trend // drift)

$$\begin{aligned} \mu : \mathcal{X} &\longrightarrow E \\ s &\longmapsto \mu(s) = \mathbb{E}\{Z(s)\}, \end{aligned}$$

- et la **fonction de covariance**

$$\begin{aligned} K : \mathcal{X} \times \mathcal{X} &\longrightarrow \mathbb{R} \\ (s_1, s_2) &\longmapsto K(s_1, s_2) = \text{Cov}\{Z(s_1), Z(s_2)\}. \end{aligned}$$

Géostatistique

Mathieu Ribatet – 12 / 70

Stationnarité stricte

Définition 4. Un processus $\{Z(s) : s \in \mathcal{X}\}$ est dit **strictement stationnaire** si pour tout $n \geq 1$ et $(s_1, \dots, s_n) \in \mathcal{X}^n$

$$\Pr\{Z(s_1) \in B_1, \dots, Z(s_n) \in B_n\} = \Pr\{Z(s_1 + h) \in B_1, \dots, Z(s_n + h) \in B_n\},$$

$h \in \mathcal{X}$ et Boréliens B_1, \dots, B_n .

☞ C'est généralement une propriété bien trop forte (et impossible à vérifier en pratique) qui font que l'on considère souvent une version assouplie.

Stationnarité faible

Définition 5. Un processus d'ordre 2 $\{Z(s) : s \in \mathcal{X}\}$ est dit **faiblement stationnaire** si pour tout $s, s_1, s_2, h \in \mathcal{X}$

$$\begin{aligned}\mathbb{E}\{Z(s)\} &= \mathbb{E}\{Z(s+h)\} \\ \text{Cov}\{Z(s_1), Z(s_2)\} &= \text{Cov}\{Z(s_1+h), Z(s_2+h)\},\end{aligned}$$

i.e.,

$$\mu(s) = \mu, \quad K(s_1, s_2) = K(s_1+h, s_2+h).$$

Proposition 2. Si $\{Z(s) : s \in \mathcal{X}\}$ est faiblement stationnaire alors

$$\mu(s) = \mu, \quad K(s_1, s_2) = K(o, s_2 - s_1) \stackrel{\text{not.}}{=} K(s_2 - s_1),$$

où o est une origine arbitraire de \mathcal{X} .

Question pour vous. . .

- Avantage de la stationnarité ?
- Inconvénient de la stationnarité ?

Géostatistique

Mathieu Ribatet – 15 / 70

Propriétés de la fonction de covariance (DM)

Soit $\{Z(s) : s \in \mathcal{X}\}$ un processus faiblement stationnaire. Sa fonction de covariance $K(\cdot)$ vérifie alors

- $K(o) = \text{Var}\{Z(s)\} \geq 0$ pour tout $s \in \mathcal{X}$;
- $K(h) = K(-h)$;
- $K(h) \leq K(o)$;
- Pour tout $n \geq 1$ et $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ et $s_1, \dots, s_n \in \mathcal{X}$,

$$\sum_{i,j=1}^n \lambda_i \lambda_j K(s_i - s_j) \geq 0.$$

Remarque. Cette dernière propriété revient à dire que $K(\cdot)$ est (semi) définie positive.

Géostatistique

Mathieu Ribatet – 16 / 70

Propriétés sur les fonctions définies positives

- Si $K_1(\cdot)$ et $K_2(\cdot)$ sont def. pos. alors $a_1K_1(\cdot) + a_2K_2(\cdot)$ l'est aussi pour $a_1, a_2 \geq 0$;
- Soient $\{K_n(\cdot)\}_{n \geq 1}$ une suite de fonctions def. pos. telle que $K_n(s) \rightarrow K(s)$ lorsque $n \rightarrow \infty$ pour tout $s \in \mathcal{X}$. Alors $K(\cdot)$ est def. pos. ;
- Si $K_1(\cdot)$ et $K_2(\cdot)$ sont def. pos. alors $K(s) = K_1(s)K_2(s)$ est def. pos.

Géostatistique

Mathieu Ribatet – 17 / 70

Isotropie

- Dans un esprit de modélisation, il sera souvent “agréable” de supposer l’isotropie

Définition 6. Un processus $\{Z(s) : s \in \mathcal{X}\}$ est dit **isotrope** si pour tout matrice de rotation^a R

$$\{Z(s) : s \in \mathcal{X}\} \stackrel{d}{=} \{Z(Rs) : s \in \mathcal{X}\}.$$

Remarque. Si $\{Z(s) : s \in \mathcal{X}\}$ est faiblement stationnaire, alors cela revient à dire

$$K(h) = K(\|h\|), \quad h \in \mathcal{X}.$$

Remarque. Un processus non isotrope est **anisotrope**.

^aUne matrice A est une matrice de rotation si $A^{-1} = A^T$ et $|A| = 1$

Géostatistique

Mathieu Ribatet – 18 / 70

Question pour vous...

- Avantage de l'isotropie ?
- Inconvénient de l'isotropie ?

Géostatistique

Mathieu Ribatet – 19 / 70

Allure générale de $K(\cdot)$

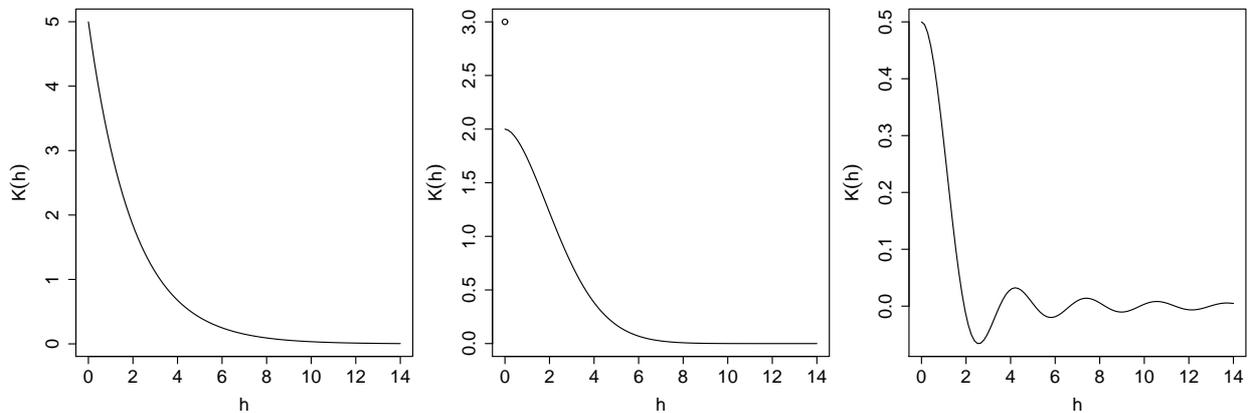


Figure 3: Quelques fonctions de covariance isotropes.

En profiter pour définir :

- portée (practical range)
- variance partielle (partial sill)
- effet pépite (nugget)
- dérivabilité en l'origine

Géostatistique

Mathieu Ribatet – 20 / 70

Continuité de $K(\cdot)$

Proposition 3. Soit un processus faiblement stationnaire $\{Z(s) : s \in \mathcal{X}\}$ de fonction de covariance $K(\cdot)$. Si $K(\cdot)$ est continu à l'origine alors elle est continue partout.

Corollaire 1. La fonction de covariance $K(\cdot)$ d'un processus faiblement stationnaire ne peut être discontinue qu'en son origine (effet pépité) et si tel est le cas alors

$$K(h) = c\delta_o(h) + K_c(h), \quad h \in \mathcal{X},$$

où $c > 0$ (la pépité) et $K_c(\cdot)$ est une fonction de covariance continue.

Continuité en moyenne quadratique

Définition 7. Un processus faiblement stationnaire $\{Z(s) : s \in \mathcal{X}\}$ est dit **continu en moyenne quadratique** si

$$\mathbb{E} [\{Z(o) - Z(h)\}^2] \rightarrow 0, \quad \|h\| \rightarrow 0.$$

Théorème 1. Un processus faiblement stationnaire $\{Z(s) : s \in \mathcal{X}\}$ est continu en moyenne quadratique si et seulement si $K(\cdot)$ est continue à l'origine.

Remarque (DM). Que devient ce théorème lorsque le processus est seulement d'ordre 2 et de moyenne constante ?

Trajectoires continues p.s. et continuité en moyenne quadratique

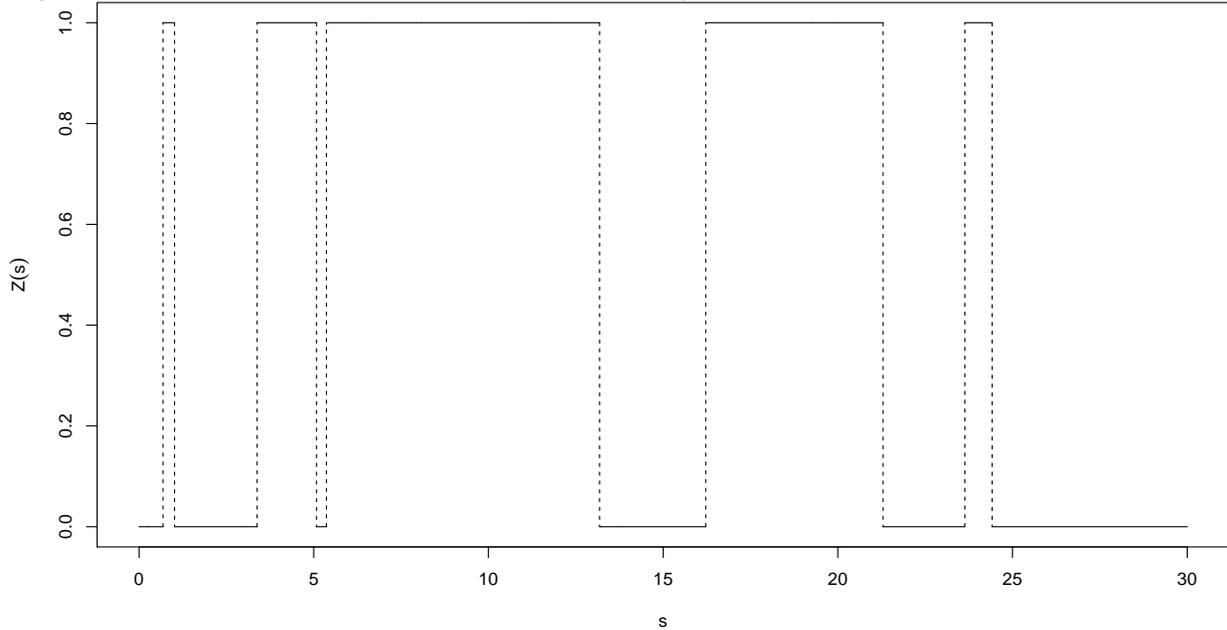


Figure 4: Réalisation d'un processus L_2 -continu et mais à trajectoires non continues.

$$Z(s) = X_{I(s)}, \quad I(s) = \inf\{i \geq 1: \xi_i \geq s\},$$

avec $\{\xi_i: i \geq 1\} \sim PPP(1)$ et $X_i \stackrel{iid}{\sim} \text{Ber}(p)$.

Dérivabilité en moyenne quadratique

Définition 8. Un processus faiblement stationnaire $\{Z(s): s \in \mathcal{X}\}$ est **dérivable** s'il existe un processus d'ordre 2 $\{D(s): s \in \mathcal{X}\}$ tel que

$$\mathbb{E} \left[\left\{ \frac{Z(s+h) - Z(s)}{\|h\|} - D(s) \right\}^2 \right] \rightarrow 0, \quad \|h\| \rightarrow 0.$$

Le processus $\{D(s): s \in \mathcal{X}\}$ est appelé **processus dérivé (en moyenne quadratique)** de $\{Z(s): s \in \mathcal{X}\}$ et sera noté naturellement $\{Z'(s): s \in \mathcal{X}\}$.

Proposition 4 (Bartlett, 1955). *Un processus faiblement stationnaire est dérivable en moyenne quadratique si et seulement si $K(\cdot)$ admet une dérivée seconde en son origine.*

Quelques familles de fonctions de corrélation isotropes

Exponentiel

$$\rho(h) = \exp\left(-\frac{h}{\lambda}\right), \quad \lambda > 0$$

Gaussien

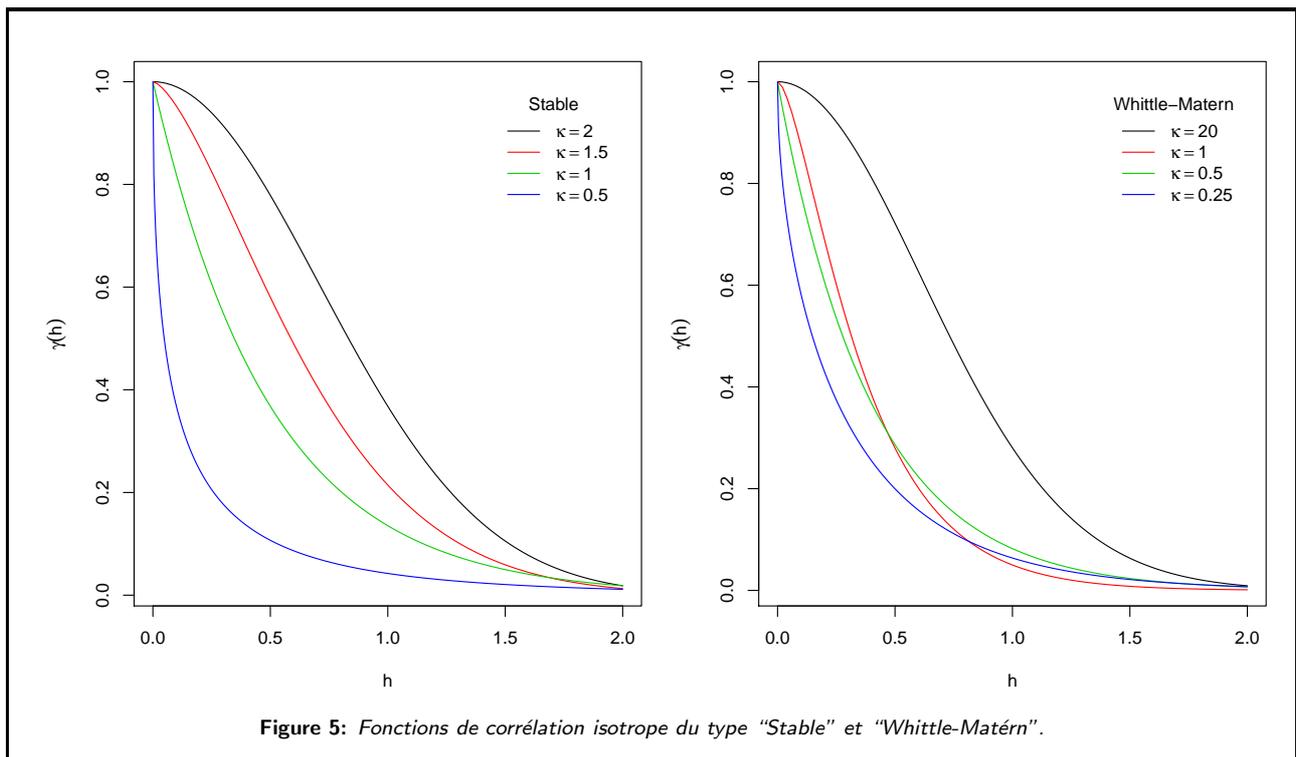
$$\rho(h) = \exp\left\{-\left(\frac{h}{\lambda}\right)^2\right\}, \quad \lambda > 0$$

Stable // Exponentiel puissance

$$\rho(h) = \exp\left\{-\left(\frac{h}{\lambda}\right)^\kappa\right\}, \quad 0 < \kappa \leq 2, \quad \lambda > 0$$

Whittle–Matérn

$$\rho(h) = \frac{2^{1-\kappa}}{\Gamma(\kappa)} \left(\frac{u}{\lambda}\right)^\kappa K_\kappa\left(\frac{u}{\lambda}\right), \quad \kappa > 0, \quad \lambda > 0.$$



Stationnarité intrinsèque

- Parfois il peut être nécessaire de relâcher (un peu) l'hypothèse de stationnarité faible

Définition 9. Un processus $\{Z(s) : s \in \mathcal{X}\}$ est dit **intrinsèquement stationnaire** ou à **accroissements stationnaires** si

$$\begin{aligned}\mathbb{E}\{Z(s) - Z(s+h)\} &= 0, \\ \text{Var}\{Z(s) - Z(s+h)\} &= 2\gamma(h) \text{ (ne dépend que de } h\text{)}\end{aligned}$$

- Stationnarité intrinsèque $\not\Rightarrow$ Stationnarité faible
- Stationnarité faible \Rightarrow Stationnarité intrinsèque
- Stationnarité intrinsèque suppose l'existence des 2 premiers moments de $\{Z(s+h) - Z(s) : s \in \mathcal{X}\}$ mais **aucunement ceux de $\{Z(s) : s \in \mathcal{X}\}$** .

Variogramme

Définition 10. Soit $\{Z(s) : s \in \mathcal{X}\}$ un processus (au moins!) à accroissements stationnaires. On appelle **(semi) variogramme** la fonction

$$\begin{aligned}\gamma : \mathcal{X} &\longrightarrow \mathbb{R}_+ \\ h &\longmapsto \gamma(h) = \frac{1}{2} \mathbb{E} [\{Z(s+h) - Z(s)\}^2].\end{aligned}$$

Remarque. Si $\{Z(s) : s \in \mathcal{X}\}$ est faiblement stationnaire alors clairement

$$\gamma(h) = K(o) - K(h).$$

En revanche l'existence d'un variogramme **n'assure pas** celle de la fonction de covariance !!!

Propriétés du variogramme (DM)

Proposition 5. Soient $\{Z(s) : s \in \mathcal{X}\}$ un processus à accroissements stationnaires et $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ tels que $\sum_{j=1}^n \lambda_j = 0$. Alors

$$\mathbb{E} \left\{ \sum_{j=1}^n \lambda_j Z(s_j) \right\} = 0, \quad \text{Var} \left\{ \sum_{j=1}^n \lambda_j Z(s_j) \right\} = - \sum_{i,j=1}^n \lambda_i \lambda_j \gamma(s_i - s_j) \geq 0$$

☞ La proposition précédente indique que sous l'hypothèse d'accroissement stationnaires, toutes les combinaisons linéaires de $Z(s_j)$ ne sont pas valides !

Remarque. Attention pour votre DM, on travaille ici sous les hypothèses d'accroissements stationnaires on ne peut donc pas écrire $\mathbb{E}\{Z(s)\}$ ou $\text{Var}\{Z(s)\}$!!!

Propriétés du variogramme

- $\gamma(o) = 0$;
- $\gamma(h) \geq 0$ pour tout $h \in \mathcal{X}$;
- $\gamma(-h) = \gamma(h)$ pour tout $h \in \mathcal{X}$;
- $\|h\|^{-2} \gamma(h) \rightarrow 0$ lorsque $\|h\| \rightarrow \infty$ (admis) ;
- La fonction $h \mapsto \exp\{-c\gamma(h)\}$, $c > 0$, est une fonction de covariance (admis).

Variogrammes paramétriques

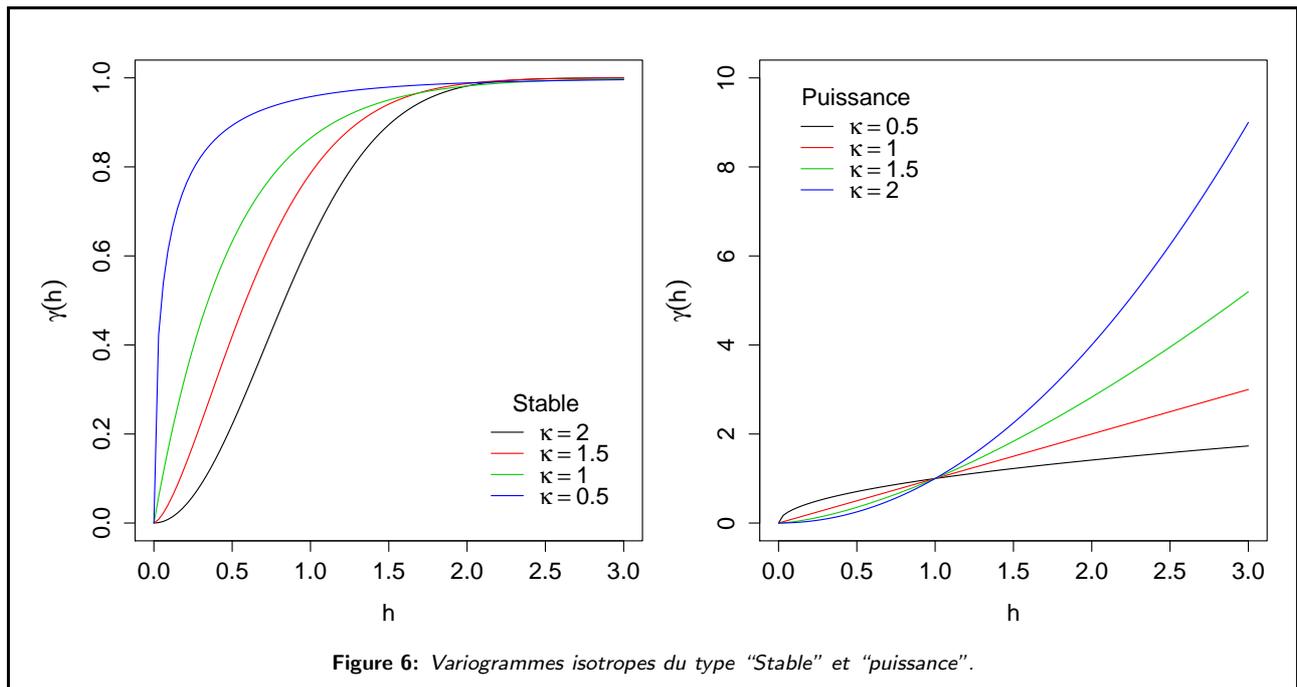
- Il existe deux types de variogrammes : bornés ou non.
- Dans le cas bornés, alors on reprend les familles paramétriques données pour les covariances puisque

$$\gamma(h) = K(o) - K(h)$$

- Si le variogramme est borné alors $\gamma(\cdot)$ hérite des propriétés déjà vues pour $K(\cdot)$.
- Un modèle paramétrique non borné fréquemment rencontré est le modèle puissance

$$\gamma(h) = \left(\frac{h}{\lambda}\right)^\kappa, \quad 0 < \kappa < 2.^a$$

^aCombiné avec un proc. gauss. on tombe sur un mouvement brownien fractionnaire (brownien classique pour $\kappa = 1$).



Données de calcium

- Données de teneur en calcium (et magnésium) du sol (entre 0 et 20cm de profondeur)
- La région d'étude a été divisée en 3 sous-régions
 - HG : parcelle pouvant être considérée comme naturelle ;
 - Bas: culture de riz (traitement au Ca le plus récent) ;
 - HD: A reçu des engrais dans le passé.
- Objectif : Modéliser spatialement la teneur de Ca encore présente qui a été utilisée par le passé pour neutraliser l'impact de l'aluminium présent dans le sol.

Format des données

	east	north	altitude	area	ca20
1	5710	4829	6.10	3	52
2	5727	4875	6.05	3	57
3	5745	4922	6.30	3	72
4	5764	4969	6.60	3	74
5	5781	5015	6.60	3	68
6	5799	5062	5.75	3	45
7	5817	5109	5.35	3	47
8	5837	5161	5.10	3	49
9	5873	5254	5.00	2	38
10	5907	5347	5.20	2	53
11	5925	5394	5.05	2	60
12	5943	5441	4.65	2	47
13	5961	5487	4.40	2	34
...					

Question pour vous...

En tant que statisticien, que feriez vous si vous deviez travailler sur ces données ?

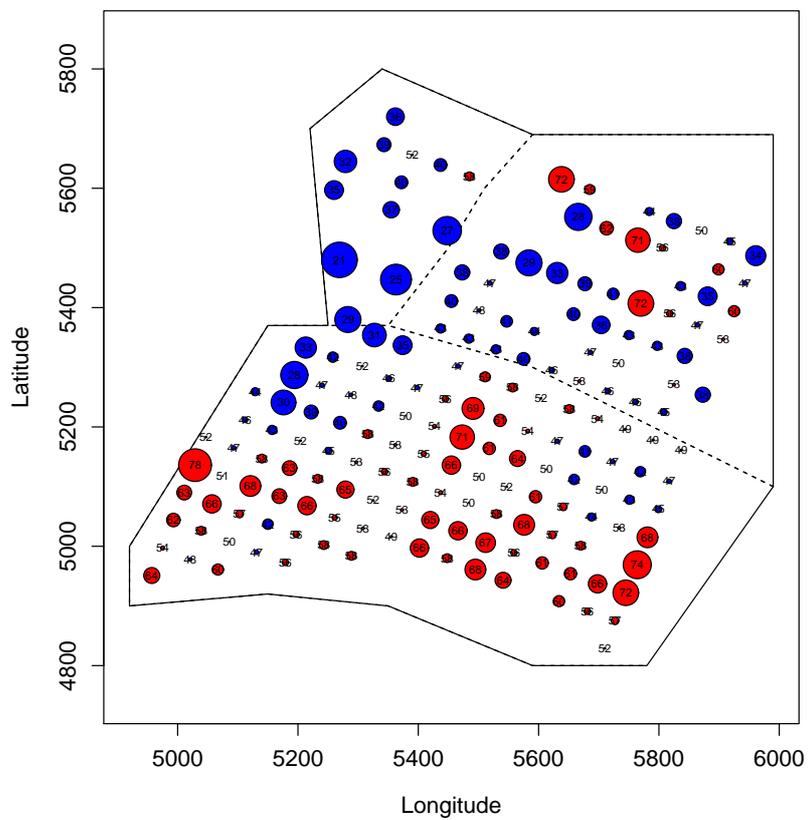
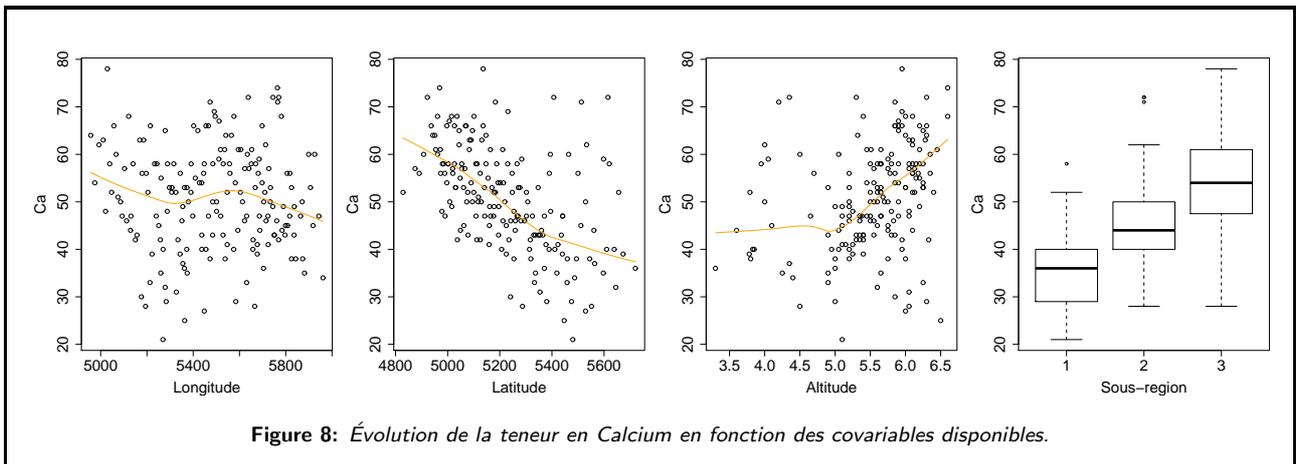


Figure 7: Données de teneur en Calcium dans le sol. (symbol plot)



Variogramme empirique

Comme

$$\gamma(h) = \frac{1}{2} \mathbb{E} [\{Z(o) - Z(h)\}^2],$$

un estimateur très naturel est

$$\hat{\gamma}(h) = \frac{1}{2n} \sum_{\ell=1}^n \{Z_{\ell}(s_i) - Z_{\ell}(s_j)\}^2, \quad h = \|s_i - s_j\|,$$

ou encore sa version agrégée (binned)

$$\hat{\gamma}(h) = \frac{1}{2nN(h, \varepsilon)} \sum_{\ell=1}^n \sum_{i,j=1}^k \{Z_{\ell}(s_i) - Z_{\ell}(s_j)\}^2 1_{\{\|s_i - s_j\| - h \leq \varepsilon\}},$$

avec

$$N(h, \varepsilon) = \sum_{i,j=1}^k 1_{\{\|s_i - s_j\| - h \leq \varepsilon\}}.$$

Variogramme empirique sur les données de Calcium

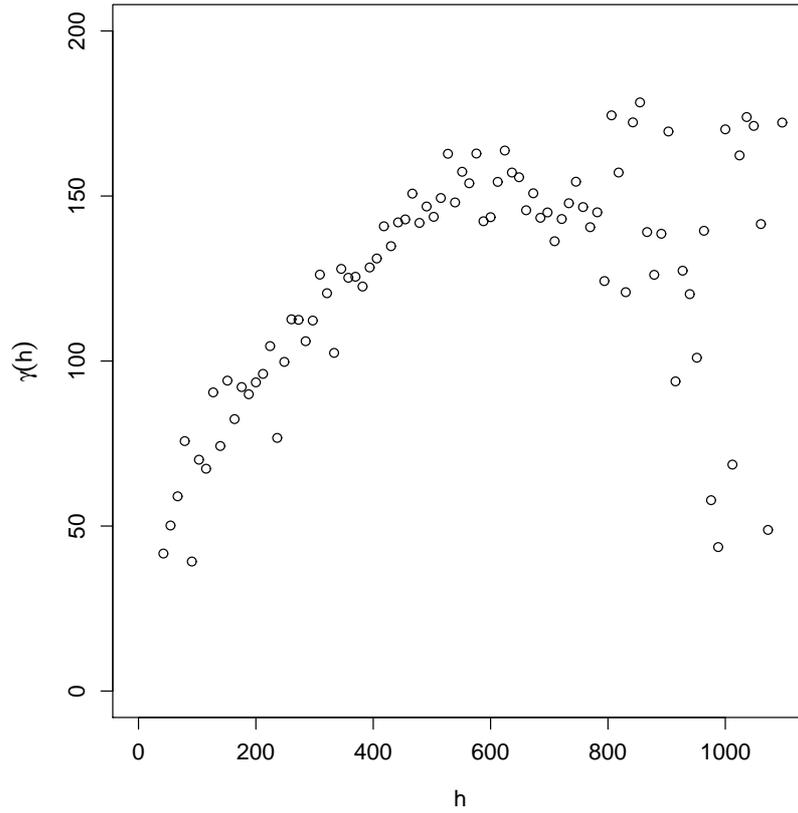


Figure 9: Variogramme empirique sur les données de Calcium.

Variogramme empirique sur les données de Calcium

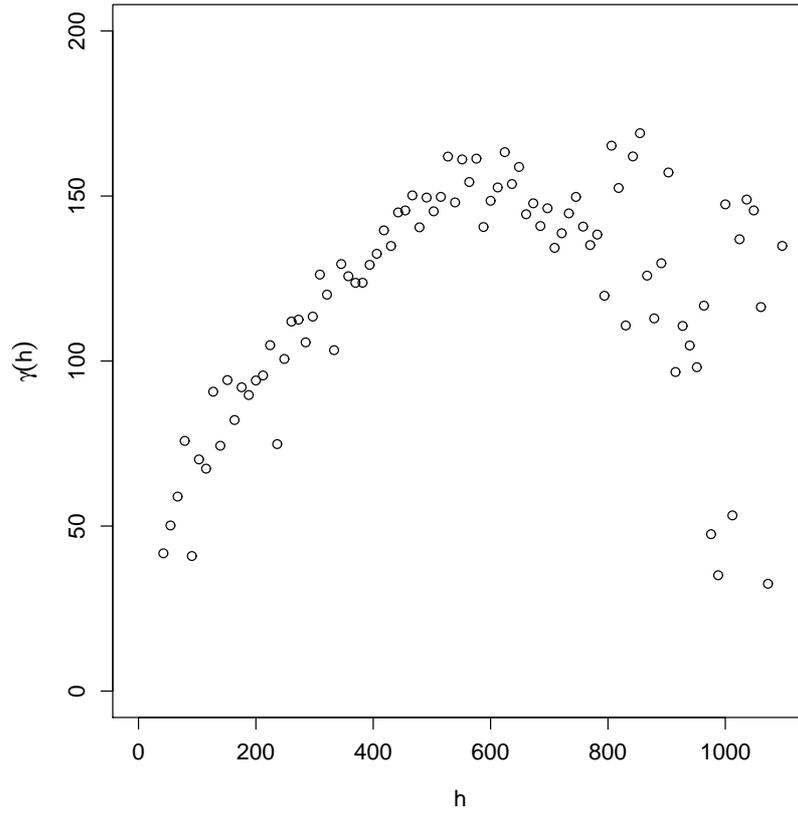


Figure 10: Variogramme empirique sur les données de Calcium (tendance longitude).

Variogramme empirique sur les données de Calcium

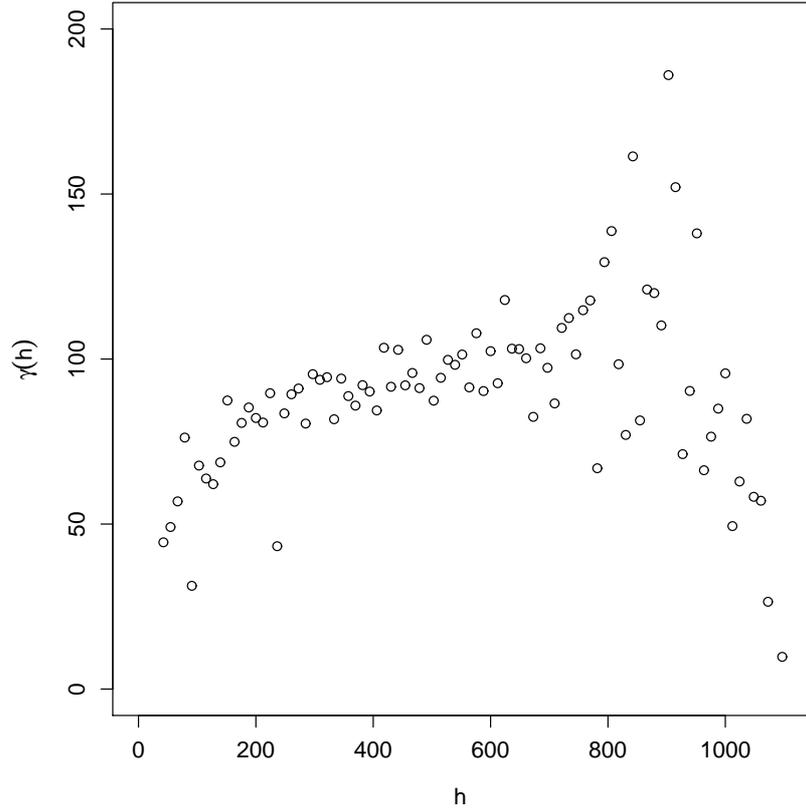


Figure 11: Variogramme empirique sur les données de Calcium (tendance latitude).

Variogramme empirique sur les données de Calcium

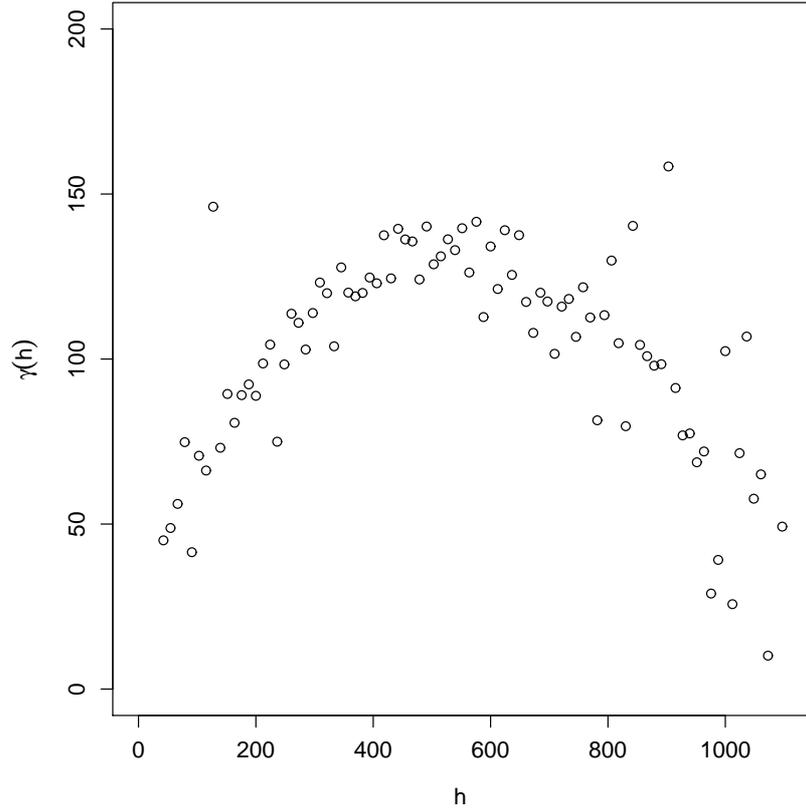


Figure 12: Variogramme empirique sur les données de Calcium (tendance altitude).

Variogramme empirique sur les données de Calcium

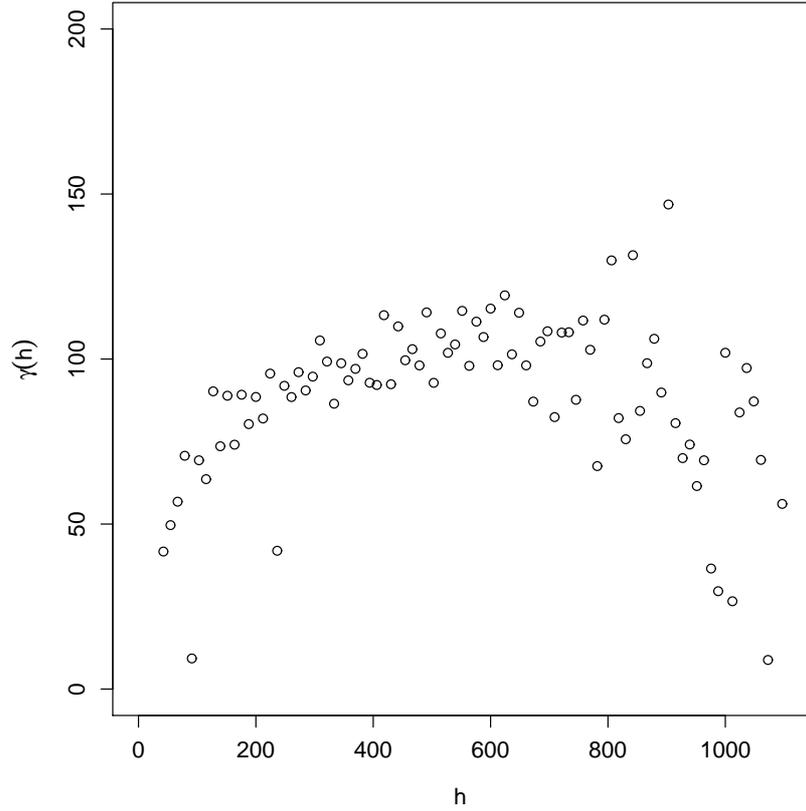


Figure 13: Variogramme empirique sur les données de Calcium (tendance sous-region).

Variogramme empirique sur les données de Calcium

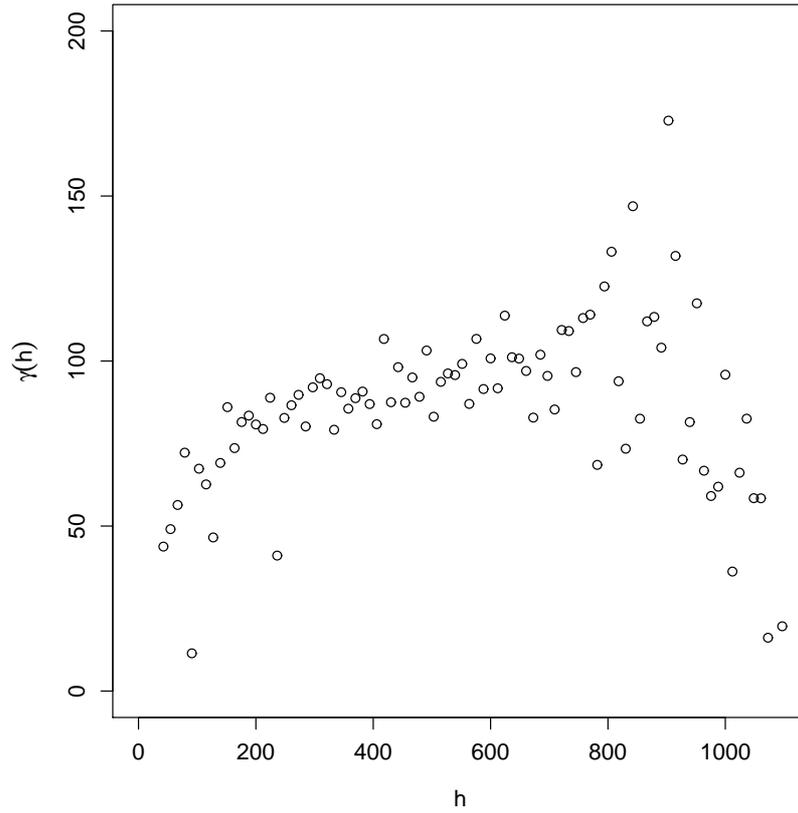


Figure 14: Variogramme empirique sur les données de Calcium (tendance sous-region + latitude).

Message à retenir

- Attention à l'effet des covariables explicatives sur le variogramme. Ces dernières "réduisent" la structure de dépendance et diminuent la variance.

Pensez à

$$Z(s) = \mathbf{X}(s)^\top \beta + \varepsilon(s), \quad s \in \mathcal{X},$$

où $\{\varepsilon(s) : s \in \mathcal{X}\}$ est un processus faiblement stationnaire de fonction de covariance $K_\varepsilon(\cdot)$. Alors clairement

$$\begin{aligned} \mathbb{E} [\{Z(s_1) - Z(s_2)\}^2] &= \mathbb{E} [\{Z(s_1) - \mu(s_1) - Z(s_2) + \mu(s_2) + \mu(s_1) - \mu(s_2)\}^2] \\ &= 2K_\varepsilon(o) + \{\mu(s_1) - \mu(s_2)\}^2 + 2K_\varepsilon(s_1 - s_2) \\ &\geq 2K_\varepsilon(o) - 2K_\varepsilon(s_1 - s_2) \\ &= 2\gamma_\varepsilon(s_1 - s_2). \end{aligned}$$

Ajustement d'un variogramme

- Le principe n'est pas vraiment nouveaux : utilisons les moindres carrés, i.e.,
 1. Choix d'une famille paramétrique de variogramme $\gamma(\cdot; \psi)$
 2. Résolution du problème de minimisation

$$\hat{\psi} = \arg \min_{\psi \in \Psi} \sum_j \{\hat{\gamma}(h_j) - \gamma(h_j; \psi)\}^2$$

Remarque. Il sera souvent préférable d'utiliser les moindres carrés pondérés, i.e.,

$$\hat{\psi} = \arg \min_{\psi \in \Psi} \sum_j \frac{N(h_j, \varepsilon)}{\gamma(h_j, \psi)^2} \{\hat{\gamma}(h_j) - \gamma(h_j; \psi)\}^2.$$

Application

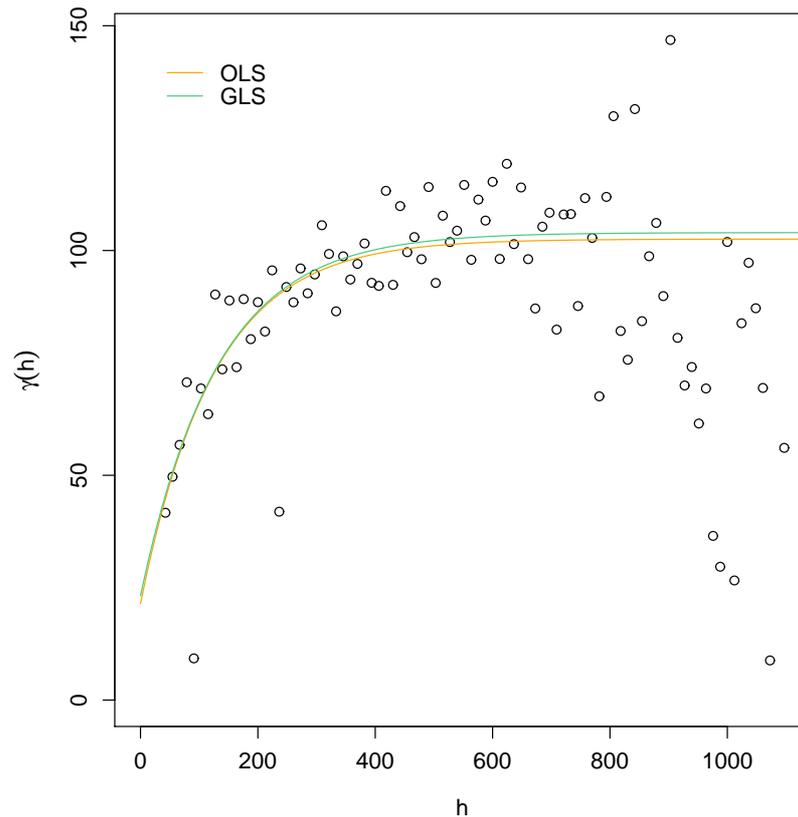
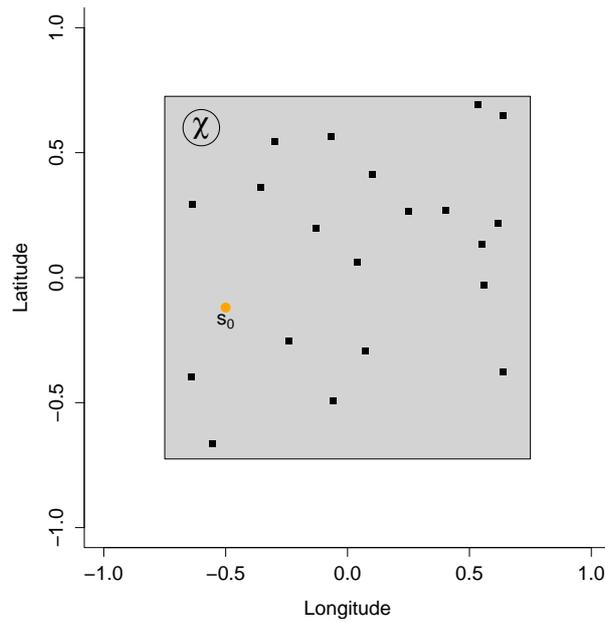


Figure 15: Ajustement par moindres carrés d'un variogramme lié à la covariance exponentielle pour les données de teneur en Ca. (tendance sous-région)

Message à retenir

- Le problème d'optimisation n'est pas très "aimable" et il sera donc souvent utile de tester les estimations en partant de différentes valeurs initiales ;
- Il est d'usage lorsqu'on suppose une famille de covariance de type Whittle–Matérn de fixer le paramètre de forme aux valeurs $\kappa = 0.25, 0.5, \dots, 2$. Ce paramètre étant particulièrement difficile à estimer.
- Il est toujours important de se demander si un effet de pépite est réellement souhaitable.

Krigeage: Objectif



- Prédire $Z(s_0)$ sachant $Z(s_1), \dots, Z(s_k)$.

Géostatistique

Mathieu Ribatet – 50 / 70

Prédicteurs linéaires

- Infinité de possibilités pour estimer $Z(s_0) \Rightarrow$ restriction au cadre des **estimateurs linéaires sans biais**, i.e.,

$$\hat{Z}(s_0) = \sum_{j=1}^k \lambda_j Z(s_j), \quad \mathbb{E}\{\hat{Z}(s_0)\} = \mu(s_0).$$

- Parmi la classe de ces estimateurs, on essaiera de déterminer le “meilleur” au sens de l'**erreur quadratique**

$$\hat{Z}(s_0) = \arg \min_{T \in Est.Lin.} \mathbb{E}[\{T - Z(s_0)\}^2].$$

Remarque. On sait que le meilleur estimateur au sens erreur quadratique est toujours $T = \mathbb{E}\{Z(s_0) \mid Z(s_1), \dots, Z(s_k)\}$. Cela dit (sauf cas gaussien par ex.), ce n'est pas toujours un estimateur linéaire...

Géostatistique

Mathieu Ribatet – 51 / 70

Krigeage ordinaire

- Supposons que $\{Z(s) : s \in \mathcal{X}\}$ est faiblement stationnaire de moyenne inconnue $\mu(s) \equiv \mu$ et de covariance (connue) $K(\cdot)$.
- Pour que l'estimateur soit sans biais il faut

$$\mathbb{E}\{\hat{Z}(s_0)\} = \sum_{j=1}^k \lambda_j \mathbb{E}\{Z(s_j)\} = \mu \Rightarrow \sum_{j=1}^k \lambda_j = 1.$$

- Et l'on souhaite donc minimiser

$$\begin{aligned} \text{Var}\{\hat{Z}(s_0) - Z(s_0)\} &= \sum_{i,j=1}^k \lambda_i \lambda_j K(s_i - s_j) + K(o) - 2 \sum_{i=1}^k \lambda_i K(s_i - s_0) \\ &= \lambda^\top K(\mathbf{s}) \lambda + K(o) - 2 \lambda^\top K(\mathbf{s} - s_0), \end{aligned}$$

sous la contrainte $\lambda^\top \mathbf{1} = 1$ avec $\lambda = (\lambda_1, \dots, \lambda_k)$ et $\mathbf{s} = (s_1, \dots, s_k)$,
 $K(\mathbf{s}) = \{K(s_i - s_j)\}_{i,j=1,\dots,k}$.

Krigeage ordinaire

Proposition 6 (DM). *L'estimateur du krigeage ordinaire est donné par*

$$\hat{Z}(s_0) = K(\mathbf{s} - s_0)^\top K(\mathbf{s})^{-1} Z(\mathbf{s}) + \frac{\mathbf{1}^\top K(\mathbf{s})^{-1} Z(\mathbf{s})}{\mathbf{1}^\top K(\mathbf{s})^{-1} \mathbf{1}} \left\{ 1 - \mathbf{1}^\top K(\mathbf{s})^{-1} K(\mathbf{s} - s_0) \right\},$$

où $Z(\mathbf{s}) = \{Z(s_1), \dots, Z(s_k)\}$.

De plus la variance de l'erreur est donnée par

$$\text{Var}\{\hat{Z}(s_0) - Z(s_0)\} = K(o) - K(\mathbf{s} - s_0)^\top K(\mathbf{s})^{-1} K(\mathbf{s} - s_0) + \frac{\{1 - \mathbf{1}^\top K(\mathbf{s})^{-1} K(\mathbf{s} - s_0)\}^2}{\mathbf{1}^\top K(\mathbf{s})^{-1} \mathbf{1}}$$

Le krigeage ordinaire en pratique

1. Calcul du variogramme empirique ;
2. Ajustement d'un variogramme paramétrique ;
3. Calculez $K(s)$ ainsi que son inverse ;
4. Pour chaque point s_0 d'un grille de \mathcal{X}
 - (a) Calcul de $K(s - s_0)$;
 - (b) Calcul de $\hat{Z}(s_0)$ et de la variance de l'erreur;
5. Faire de jolies cartes ;

Géostatistique

Mathieu Ribatet – 54 / 70

Quantité de Calcium contenu dans le sol

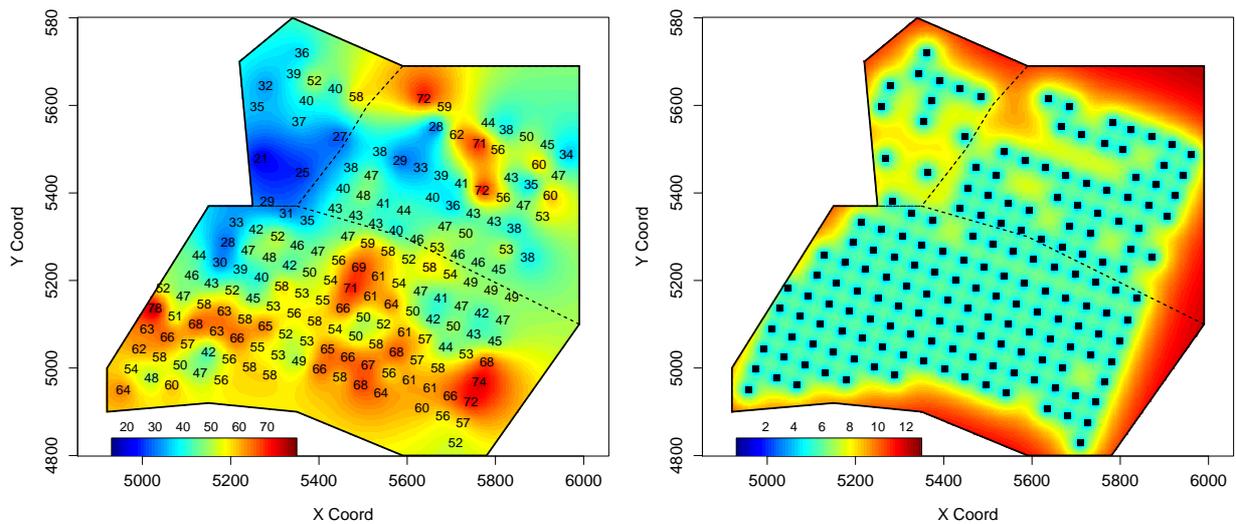


Figure 16: Krigeage ordinaire et erreur de prédiction pour les données de teneur en Ca.

Géostatistique

Mathieu Ribatet – 55 / 70

Question pour vous. . .

- Que pensez vous de ce que nous venons de faire ?

Géostatistique

Mathieu Ribatet – 56 / 70

Krigeage universel

- Nous supposons désormais que

$$Z(s) = \mathbf{X}(s)^\top \beta + \varepsilon(s), \quad s \in \mathcal{X},$$

où $\{\varepsilon(s) : s \in \mathcal{X}\}$ est un processus faib. stat. de moyenne nulle et de covariance (connue) $K(\cdot)$, β vecteur de paramètre à estimer et $\mathbf{X}(s)$ vecteur de covariables au point $s \in \mathcal{X}$.

- $\hat{Z}(s_0) = \sum_{j=1}^k \lambda_j Z(s_j)$ est sans biais :

$$\begin{aligned} \mathbb{E}\{\hat{Z}(s_0) - Z(s_0)\} &= \sum_{j=1}^k \lambda_j \mathbf{X}(s_j)^\top \beta - \mathbf{X}(s_0)^\top \beta \\ &= \left\{ \sum_{j=1}^k \lambda_j \mathbf{X}(s_j) - \mathbf{X}(s_0) \right\}^\top \beta \\ &= 0, \quad \text{pour tout } \beta. \end{aligned}$$

Géostatistique

Mathieu Ribatet – 57 / 70

Krigeage universel

- L'équation précédente impose donc les p contraintes suivantes

$$\sum_{j=1}^k \lambda_j \mathbf{X}_\ell(s_j) - \mathbf{X}_\ell(s_0), \quad \ell = 1, \dots, p,$$

où $\mathbf{X}(s) = \{\mathbf{X}_1(s), \dots, \mathbf{X}_p(s)\}$ —i.e., p covariables.

- On souhaite alors minimiser

$$\text{Var}\{\hat{Z}(s_0) - Z(s_0)\}$$

avec les p contraintes précédentes.

Krigeage universel

- Après quelques calculs pénibles mais simples, on trouve alors pour $\lambda = (\lambda_1, \dots, \lambda_k)$

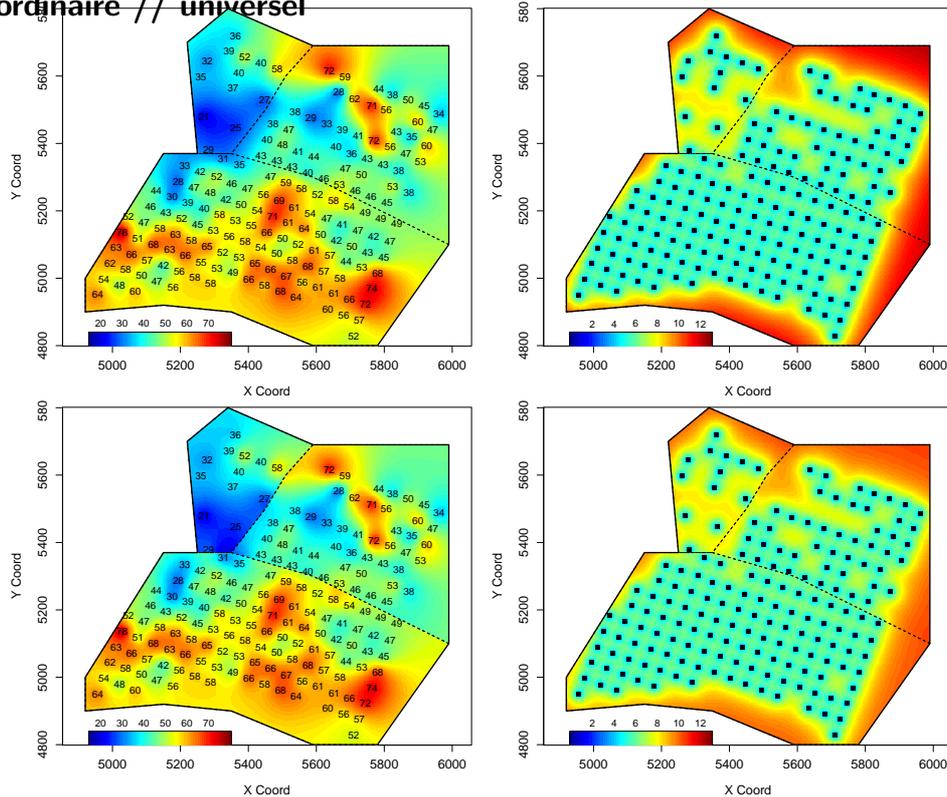
$$\lambda = K(\mathbf{s})^{-1}K(\mathbf{s} - s_0) - K(\mathbf{s})^{-1}\mathbf{X}(\mathbf{s})\{\mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1}\mathbf{X}(\mathbf{s})\}^{-1}\{\mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1}K(\mathbf{s} - s_0) - \mathbf{X}(s_0)\}.$$

- La variance de l'erreur de l'estimateur $\text{Var}\{\hat{Z}(s_0) - Z(s_0)\}$ est donnée par

$$K(o) - K(\mathbf{s} - s_0)^\top K(\mathbf{s})^{-1}K(\mathbf{s} - s_0) + \{\mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1}K(\mathbf{s} - s_0) - \mathbf{X}(s_0)\}^\top \{\mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1}\mathbf{X}(\mathbf{s})\}^{-1}\{\mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1}K(\mathbf{s} - s_0) - \mathbf{X}(s_0)\}.$$

Remarque. On retombe sur le krigeage ordinaire en posant $\mathbf{X}(s) \equiv 1$.

Krigeage ordinaire // universel



Géostatistique

Mathieu Ribatet – 60 / 70

Modélisation par processus gaussiens

- Ce que nous avons vu précédemment est la vision **École des Mines de Paris**.
- L'approche anglo-saxonne consiste à poser un modèle et non pas seulement supposer une stationnarité faible ou intrinsèque.
- Le plus souvent on supposera que

$$\Phi\{Z(s)\} = f(s; \beta) + \varepsilon(s), \quad s \in \mathcal{X},$$

où $\Phi(\cdot)$ est une fonction lien supposée connue (comme pour les GLM), $\{\varepsilon(s) : s \in \mathcal{X}\}$ un processus Gaussien centré de covariance $K(\cdot)$ et $f(\cdot; \beta)$ une fonction paramétrée par un vecteur de paramètre β .

☞ Nous nous restreindrons au cas où $\Phi : z \mapsto z$ et $f(s; \beta) = \mathbf{X}(s)\beta$.

Géostatistique

Mathieu Ribatet – 61 / 70

Vraisemblance

La log-vraisemblance est donnée par

$$\ell(\psi) = -\frac{nk}{2} \log 2\pi - \frac{n}{2} \log |K(\mathbf{s})| - \frac{1}{2} \sum_{\ell=1}^n \{z_{\ell}(\mathbf{s}) - \mathbf{X}(\mathbf{s})\beta\}^{\top} K(\mathbf{s})^{-1} \{z_{\ell}(\mathbf{s}) - \mathbf{X}(\mathbf{s})\beta\}$$

Remarque. Numériquement il sera plus efficace de calculer les formes quadratiques plus haut par la formule

$$\begin{aligned} y^{\top} K(\mathbf{s})^{-1} y &= y^{\top} \{C(\mathbf{s})C(\mathbf{s})^{\top}\}^{-1} y \\ &= y^{\top} C(\mathbf{s})^{-\top} C(\mathbf{s})^{-1} y \\ &= x(\mathbf{s})^{\top} x(\mathbf{s}), \end{aligned}$$

où $C(\mathbf{s})$ est la décomposition de Cholesky de $K(\mathbf{s})$ et $x(\mathbf{s})$ la solution du système **triangulaire** $C(\mathbf{s})x(\mathbf{s}) = y$. De plus $|K(\mathbf{s})| = |C(\mathbf{s})|^2 = \prod_{j=1}^k C(\mathbf{s})_{jj}^2$.

Géostatistique

Mathieu Ribatet – 62 / 70

DM

Écrivez une fonction R calculant l'estimateur du maximum de vraisemblance d'un processus Gaussien défini plus haut.

Géostatistique

Mathieu Ribatet – 63 / 70

Rappel : Lois gaussiennes conditionnelles

Proposition 7. Soit $Z(\mathbf{s}, \mathbf{s}_0)$ un vecteur Gaussien de $\mathbb{R}^{k \times k_0}$ d'espérance $\mu(\mathbf{s}, \mathbf{s}_0)$ et matrice de covariance $K(\mathbf{s}, \mathbf{s}_0)$. Alors

$$Z(\mathbf{s}_0) | \{Z(\mathbf{s}) = z(\mathbf{s})\} \sim N \left\{ \tilde{\mu}(\mathbf{s}_0), \tilde{K}(\mathbf{s}_0) \right\},$$

avec

$$\begin{aligned} \tilde{\mu}(\mathbf{s}_0) &= \mu(\mathbf{s}_0) + K(\mathbf{s}_0, \mathbf{s})K(\mathbf{s})^{-1}\{z(\mathbf{s}) - \mu(\mathbf{s})\} \\ \tilde{K}(\mathbf{s}_0) &= K(\mathbf{s}_0) - K(\mathbf{s}_0, \mathbf{s})K(\mathbf{s})^{-1}K(\mathbf{s}_0, \mathbf{s})^\top, \end{aligned}$$

où nous avons utilisé la décomposition suivante

$$K(\mathbf{s}_0, \mathbf{s}) = \begin{bmatrix} K(\mathbf{s}_0) & K(\mathbf{s}_0, \mathbf{s}) \\ K(\mathbf{s}_0, \mathbf{s})^\top & K(\mathbf{s}) \end{bmatrix}.$$

Lien avec le Krigeage

- L'estimateur $\hat{Z}(s_0) = \mathbb{E}\{Z(s_0) | Z(\mathbf{s})\}$ minimise l'erreur quadratique moyenne. Ainsi dans le cas gaussien faib. stat.,

$$\hat{Z}(s_0) = \tilde{\mu}(s_0) = \mu(s_0) + K(\mathbf{s} - s_0)^\top K(\mathbf{s})^{-1}\{Z(\mathbf{s}) - \mu(\mathbf{s})\}.$$

- Pour $\mu(\mathbf{s}) = \mathbf{X}(\mathbf{s})^\top \hat{\beta}$, et à covariance connue, on sait que $\hat{\beta} = \{\mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} \mathbf{X}(\mathbf{s})\}^{-1} \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} Z(\mathbf{s})$,

$$\begin{aligned} \hat{Z}(s_0) &= \mathbf{X}(s_0)^\top \{ \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} \mathbf{X}(\mathbf{s}) \}^{-1} \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} Z(\mathbf{s}) + \\ &\quad K(\mathbf{s} - s_0)^\top K(\mathbf{s})^{-1} \left[Z(\mathbf{s}) - \mathbf{X}(\mathbf{s})^\top \{ \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} \mathbf{X}(\mathbf{s}) \}^{-1} \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} Z(\mathbf{s}) \right] \\ &= \left(\mathbf{X}(s_0)^\top \{ \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} \mathbf{X}(\mathbf{s}) \}^{-1} \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} + \right. \\ &\quad \left. K(\mathbf{s} - s_0)^\top K(\mathbf{s})^{-1} \left[\mathbf{1} - \mathbf{X}(\mathbf{s})^\top \{ \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} \mathbf{X}(\mathbf{s}) \}^{-1} \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} \right] \right) Z(\mathbf{s}) \\ &= \left(\left\{ \mathbf{X}(s_0)^\top - K(\mathbf{s} - s_0) K(\mathbf{s})^{-1} \mathbf{X}(\mathbf{s})^\top \right\} \{ \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} \mathbf{X}(\mathbf{s}) \}^{-1} \mathbf{X}(\mathbf{s})^\top K(\mathbf{s})^{-1} + \right. \\ &\quad \left. K(\mathbf{s} - s_0)^\top K(\mathbf{s})^{-1} \right) Z(\mathbf{s}). \end{aligned}$$

- On retombe sur l'estimateur du krigeage universel !

Limitation de la vraisemblance

- La vraisemblance fait intervenir une décomposition de Cholesky—coût algorithmique $O(k^3)$.
- Ceci restreint son utilisation au cas où $k \leq 3000$.
- D'autres approches sont possibles lorsque $k > 3000$:
 - Processus Gaussien avec $K(\mathbf{s})$ creuse ;
 - Utilisation de vraisemblances composites.

3. Simulations de processus gaussiens

Motivations

- Les simulations sont souvent utiles afin d'estimer par ex.

$$\Psi = T [\mathbb{E} \{Z(\cdot) \mid Z(\mathbf{s}) = \mathbf{z}\}],$$

où T est une fonctionnelle supposée connue.

- Puisque le krigeage $\hat{Z}(\cdot)$ est un estimateur sans biais de $\mathbb{E}\{Z(\cdot) \mid Z(\mathbf{s})\}$, si T est **linéaire** alors l'estimateur $\hat{\Psi} = T\{\hat{Z}(\cdot)\}$ vérifie

$$\mathbb{E}(\hat{\Psi}) = T [\mathbb{E}\{\hat{Z}(\cdot)\}] = T [\mathbb{E}\{Z(\cdot) \mid Z(\mathbf{s}) = \mathbf{z}\}] = \Psi.$$

- Ce n'est plus vrai si T n'est pas linéaire ! Une solution consiste donc à estimer Ψ par des méthodes de type Monte-Carlo.

Simulations non conditionnelles

- Il s'agit ici de simuler selon la loi de $\{Z(s) : s \in \mathcal{X}\}$.
- Trois grandes approches :
 - Approche directe via une décomposition de Cholesky ;
 - Approche par matrices circulantes et FFT ;
 - Approche par bandes tournantes.
- Nous allons

Simulations conditionnelles