

Chapitre XXII La Covariance

Covariogramme géométrique

Version probabiliste :

- ensembles
- fonctions
- propriétés de la covariance

Théorie intrinsèque :

- variogramme

Covariogramme géométrique

- Soit dans \mathbb{R}^n le vecteur \vec{h} d'origine 0, de module h et de direction α . Prenons pour élément structurant B le *bipoint* $\{ 0, h \}$ et considérons l'érosion

$$X \ominus B = X \cap X_{-h} = \{ x \in \mathbb{R}^n, x \in X \cap X_{-h} \}$$

$X \ominus B$ est l'ensemble des points communs à X et à son translaté par le vecteur \vec{h} .

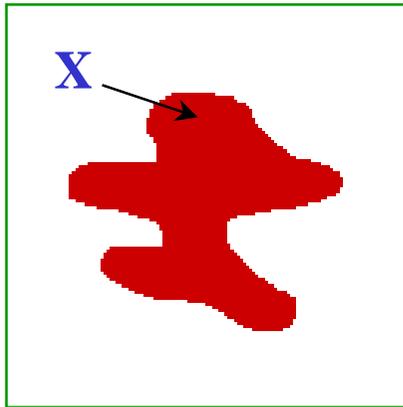
- *Définition* : On appelle *covariogramme géométrique* de X la mesure de Lebesgue de $X \cap X_{-h}$, et l'on note:

$$\vec{K}(h) = K_{\alpha}(h) = \text{Mes} [X \cap X_{-h}] = \text{Mes} [X \cap X_h]$$

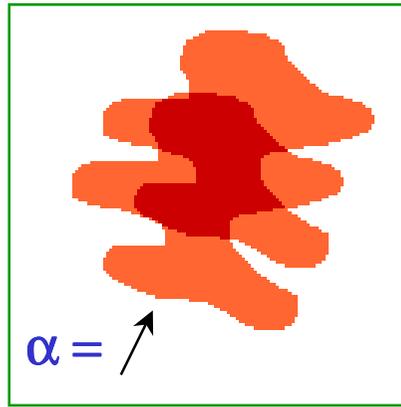
- En particulier $K_{\alpha}(\mathbf{0}) = \text{Mes } X$ et $\int \vec{K}(h) dh = [\text{Mes } X]^2$

L'intégrale du covariogramme vaut le carré de la surface (du volume) de X .

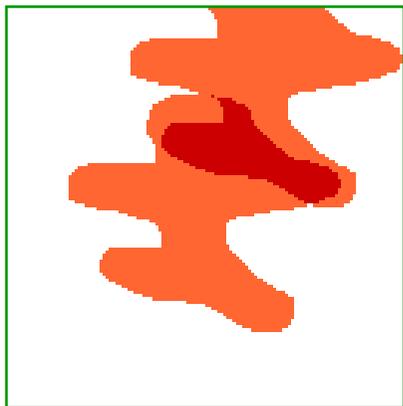
Erosion par un bipoint : exemple



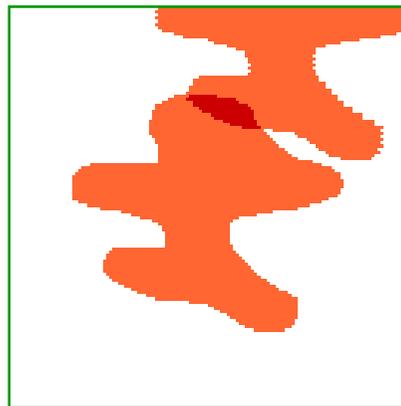
$h = 0$



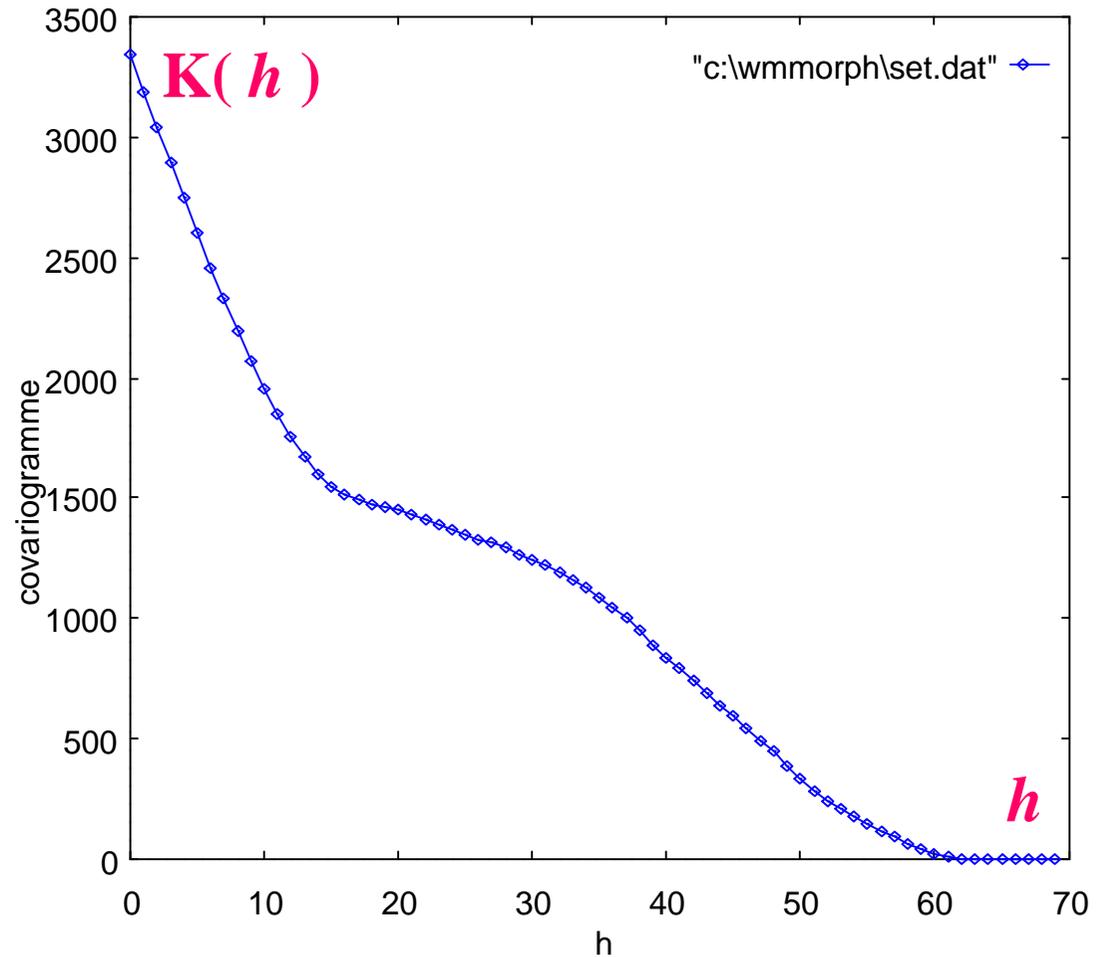
$h = 20$



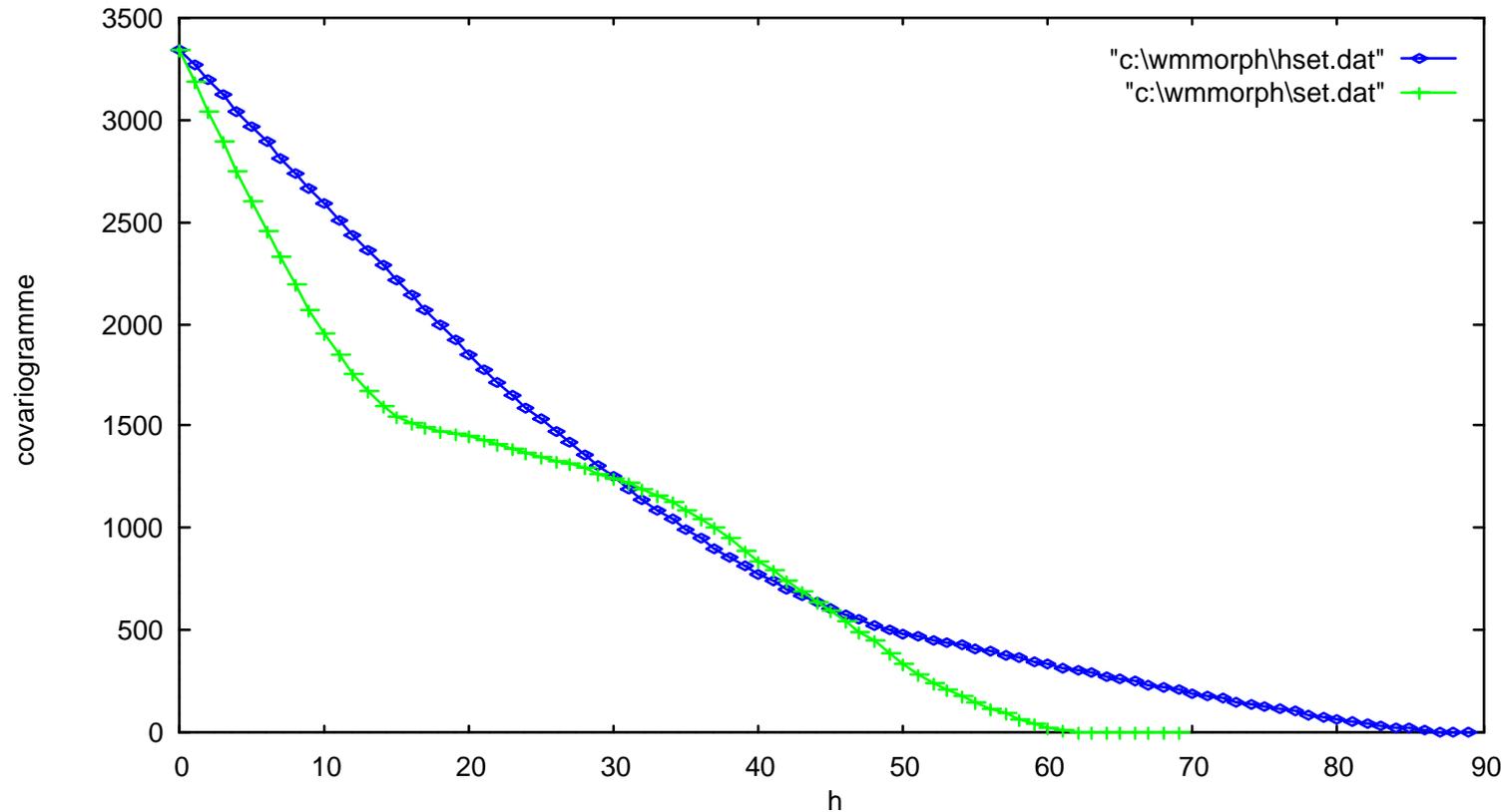
$h = 40$



$h = 60$

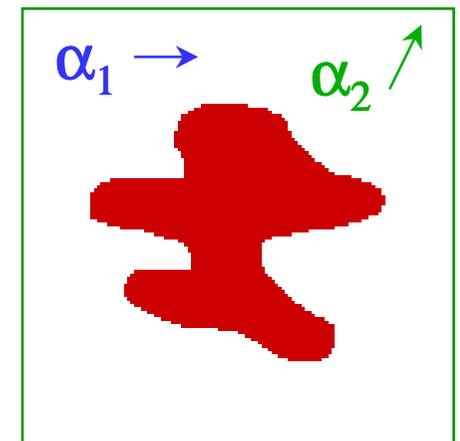


Anisotropies du covariogramme



$$\alpha_1 = 0^\circ$$

$$\alpha_2 = 60^\circ$$



- La tangente à l'origine $-K_\alpha'(0) = - [\partial K_\alpha(h) / \partial h]_{h=0}$ n'est autre que la **variation diamétrale** de X dans la direction α , et dans le cas convexe son contour apparent. On retrouve ici la **formule de Crofton**:

$$-\int K_\alpha'(0) d\alpha = 2U(X) \quad (\text{dans } R^2) \quad ; \quad -\int K_\alpha'(0) d\alpha = \pi S(X) \quad (\text{dans } R^3)$$

Versions probabilistes

- Associons à l'ensemble X sa **fonction indicatrice** f . On a les égalités:

$$K_{\alpha}(h) = \text{Mes}\{X \cap X_h\} = \int f(x) \cdot f(x+h) dx$$

la sommation étant étendue à tout l'espace. Si maintenant f n'est plus une indicatrice, mais une **fonction numérique** mesurable, l'intégrale ci-dessus *définit* son covariogramme.

- Il arrive que l'objet représenté par f
 - puisse être *reproduit à volonté* (e.g. un cristal de quartz)
 - ou, même s'il est unique, *se répète indéfiniment* à travers l'espace (e.g. une mosaïque de champs vue d'avion).
- On l'interprète alors comme une réalisation de **fonction aléatoire**. Dans le premier cas, pour f *p.s.* de carré sommable, le covariogramme devient

$$K_{\alpha}(h) = E[\text{Mes}\{X \cap X_h\}] \quad K_{\alpha}(h) = E[\int f(x) \cdot f(x+h) dx]$$

(ensembles) *(fonctions)*

Cadre probabiliste stationnaire

- Lorsque l'objet d'étude est grand par rapport au masque de mesures, il est préférable de l'interpréter en termes d'ensemble fermé ou de fonction aléatoire **stationnaire**. Le $K_\alpha(h)$ est alors remplacée par la **covariance**

$$C(h, \alpha) = E [(f(x) - m) \cdot (f(x+h) - m)] = E [f(x) \cdot f(x+h)] - m^2 .$$

où $m = E [f(x)]$ l'espérance du processus.

- Pour $h = 0$, il vient $C(0, \alpha) = E [(f(x) - m)]^2 = \sigma^2$

La valeur à l'origine de la covariance est égale à la **variance du processus**. Dans le cas d'un ensemble de proportion volumique p , on a

$$C(0, \alpha) = p(1 - p)$$

- Pour $h = \infty$, il n'y a plus de corrélation entre x et $x+h$; donc

$$C(h, \alpha) = E [(f(x) - m)] \cdot E[(f(x+h) - m)] = 0$$

Pour plus de clarté, nous examinons ci-dessous d'abord le cas des ensembles, puis celui des fonctions.

Covariance ensembliste

- Lorsque l'objet d'étude est modélisé par un **fermé aléatoire stationnaire** X , la notion qui correspond directement au covariogramme $K_\alpha(h)$ est la covariance **non centrée**

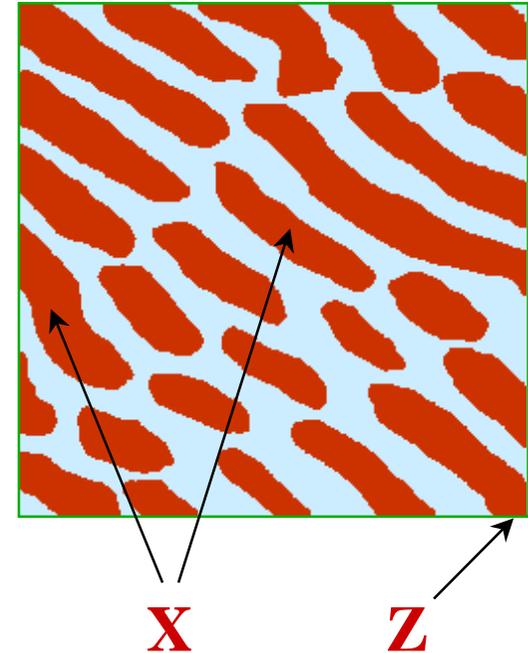
$$C_1(h, \alpha) = C(h, \alpha) + p^2 = \text{Prob} \{ \mathbf{x} \in X \cap X_h \}.$$

- Expérimentalement, la covariance ensembliste non centrée dans un champ Z s'estime comme le rapport des implantations favorables du doublet B

$$N = A [(X \cap Z) \ominus B]$$

sur le total de ses implantations dans Z

$$D = A [Z \ominus B] \quad \text{i. e.} \quad C_1(h, \alpha)^* = N/D .$$



Exemple d'objet X dont les dimensions excèdent celles du champ Z .

Propriétés de la covariance ensembliste

On retrouve les propriétés du $K(h)$, enrichies d'interprétations probabilistes :

- A l'**origine** l'on a

$$C_1(\mathbf{0}, \alpha) = A_A(\mathbf{X}) = p \quad (\text{quel que soit } \alpha) ;$$

- Quand h est **très grand**, les évènements $x \in X$ et $x \in X_h$ deviennent indépendants, d'où

$$\lim C_1(h, \alpha) = p^2 = [A_A(\mathbf{X})]^2 \quad .$$

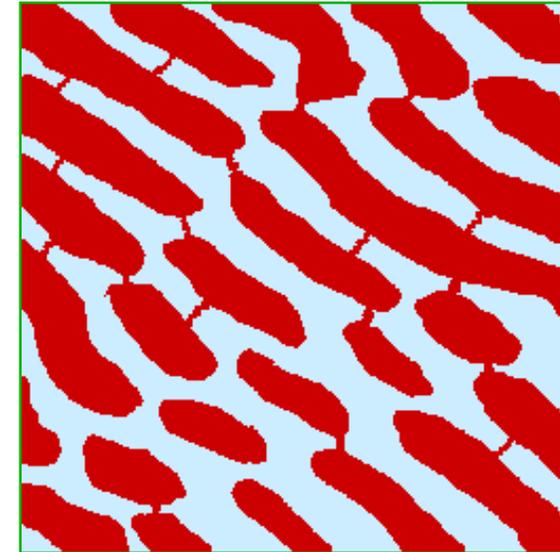
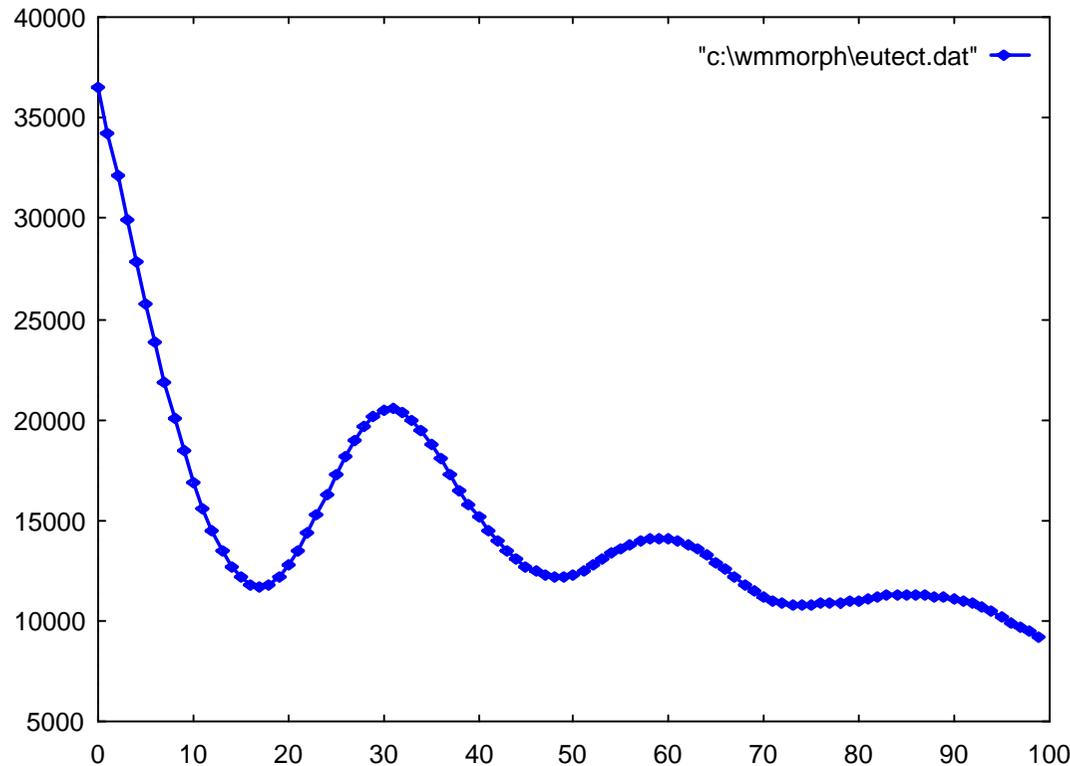
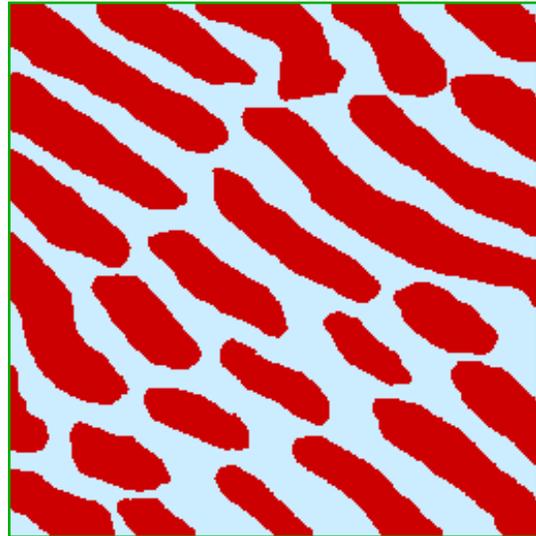
- La moyenne directionnelle de la tangente à l'origine vaut le **périmètre (la surface) spécifique**

$$-\int C_1'(\mathbf{0}, \alpha) d\alpha = (2/\pi)U_L(\mathbf{X}) = S_V(\mathbf{X})/4$$

(dans R^2) (dans R^3)

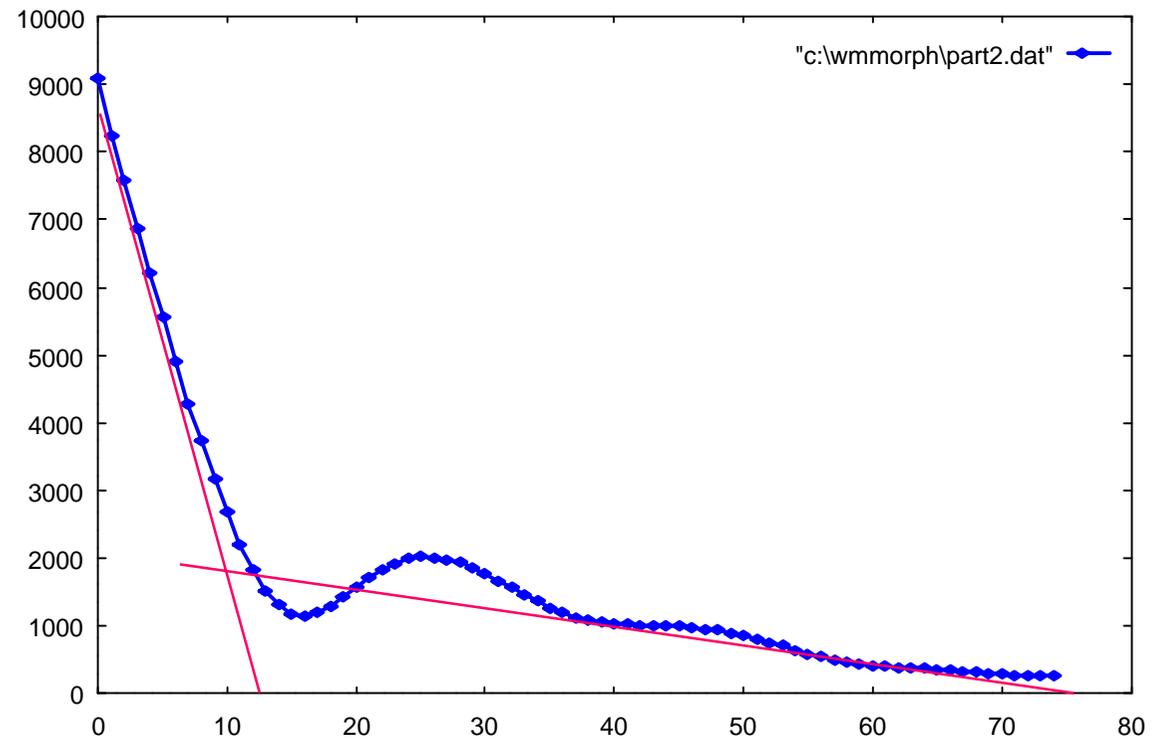
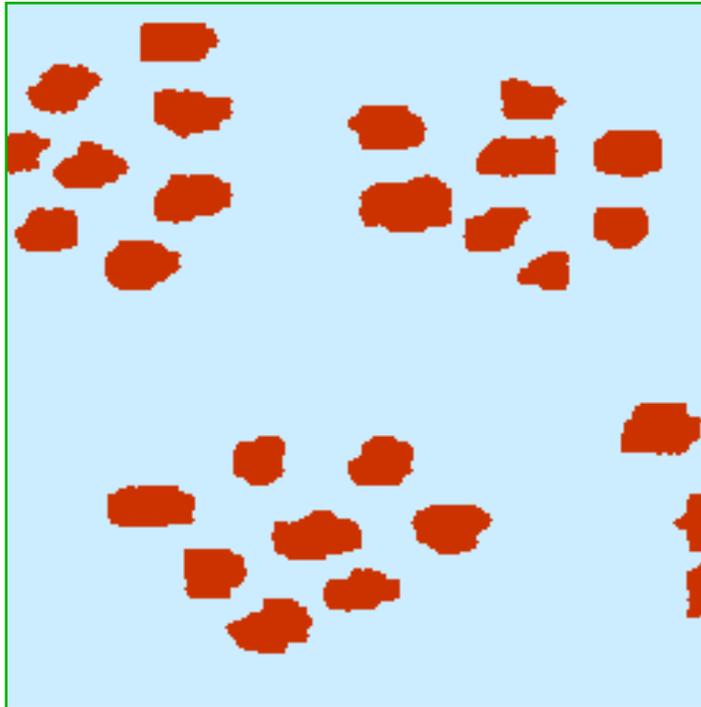
- La **portée** $a = 1/p(1-p) \int [C_1(h) - p^2] dh$ a une **signification ergodique**
(cf. plus bas XII-21)

Pseudo-périodicités et covariance



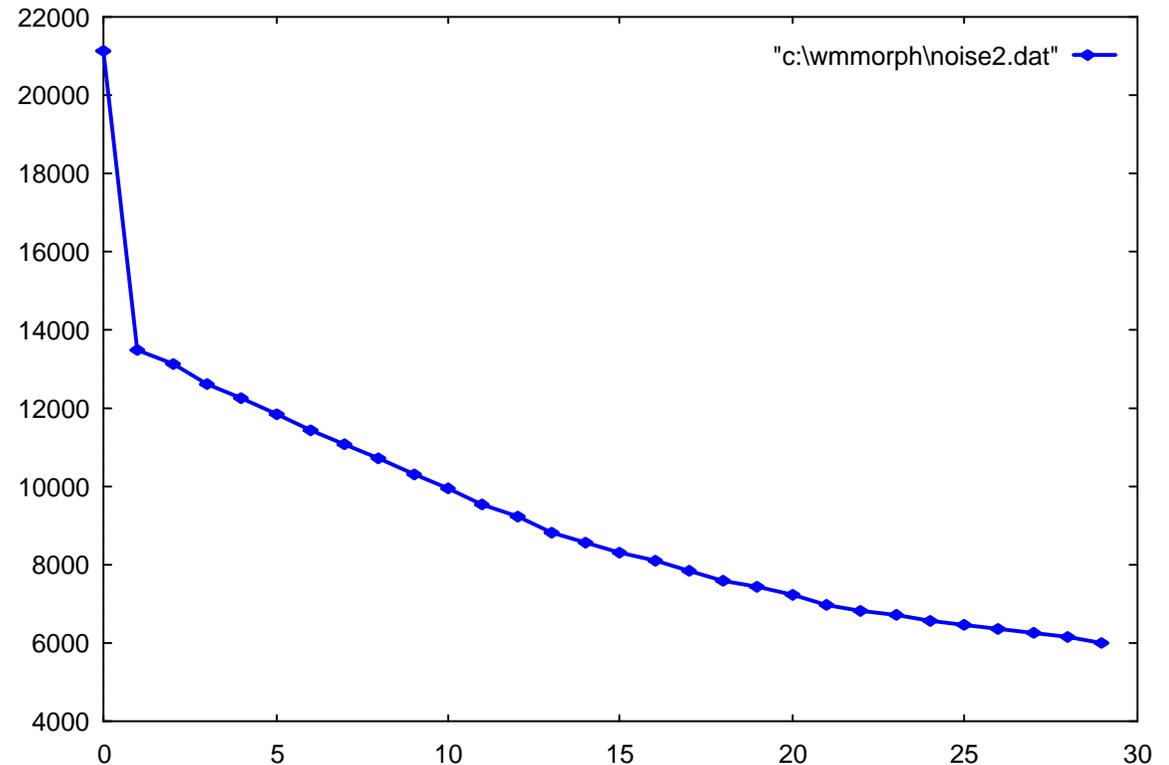
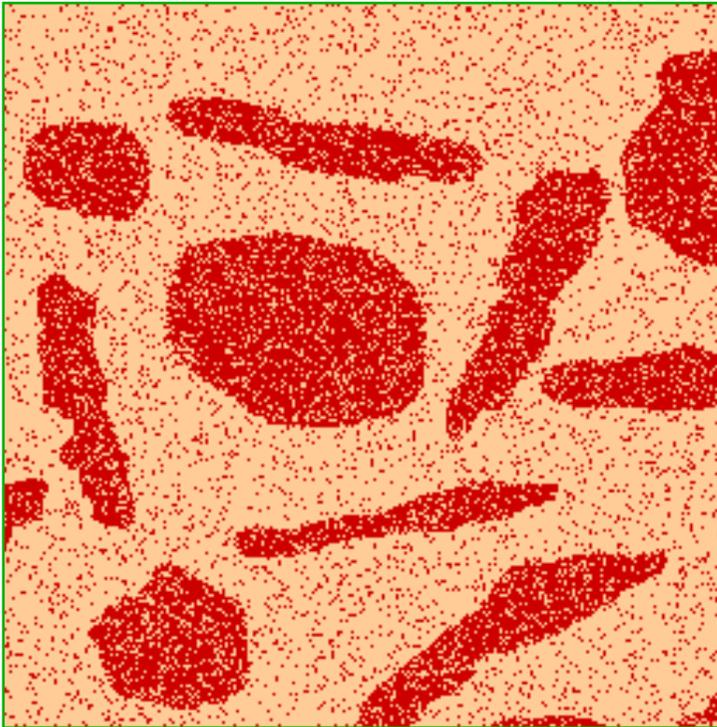
- Les maxima de la covariance, s'ils sont répétés (ici aux abscisses 30, 60, 90) indiquent des (pseudo) **périodes** de l'objet;
- En revanche la covariance est **aveugle à la connexité** : les deux images présentées ont deux covariances graphiquement indiscernables

Groupements et covariance



- Les objets d'échelles différentes (ici, les particules et leurs groupes) ont des covariances qui se superposent (en rouge, les tangentes à l'origine associées)
- En plus, apparaît ici une oscillation due à la valeur égale des distances entre particules dans leurs paquets.

Bruit et covariance

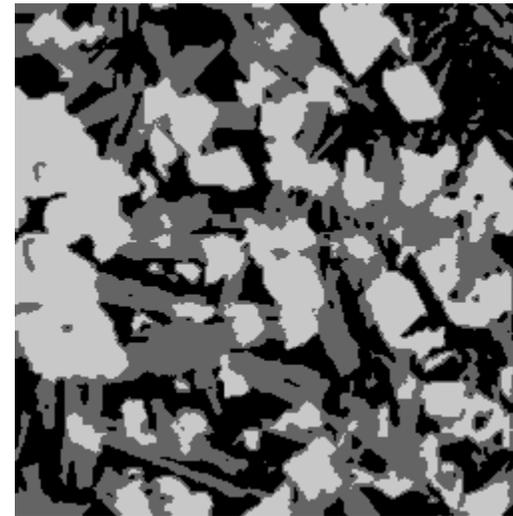


- Cette situation est limite de la précédente: la plus petite échelle est réduite à du **bruit ponctuel**, impulsionnel, de type Poisson.
- En digital, la portée du bruit vaut donc 1, et sa **variance** $p_0(1-p_0)$ est égale à l'ordonnée du décrochement.

Covariance rectangle

- La division de l'espace en un ensemble et son complémentaire est parfois remplacée par une partition en n phases $X_1, X_2 \dots X_n$. L'étude de leurs relations spatiales s'effectue alors par la covariance rectangle

*Section polie
à trois phases*



*Partition
correspondante*

Définition : la covariance rectangle est la probabilité que x soit dans la phase X_i et $x+h$ dans X_j . On note :

$$C_{ij}(\mathbf{h}, \alpha) = \text{Prob} \{ \mathbf{x} \in X_i, \mathbf{x}+\mathbf{h} \in X_j \}.$$

Propriétés de la covariance rectangle

- Lorsque les phases X et X^c sont **complémentaires**

$$C_{10}(\mathbf{h}, \alpha) = A_A(X) - C_{11}(\mathbf{h}, \alpha)$$

où $C_{11}(\mathbf{h}, \alpha)$ est la covariance de la phase X_1

- Pour une structure **multiphasée**

$$A_A(X_i) = \sum C_{ij}(\mathbf{h}, \alpha) \quad \text{et} \quad i \neq j \Rightarrow C_{ij}(\mathbf{0}, \alpha) = 0$$

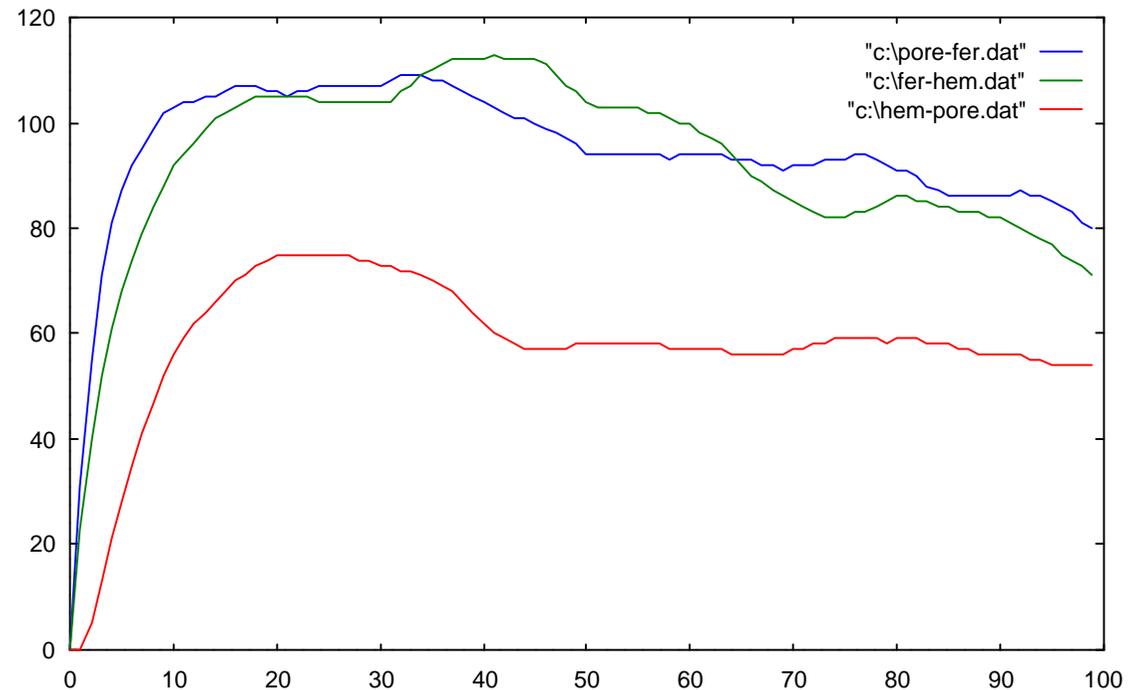
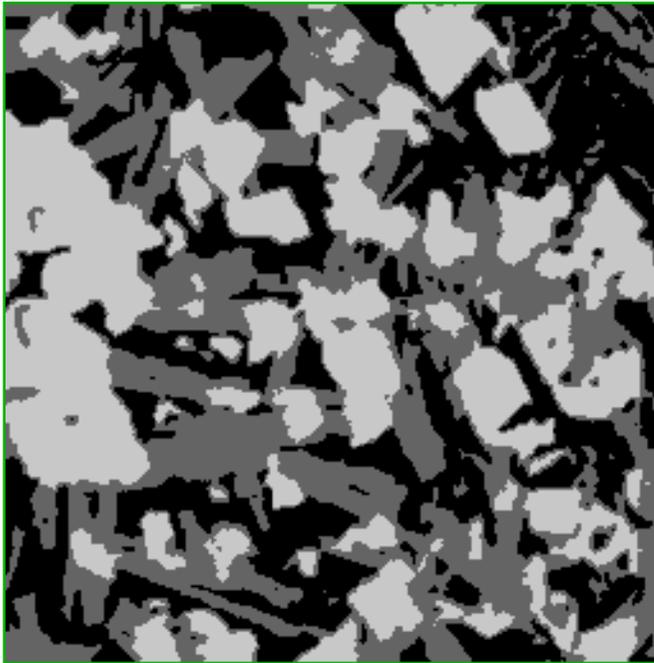
- Pour **h grand**, les événements $x \in X_i$ et $x+h \in X_j$ sont indépendants, d'où

$$\lim C_{ij}(\mathbf{h}, \alpha) = [A_A(X_i)] [A_A(X_j)]$$

- La moyenne directionnelle des dérivées à l'origine est liée au **périmètre (surface) spécifique** entre les phases i et j :

$$-\int C_{ij}'(\mathbf{0}, \alpha) d\alpha = (2/\pi) U_L(X_i / X_j) = S_V(X_i / X_j) / 4$$

Exemple métallographique



- Echantillon d'aggloméré de minerai de fer.
- Gris clair: *hématite* ;
gris sombre : *ferrite* ;
noir : *pores*

Les deux covariances de la ferrite sont semblables. Celle entre pores et hématite

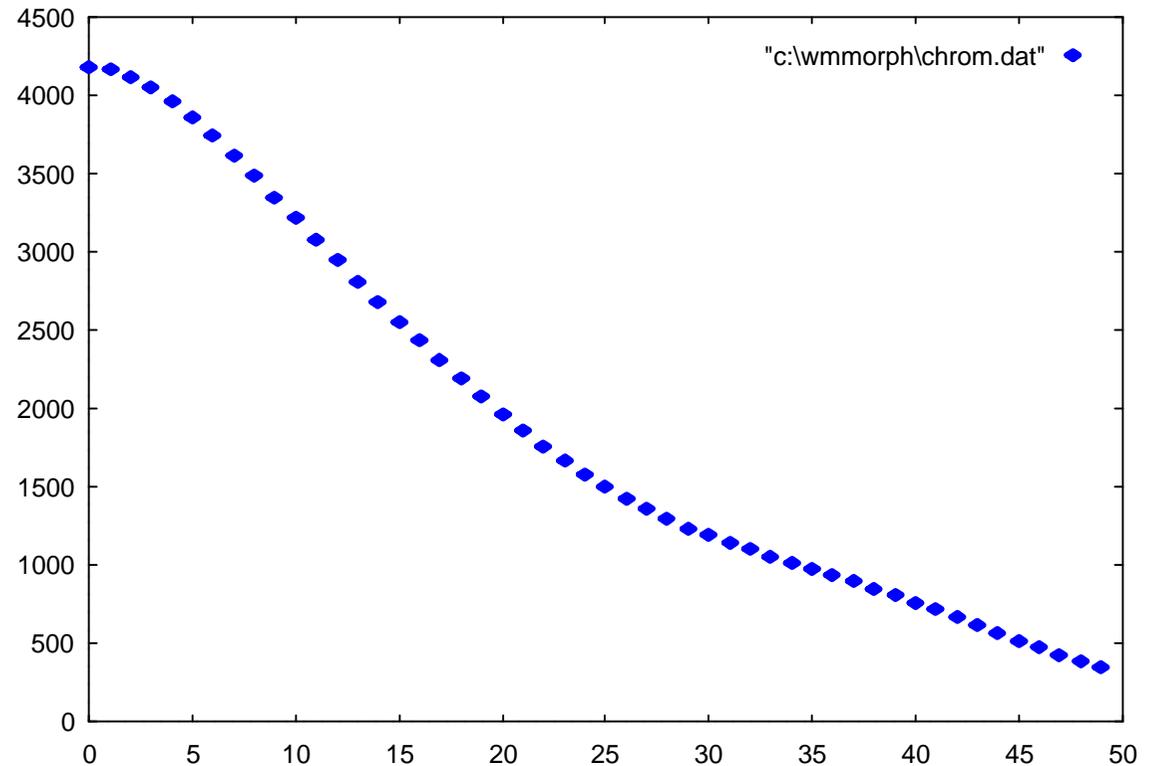
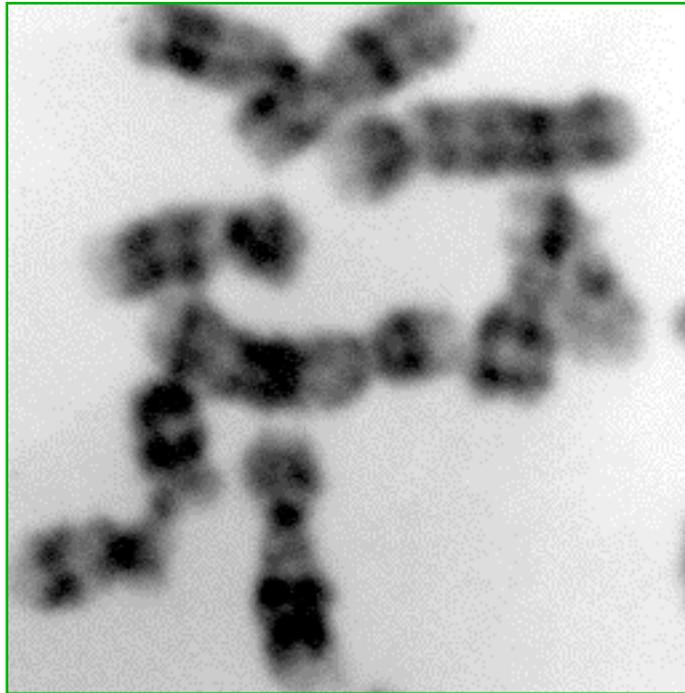
- démarre plus bas : *faible surface de contact*
- présente un net effet de trou à $h = 50$: *halo de ferrite autour de l'hématite*

Passage des ensembles aux fonctions

S'agissant de fonctions aléatoires, le passage **binaire** → **numérique** se manifeste essentiellement :

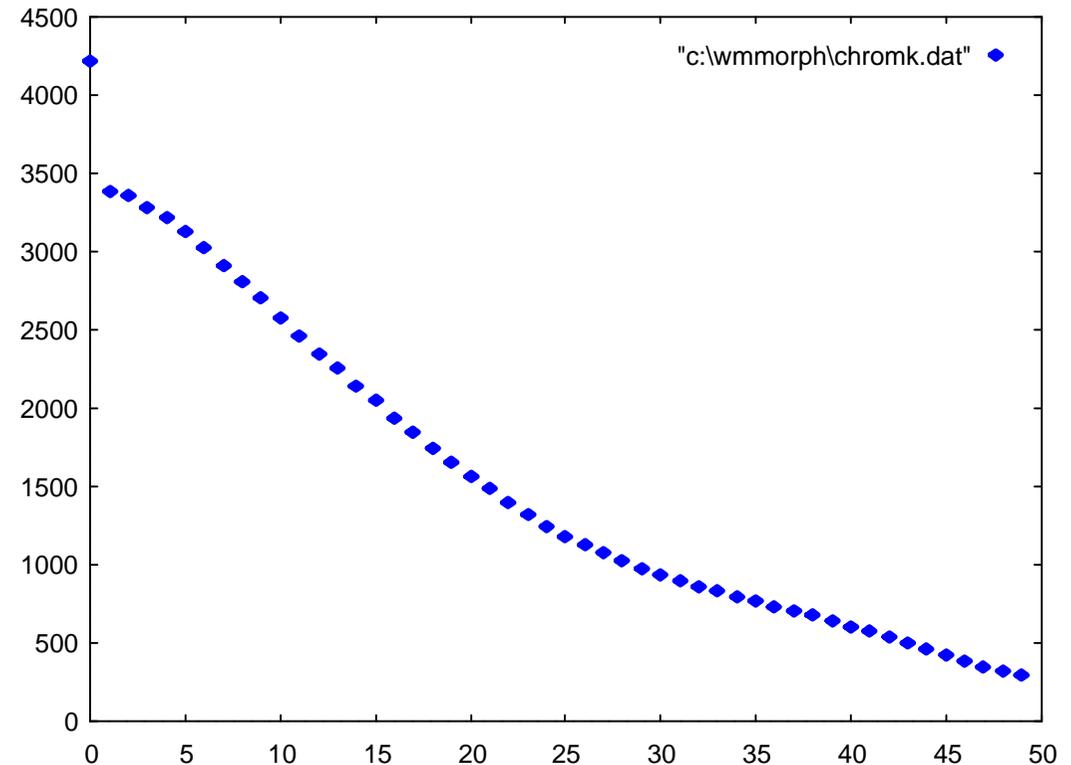
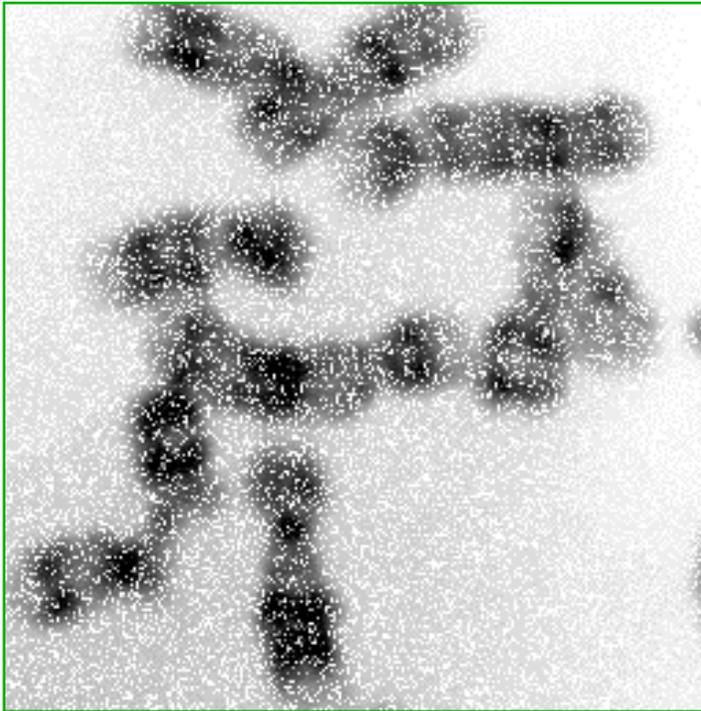
- *sur le comportement au voisinage de l'origine* :
 - les fonctions peuvent être dérivables en *m.q.*,
 - mais pas les indicatrices ensemblistes;
- *sur le comportement à l'infini*:
 - aussi grand soit le champ d'étude, la variance d'un ensemble aléatoire reste bornée par
$$\sup \{ p(1-p), p \in [0,1] \} = 0,25$$
 - A l'inverse, certains phénomènes physiques peuvent avoir une capacité de **dispersion quasi infinie**, *i.e.* voir leur variance augmenter sans limite avec la taille du domaine d'expérimentation (minéralisations, pluie...)

Comportements près de l'origine (I)



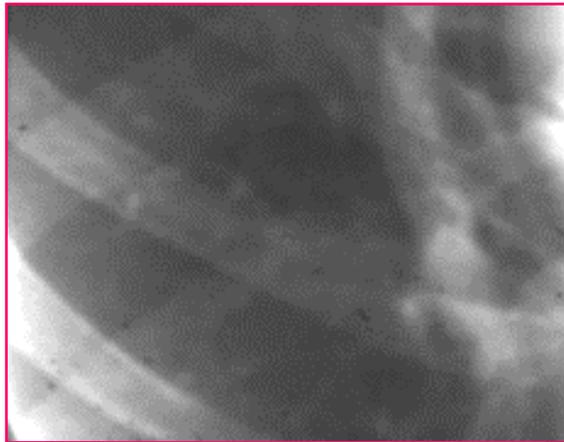
- La covariance d'une fonction **dérivable en moyenne quadratique** est dérivable à l'origine, et par symétrie sa dérivée est nulle.
- A noter que l'hypothèse de stationnarité, manifestement en défaut ici, n'est pas gênante pour l'étude au voisinage de l'origine.

Comportements près de l'origine (II)

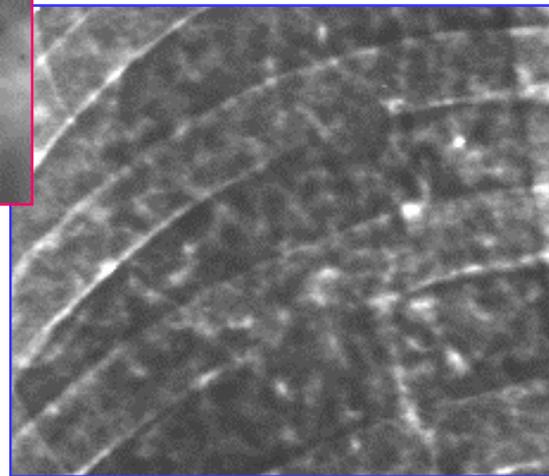


- La présence d'un bruit poissonnien se traduit par un **décrochement** de la covariance à l'origine, dit «**effet de pépite**» dont la valeur égale la variance du bruit.
- En revanche, s'il y a **indépendance** entre le bruit et le phénomène, le reste de la covariance n'est pas affecté: robustesse très utile en pratique.

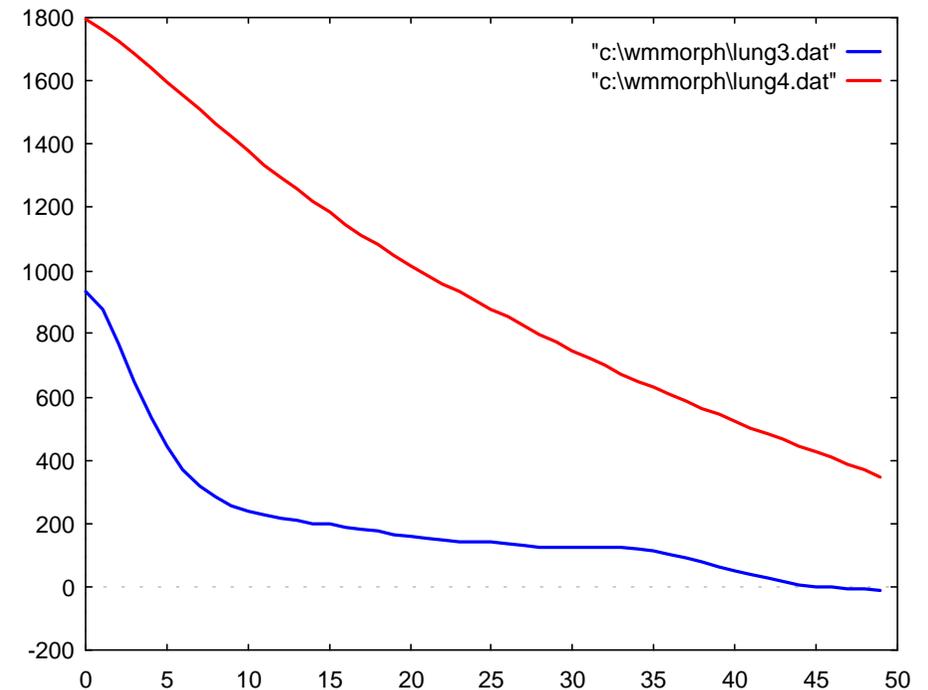
Comportements près de l'origine (III)



a) poumon sain



b) poumon de fumeur



- Les covariances des fonctions **continues en moyenne quadratique** (e.g. indicatrices) ne sont pas dérivables à l'origine (tangente oblique).
- De plus, la covariance **somme les structures** indépendantes (e.g. l'effet de pépite). Dans la radio de poumon du fumeur apparaît la superposition de la structure des côtes (8μ), et des alvéoles pulmonaires sombres (35μ).

Fonctions aléatoires de dilution

- Quel est le lien entre une forme de base (Ex. le nodule pulmonaire) et la structure stationnaire qu'il engendre ? Peut-on construire une fonction aléatoire stationnaire de même covariance qu'un covariogramme donné? Comment justifier les superpositions d'échelles sur les covariances ?
- Pour répondre à ces questions, le modèle le plus simple (mais pas le seul) consiste à
 - partir de **points poissonniens** $\{x_i, i \in I\}$ dans \mathbb{R}^n , de variance λ
 - implanter indépendamment en chaque point x_i , une réalisation f_i d'un **objet primaire** de covariogramme $K_\alpha(h)$,
 - et faire la **somme** $f = \sum f_i$ de toutes ces objets primaires.
- La covariance de la fonction aléatoire f , stationnaire, vaut alors

$$C(h, \alpha) = \lambda K_\alpha(h)$$

et celle de la somme $f = f' + f''$ de deux fonctions indépendantes de ce type

$$\lambda' K'_\alpha(h) + \lambda'' K''_\alpha(h).$$

Covariance et analyse spectrale

Passons au comportement à l'infini, en commençant par un théorème célèbre.

Théorème (Wiener-Khinchine) : La covariance $C(\mathbf{h})$ d'une fonction aléatoire stationnaire f est la transformée de Fourier de son spectre d'énergie $\phi(\mathbf{v})$, donc d'une mesure ≥ 0 .

Corollaire 1: Cette dernière condition équivaut à dire que $C(\mathbf{h})$ est de type **défini positif**, i.e. pour tous poids λ_i et tous points \mathbf{x}_i , l'on a

$$\sum \lambda_i \lambda_j C(\mathbf{x}_i - \mathbf{x}_j) \geq 0.$$

Corollaire 2 : Comme la transformation de Fourier échange les propriétés à l'origine et à l'infini, il vient en particulier :

$$\int \phi(\mathbf{v}) d\mathbf{v} = C(\mathbf{0}) = \mathbf{E} [(f(\mathbf{x}) - m)]^2 = \sigma_0^2$$

La somme des énergies associées à toutes les fréquences est donc égale à la **variance ponctuelle** σ_0^2

Variances et portée

- Inversement, la **portée a** est associée à la fréquence nulle :

$$a \ C(0) = \int C(h) \ dh = \phi(0) .$$

- Soit alors Z un champ de grande taille, et $f_Z(x) = \int_Z f(x-y)dy$, il vient

$$Z \rightarrow \infty \quad \Rightarrow \quad \sigma^2_Z = E [(f_Z(x) - m)]^2 \rightarrow [a / \text{Mes } Z] C(0)$$

Vis à vis de la variance σ^2_Z d'un échantillon moyen Z , la portée apparaît donc asymptotiquement comme **l'unité de taille** du phénomène.

- Plus généralement, que Z soit grand ou pas, si $\sigma^2_{0/Z}$ désigne la variance ponctuelle dans Z , on montre que

$$\sigma^2_{0/Z} = \sigma^2_0 - \sigma^2_Z ;$$

Cette identité remarquable, dite **formule de Krige**, constitue le point de départ de la **Géostatistique**.

Théorie intrinsèque, variogramme

- Lorsque le phénomène étudié est très étendu à travers l'espace, il arrive que, quand Z augmente, la variance expérimentale $[\sigma^2_{0/z}]^*$ d'un point dans Z croisse indéfiniment avec Z , et cela à toutes nos échelles d'analyse.
- D'après la formule de Krige, cela signifie que le modèle stationnaire est inadapté: $\sigma^2_0 = C(0)$ étant infini, la covariance n'existe pas.
- **Modèle intrinsèque** (*G.Matheron*) : Mais il se peut que les *accroissements* de f existent et aient un sens. Les supposant stationnaires, on étudie alors leur variance, ou variogramme

$$\gamma(\mathbf{h}, \alpha) = E [f(\mathbf{x}) - f(\mathbf{x}+\mathbf{h})]^2$$

- **Propriétés** : On a $\gamma(0, \alpha) = 0$, et

$$\sigma^2_0 < \infty, \Rightarrow \gamma(\mathbf{h}, \alpha) = 2[C(0) - C(\mathbf{h})]$$

Quand $\sigma^2_0 = \infty$, la notion de portée s'estompe, et de nouveaux comportements à l'infini apparaissent.

Exemple : marches poissonniennes (I)

- On objectera : mais de toutes façons, toute donnée numérique est **finie** ! Voyons cela de plus près, dans le cas d'un modèle qui se calcule bien.
- **Définition** : On engendre des **marches poissonniennes** en implantant
 - un processus de points de Poisson dans \mathbb{R}^1 , de variance λ ,
 - et un saut d'amplitude δ en chaque point de Poisson. La variable δ suit une loi de moyenne nulle et de variance σ^2 .
- **Propriétés** : Dans ce modèle, seuls les accroissements sont définis; ils sont de moyenne nulle, et de variogramme

$$\gamma(\mathbf{h}) = \lambda \sigma^2 |\mathbf{h}|$$

Exemple : marches poissonniennes (II)

le **variogramme** expérimental $2\gamma^*(h) = 1/(L-h) \int_0^{L-h} [f(x+h) - f(x)]^2 dx$

a pour espérance $E[2\gamma^*(h)] = \gamma(h) = \lambda \sigma^2 |h|$.

- Mais on peut toujours poser $f_L = 1/L \int_L f(x) dx$, puis

$$\text{Cov}^*(h) = 1/(L-h) \int_0^{L-h} (f(x+h) - f_L)(f(x) - f_L) dx.$$

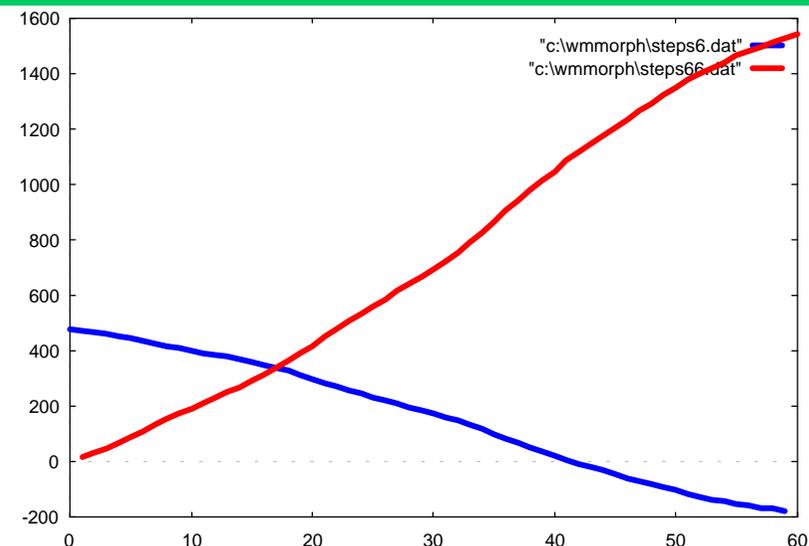
- Cette «**covariance**» expérimentale admet pour espérance

$$E[\text{Cov}^*(h)] = L/3 - 4/3 \cdot h + 2/3 \cdot h^2 / L \quad (0 \leq h \leq L);$$

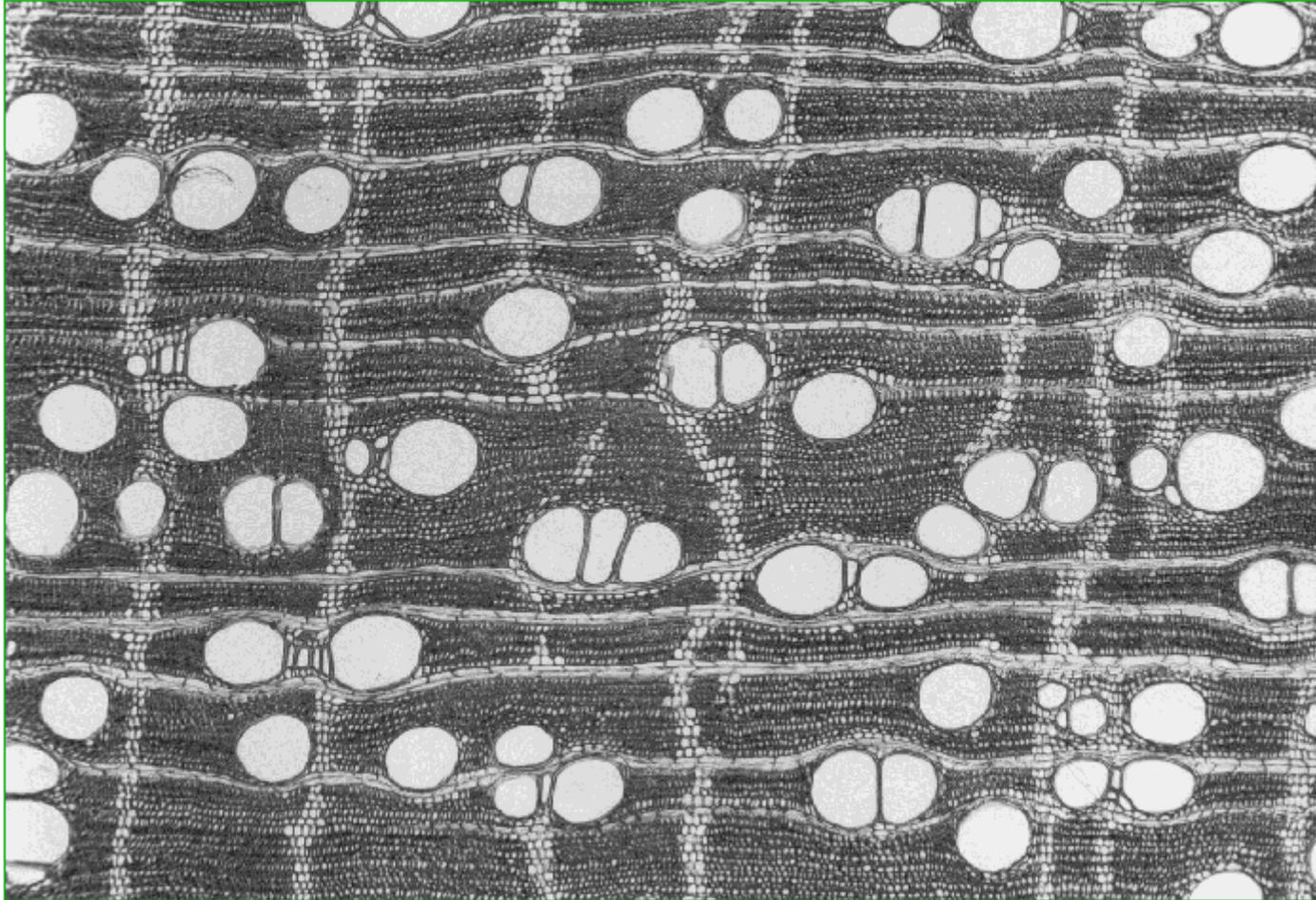
La vraie structure du phénomène est complètement faussée.



*Simulation de marches,
sa « covariance », son variogramme*



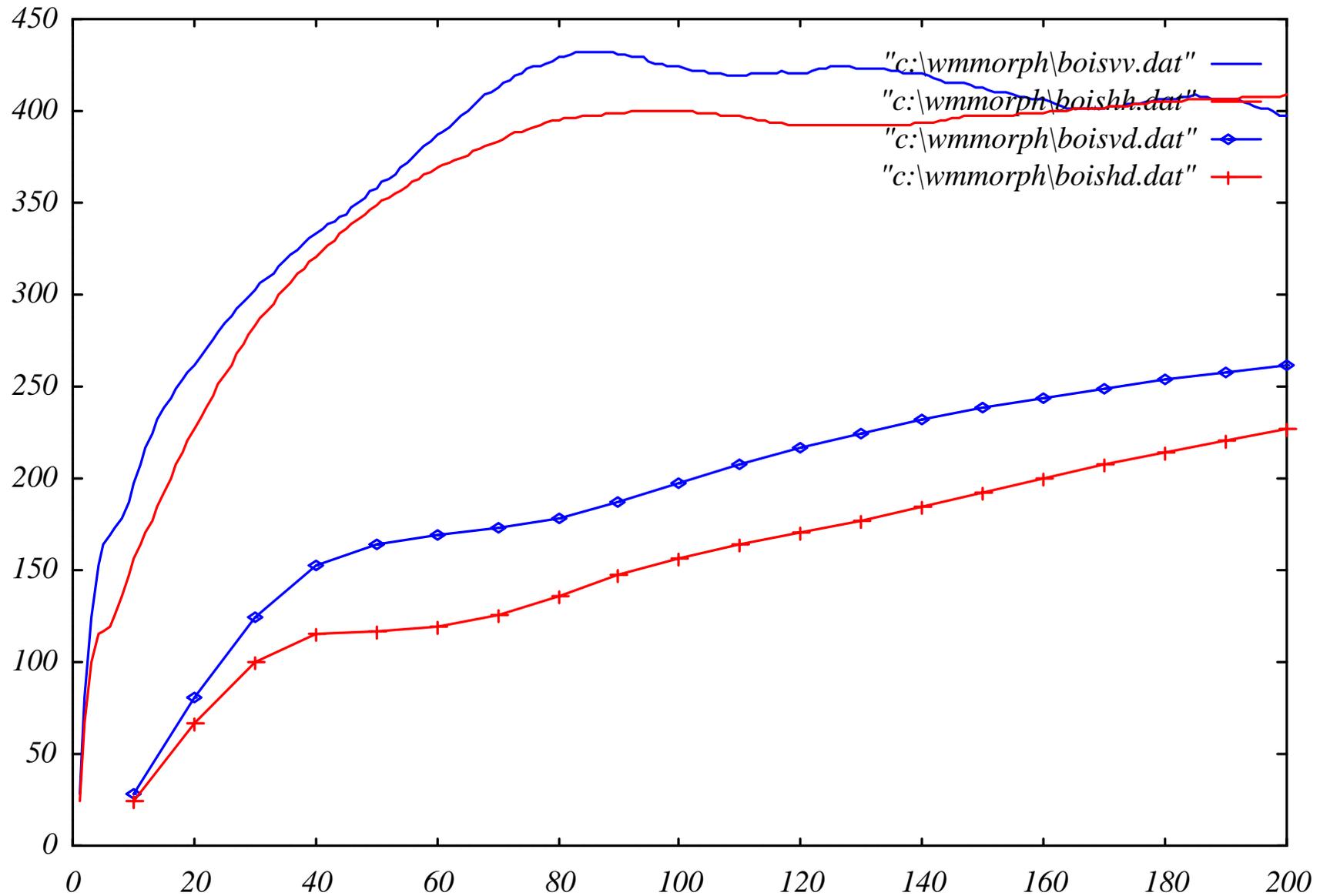
Clap de fin (I)



Lame mince de hêtre, perpendiculaire à l'axe du tronc. La direction horizontale est celle des rayons partant du centre de l'arbre; la verticale celle des anneaux annuels.

Dimensions : 1506×1016 pixels, avec 1 pixel = 1 micron.

Clap de fin (II)



Variogrammes vertical (bleu) et horizontal (rouge) des pas 1 à 200 , accompagnés de leurs débuts, avec une échelle dix fois plus large en x ... Interprétez