

# Introduction à la géostatistique et variogrammes

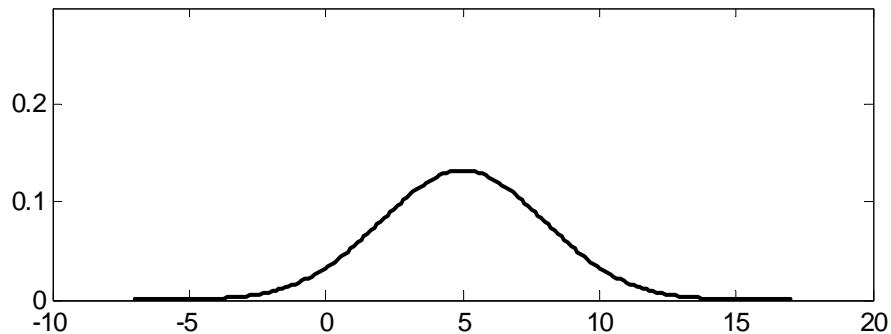
Automne 2003

# Plan

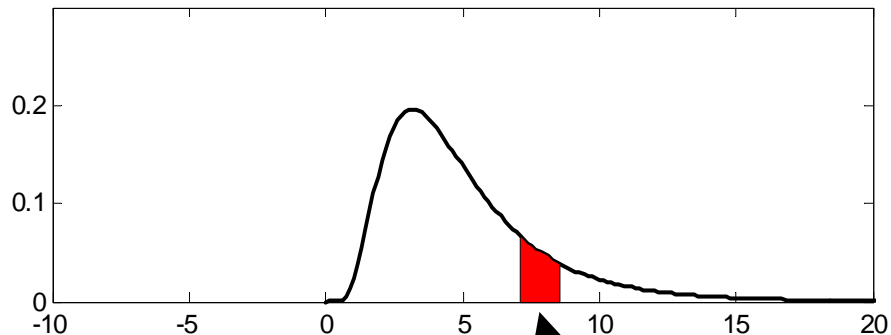
- Rappels statistiques
  - 1 v.a.
  - 2 v.a.
- Point de vue géostat
  - Gisement vs modèle stat
- Historique
- Effet support
- Effet information
- Géostatistique linéaire
  - Hypothèse de stationnarité
  - Variogramme expérimental
  - Modèles
  - Problèmes et stratégie de modélisation

Une v.a. (continue) est entièrement caractérisée par sa fonction de densité

Loi normale,  $m=5$ ,  $\sigma=3$



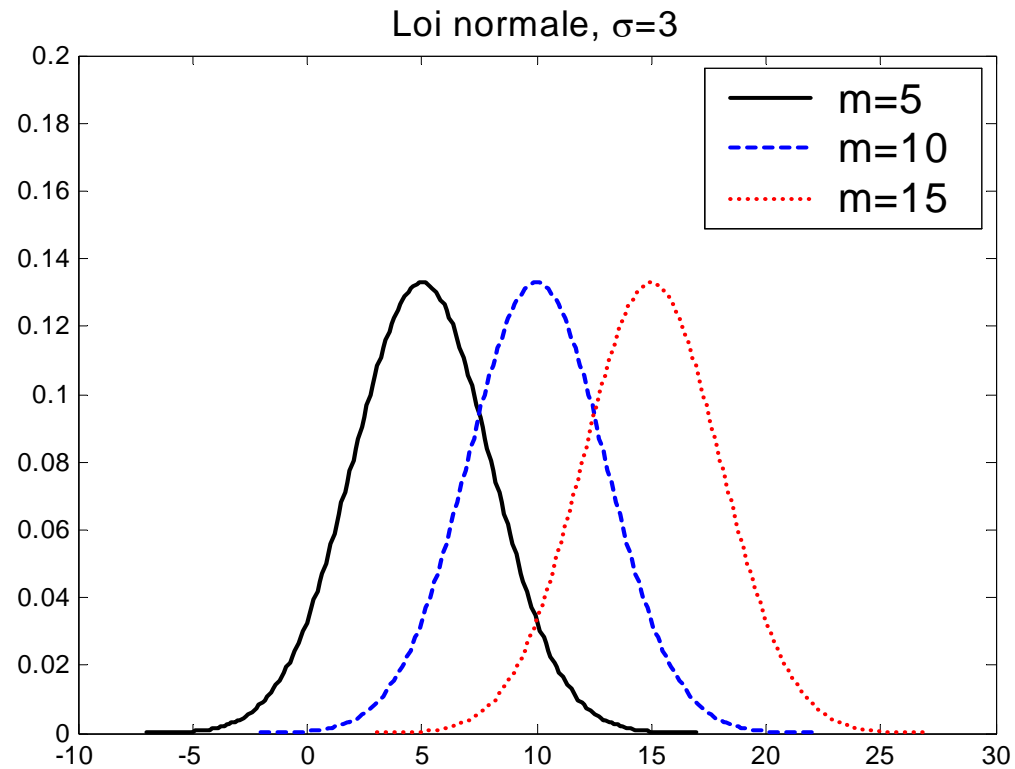
Loi lognormale,  $m=5$ ,  $\sigma=3$



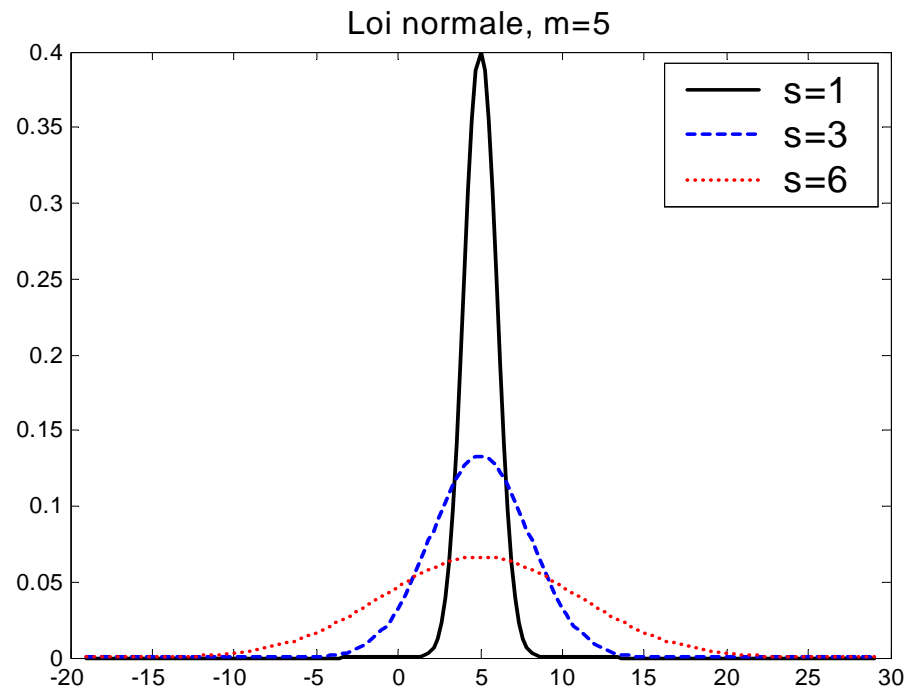
Intégrale: probabilité

## Résumer une distribution par certaines statistiques

Tendance centrale (moyenne)

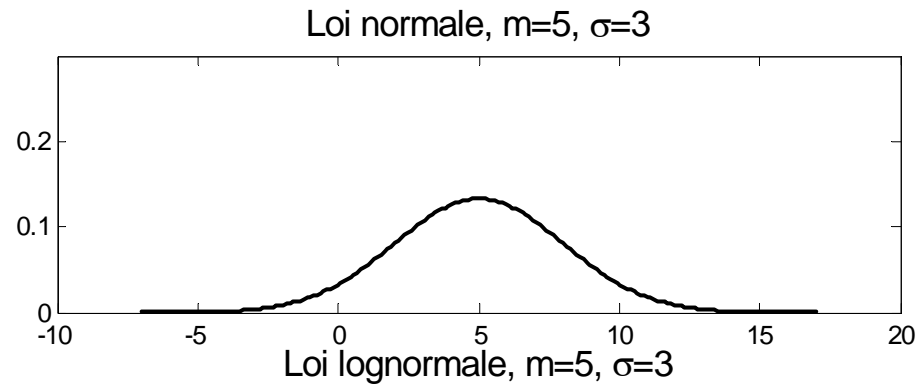


## Dispersion autour de la moyenne (écart-type)

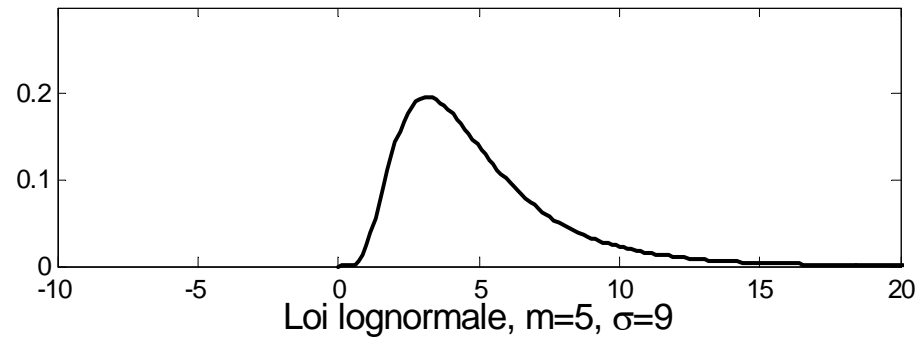


# Asymétrie

= 0



= 0.44



= 3.84

Le type de loi peut avoir une grande influence sur les ressources et la valeur d'un gisement.

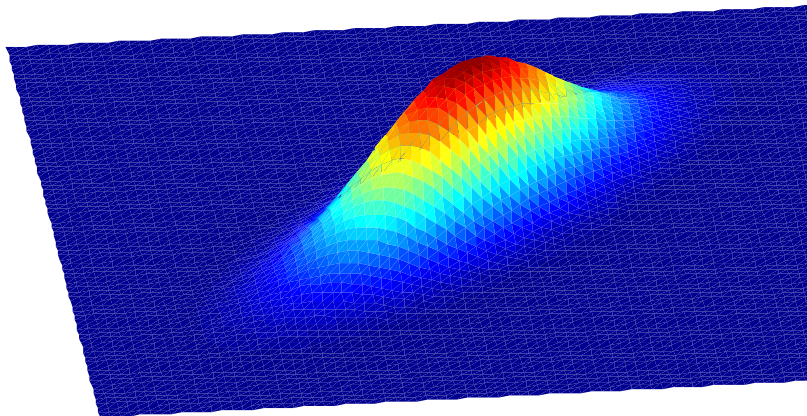
Exemple : gisement  $m=2$ ,  $\sigma=2$ , t.coupure = 1

	Normale	Lognormale
Tonnage/T0	0.69	0.66
Teneur	3.02	2.70
Profit conv	1.40	1.13

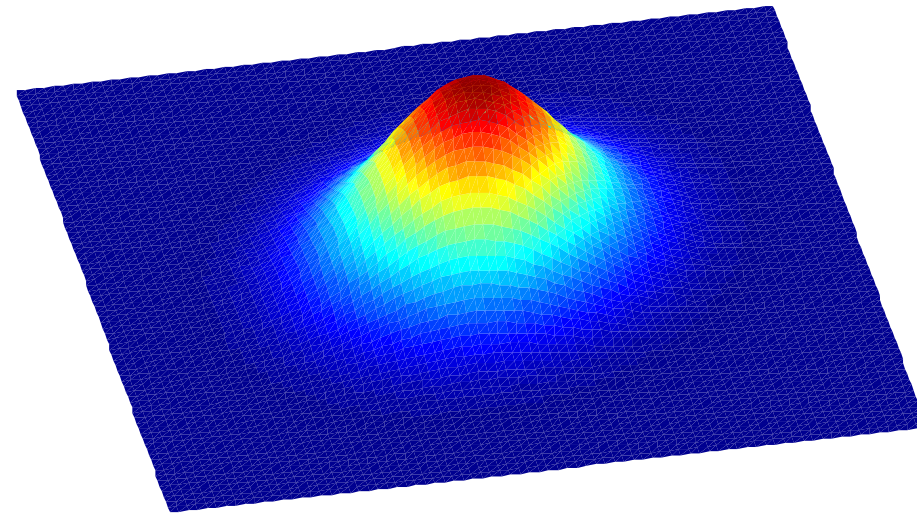
# Deux v.a.

- Un couple de v.a.  $X$  et  $Y$  (continues) est entièrement caractérisé par la loi de densité conjointe  $f(x,y)$

Loi binormale,  $\sigma_1=2$   $\sigma_2=5$   $\rho=0.8$



Loi binormale,  $\sigma_1=2$   $\sigma_2=5$   $\rho=0.2$



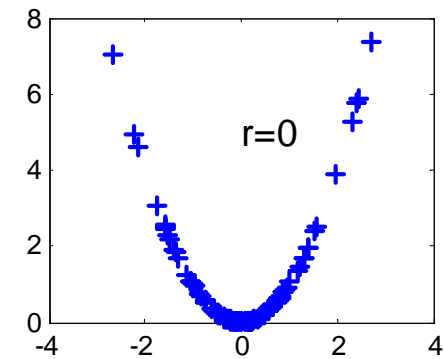
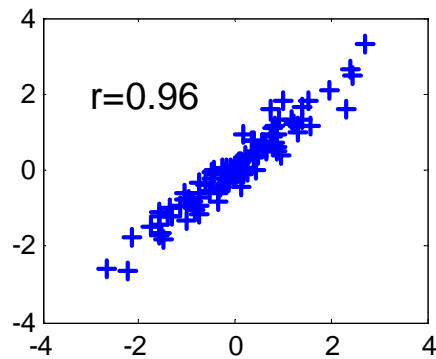
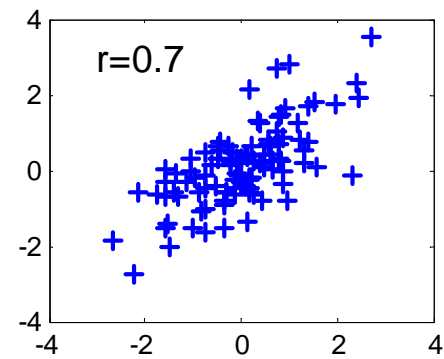
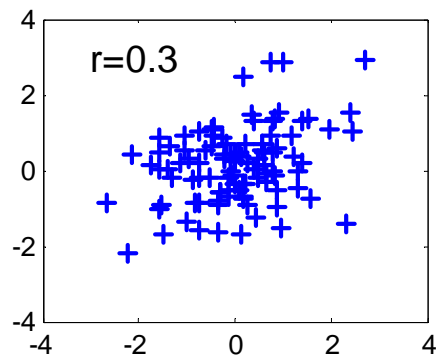
On peut résumer une distribution bivariable par différentes statistiques dont :

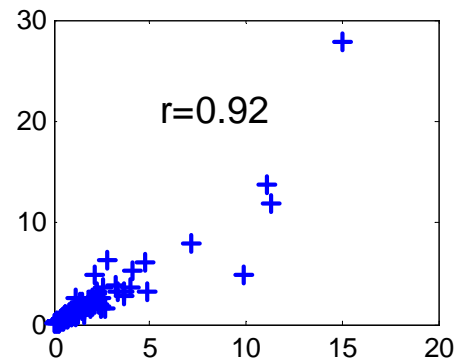
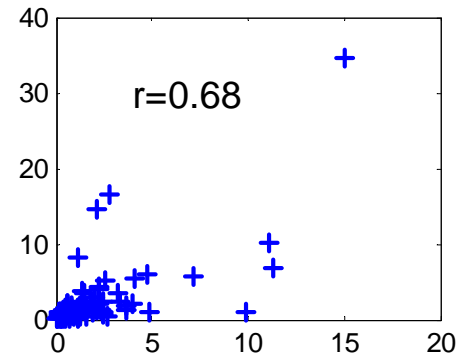
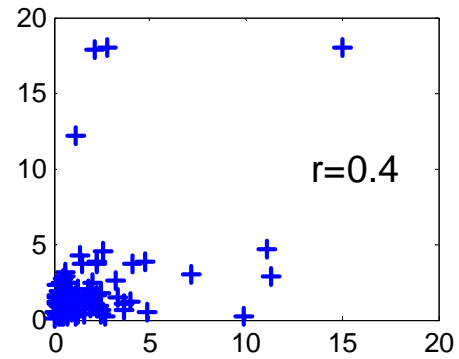
- moyennes des 2 variables
- écart-types (ou variances) des 2 variables
- corrélation (ou covariance) entre les 2 variables

La covariance mesure le degré d'association entre 2 v.a.

La corrélation est la covariance entre 2 v.a. normalisées pour présenter un écart-type de 1

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$





# Espérance mathématique

- Notion fondamentale

$$E[X] = \int x f_X(x) dx$$

$$E[g(X)] = \int g(x) f_X(x) dx$$

$$E[g(X, Y)] = \int \int g(x, y) f_{X,Y}(x, y) dx dy$$

- Si  $g(X) = (X - m)^2 \Rightarrow E[g(X)] = \text{Var}(X)$

- Si  $g(X, Y) = (X - m_x)(Y - m_y) = \text{Cov}(X, Y)$

# Propriétés de l'espérance mathématique

- E est un opérateur linéaire =>

$$E[c g(X)] = c E[g(X)]$$

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)]$$

$$E[g(X,Y) + h(X,Y)] = E[g(X,Y)] + E[h(X,Y)]$$

En particulier

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2 * \text{Cov}(X_1, X_2)$$

$$\text{Var}(aX_1 + bX_2) = a^2 \text{Var}(X_1) + b^2 \text{Var}(X_2) + 2ab * \text{Cov}(X_1, X_2)$$

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j)$$



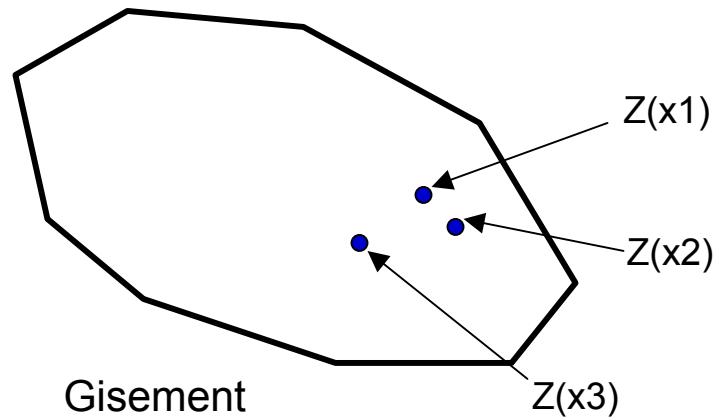
Une des expressions  
qui revient le plus  
souvent en géostat

$$\text{Var}\left(c \int z(x) dx\right) = c^2 \iint \text{Cov}(z(x), z(y)) dx dy$$

← Une autre expression  
qui revient souvent en  
géostat

$$\text{Cov}\left(a \int z(x) dx, b \int z(y) dy\right) = ab \iint \text{Cov}(z(x), z(y)) dx dy$$

# Point de vue de la géostat



Gisement  $\Rightarrow$  infinité de points ou très grand nombre de « quasi-points »

$x$  : emplacement géographique

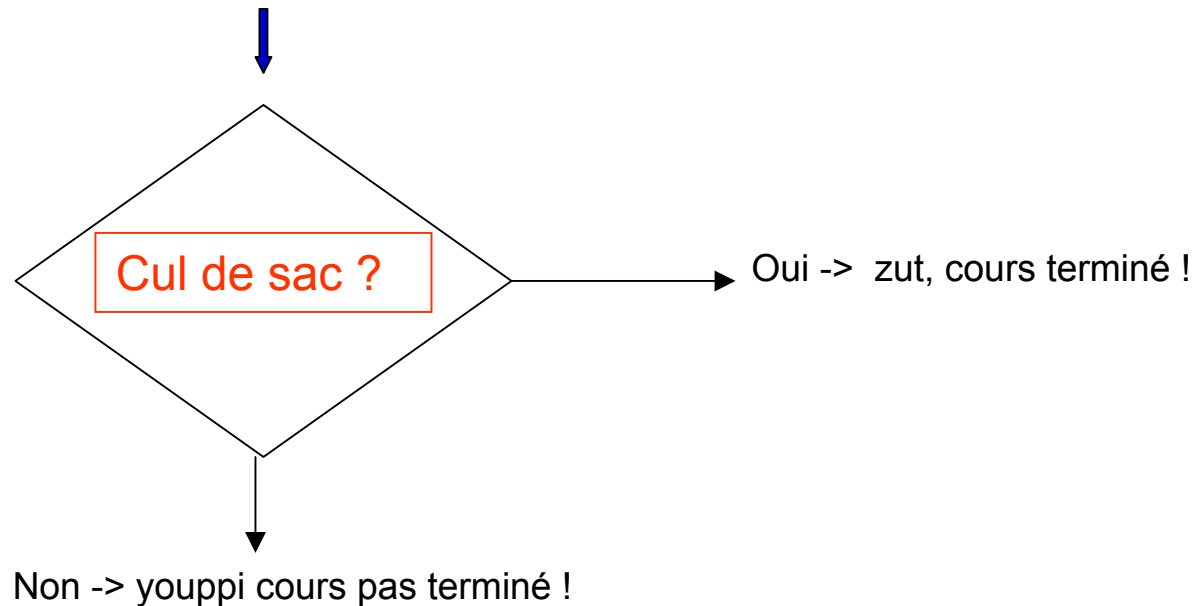
chaque point  $\rightarrow$  teneur  $\rightarrow Z(x)$

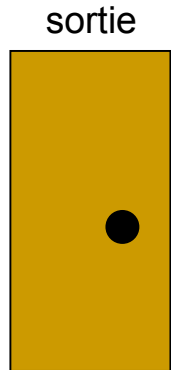
chaque teneur  $\rightarrow$  v. a. (ensemble forme une fonction aléatoire (de  $x$ ))

Impossible d'estimer à partir des données la loi de densité conjointe

Impossible d'estimer à partir des données la loi de densité bivariable

Impossible d'estimer à partir des données la loi de densité d'une variable





-Hypothèses

-Questions auxquelles le modèle permet de répondre

Hypothèse



Gamme de questions



Généralité



Hypothèse



-Simulations

-Méthodes non-linéaires

-Méthodes linéaires

# Historique de la géostatistique

- 1930-1950 Théorie des fonctions aléatoires (Kolmogorov, Wiener)
- 1955 Daniel Krige : approche empirique (régression) pour corriger problèmes de biais conditionnel observé dans les mines

Pourquoi moins que prévu ?

Comment prévoir et tenir compte de l'effet support ?

- 1960-1970 Matheron (mines), Matérn (foresterie), Gandin (météorologie) développent ensemble d'outils => naissance de la géostat linéaire stationnaire. Réponse aux questions de Krige. Matheron donne le nom de « krigeage » à la méthode d'estimation qu'il développe.
- 1970 Polytechnique est la 1ère Université en A. du N. à enseigner la géostat (M. David)

# Historique (suite)

- 1973 géostat linéaire non-stationnaire \*
- 1975 géostat non-linéaire
- 1977 1er livre en anglais de géostat (M. David)
- 1980 géostat linéaire multivariable \*
- 1985 simulations \*

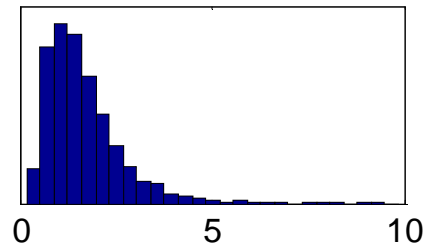
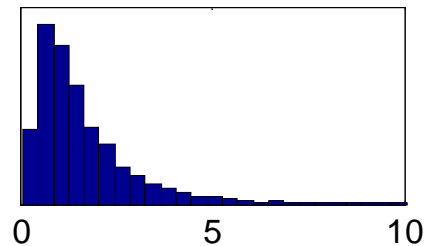
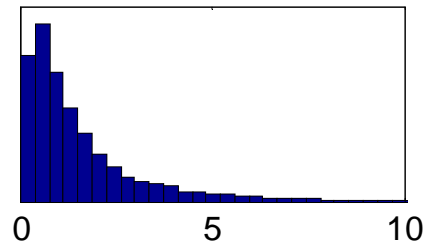
\* *Domaines encore actifs de recherche*

- Aujourd'hui, la géostat est appliquée dans une foule de domaines (mines, pétrole, foresterie, agriculture, environnement, hydrogéologie, géotechnique, pêches, biologie, biomédical,...)

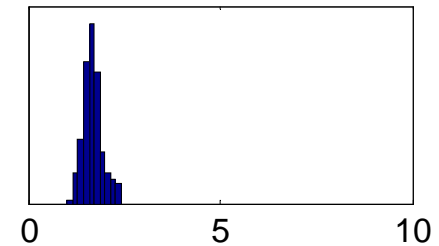
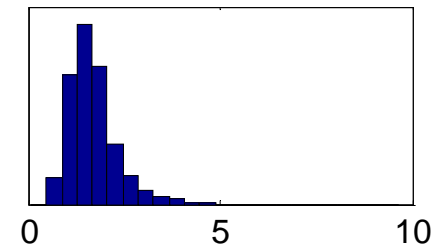
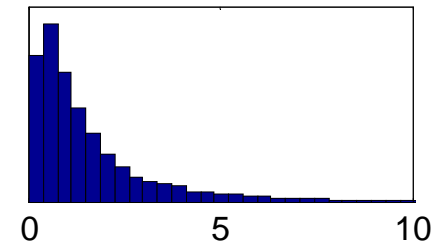
# Effet support



Gisement A

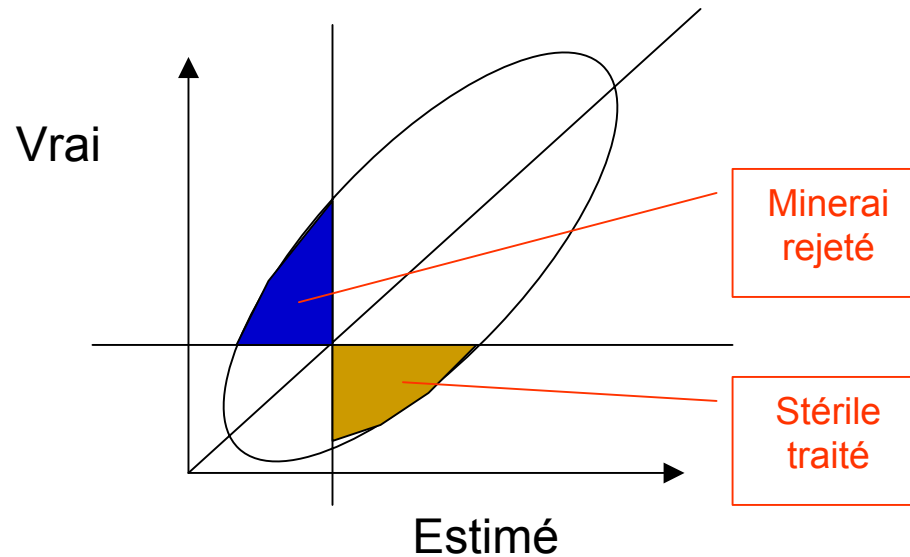


Gisement B



Comment prévoir ces comportements différents ?  
Quel est l'impact \$\$ ?

# Effet information



Prévoir l'étendue des plages d'erreur et pertes en \$ ?  
Évaluer \$ en information pour réduire les pertes ?

On mine à partir d'estimés mais on récolte des valeurs vraies !

### À tonnage extrait égal

⇒ on récupère toujours moins de métal avec des gros blocs qu'avec des petits blocs (effet support)

⇒ on récupère toujours moins de métal avec des estimés qu'avec les vraies valeurs (effet information)

### *La géostatistique permet théoriquement*

- Prévoir l'ampleur de ces effets
- Minimiser ces effets
- Prendre des décisions éclairées au vu de ces effets

# Géostatistique linéaire

## ■ Questions

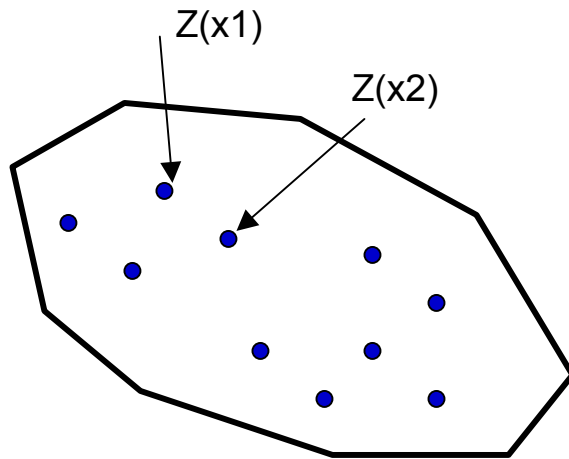
- Estimation de teneurs ponctuelles ou blocs
- Précision associée à ces estimations

## ■ Hypothèse

- Stationnarité du second ordre

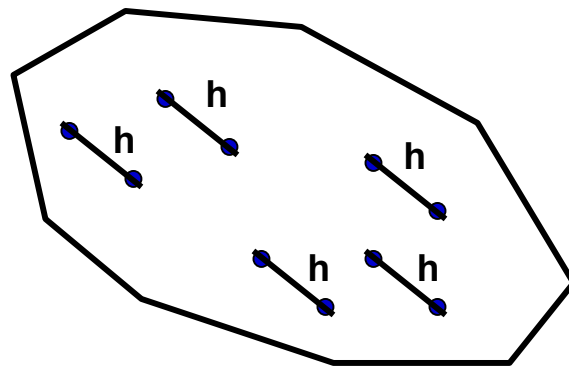
Les caractéristiques sont moyenne, variance et covariance

Deux paires de points espacés d'un même vecteur « h » ont des caractéristiques semblables



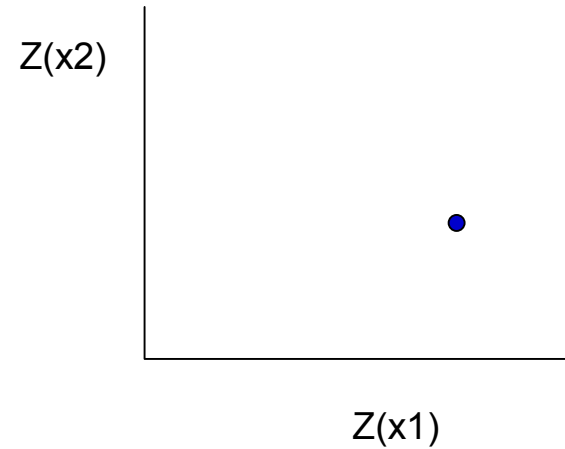
Gisement

+ hypothèse stationnarité



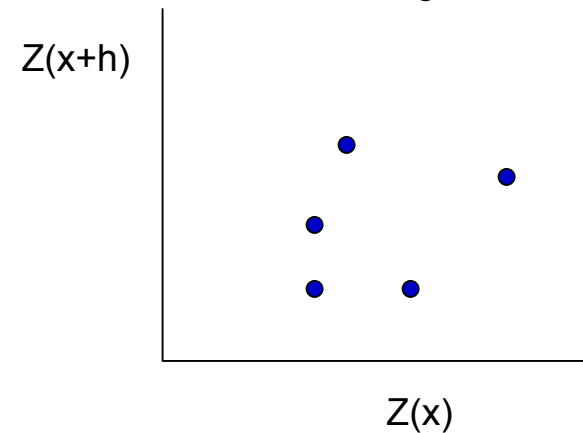
Gisement

Diagramme binaire

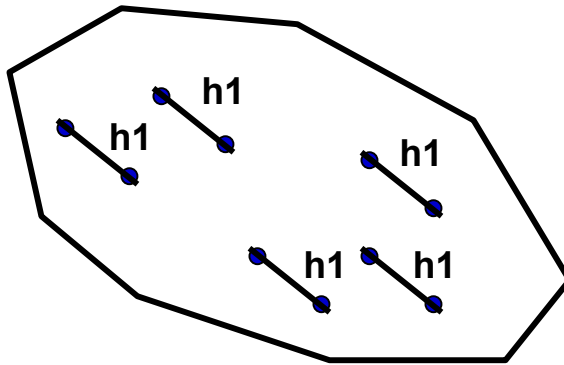


Moyenne ? Variance ? Covariance?

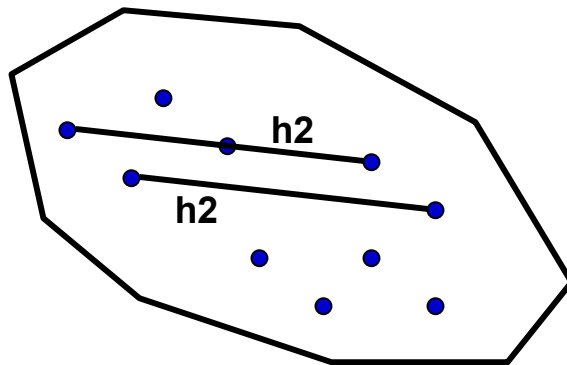
« h scattergram »



Faire varier « h »

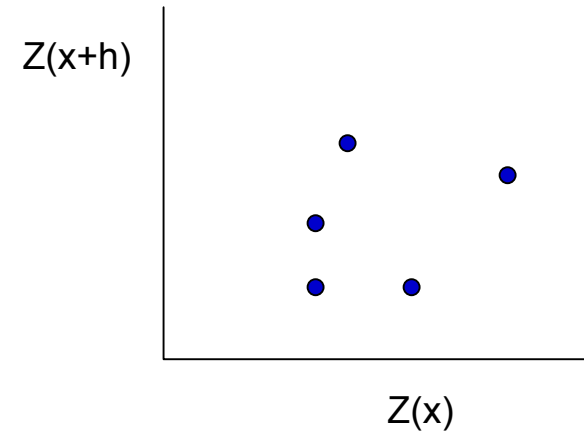


Gisement

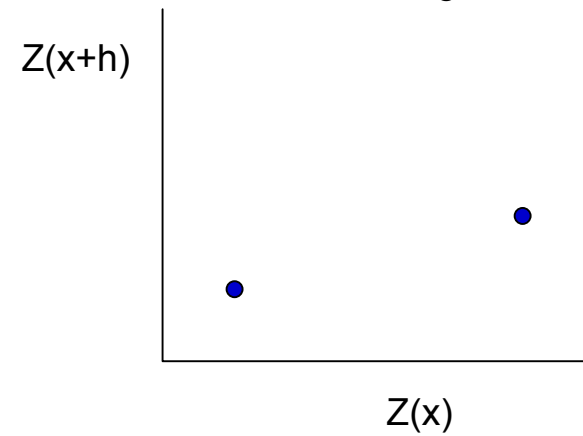


Gisement

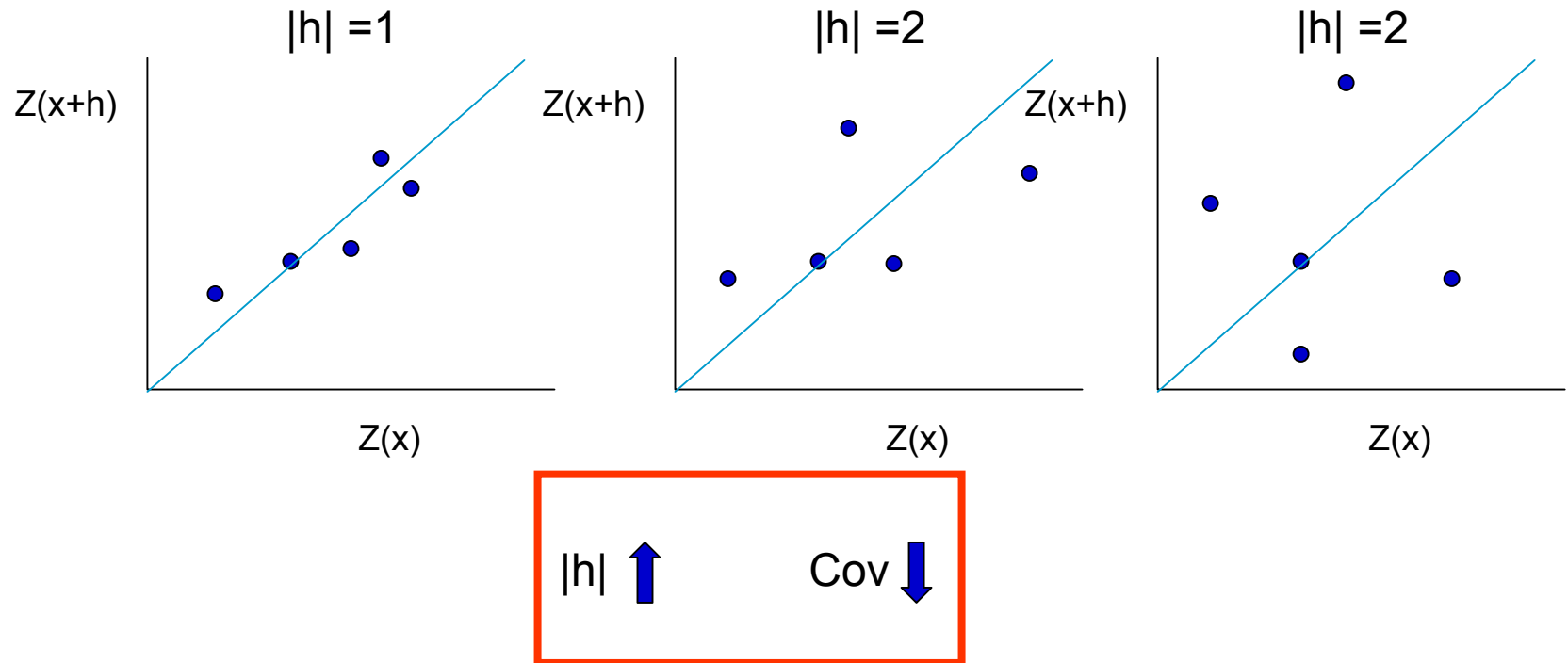
« h1 scattergram »



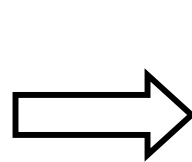
« h2 scattergram »



«  $h$  » peut varier en direction et en module.

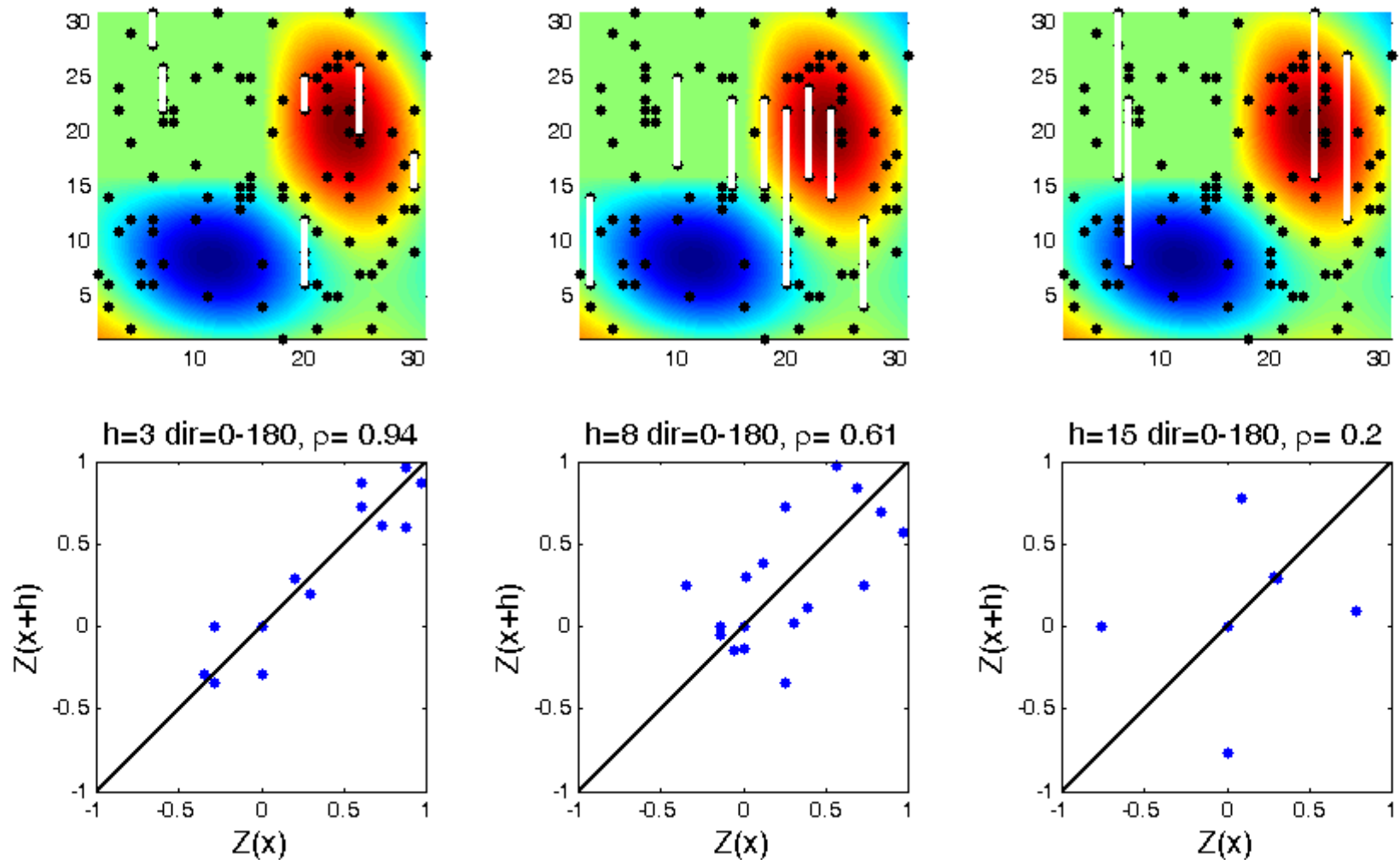


On cherche si possible avoir au moins 30 points sur chaque diagramme

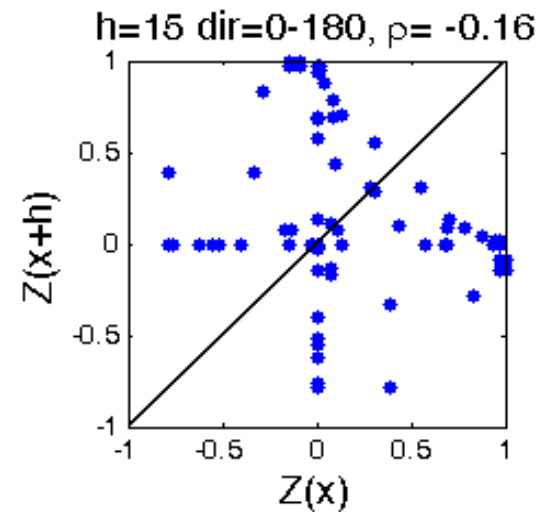
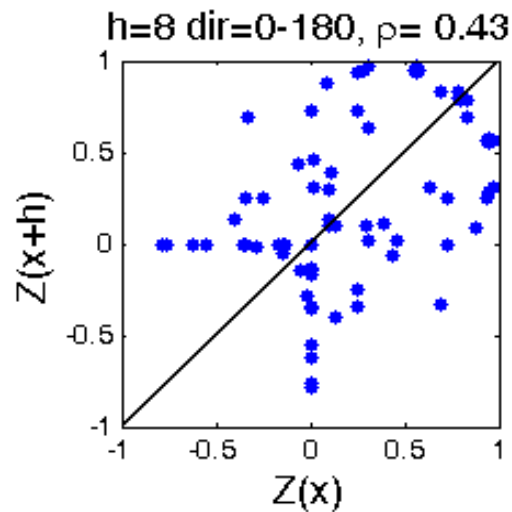
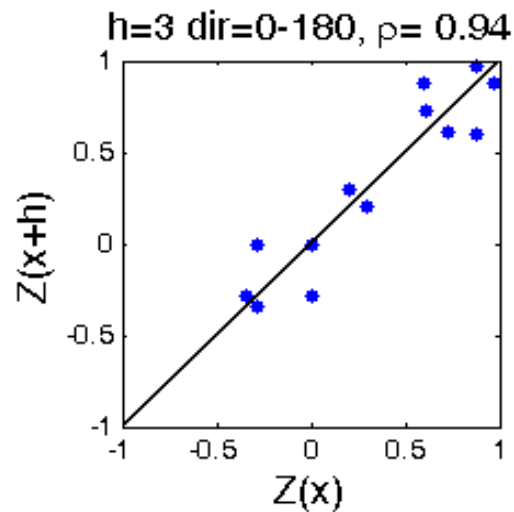
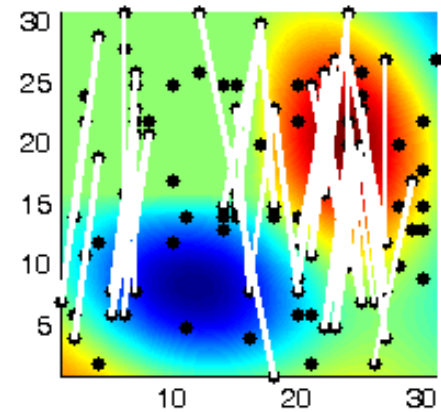
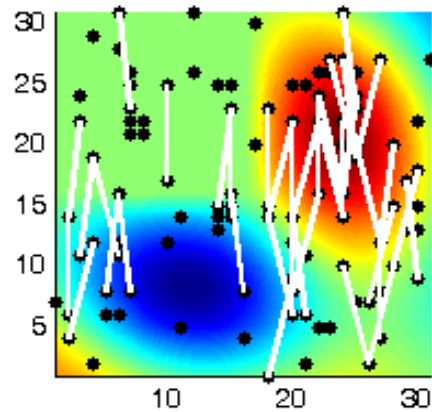
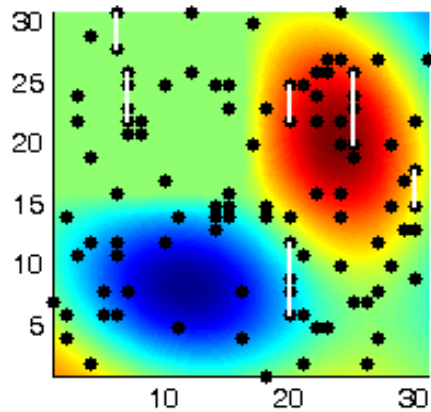


- Tolérance sur la direction  
- Tolérance sur le module

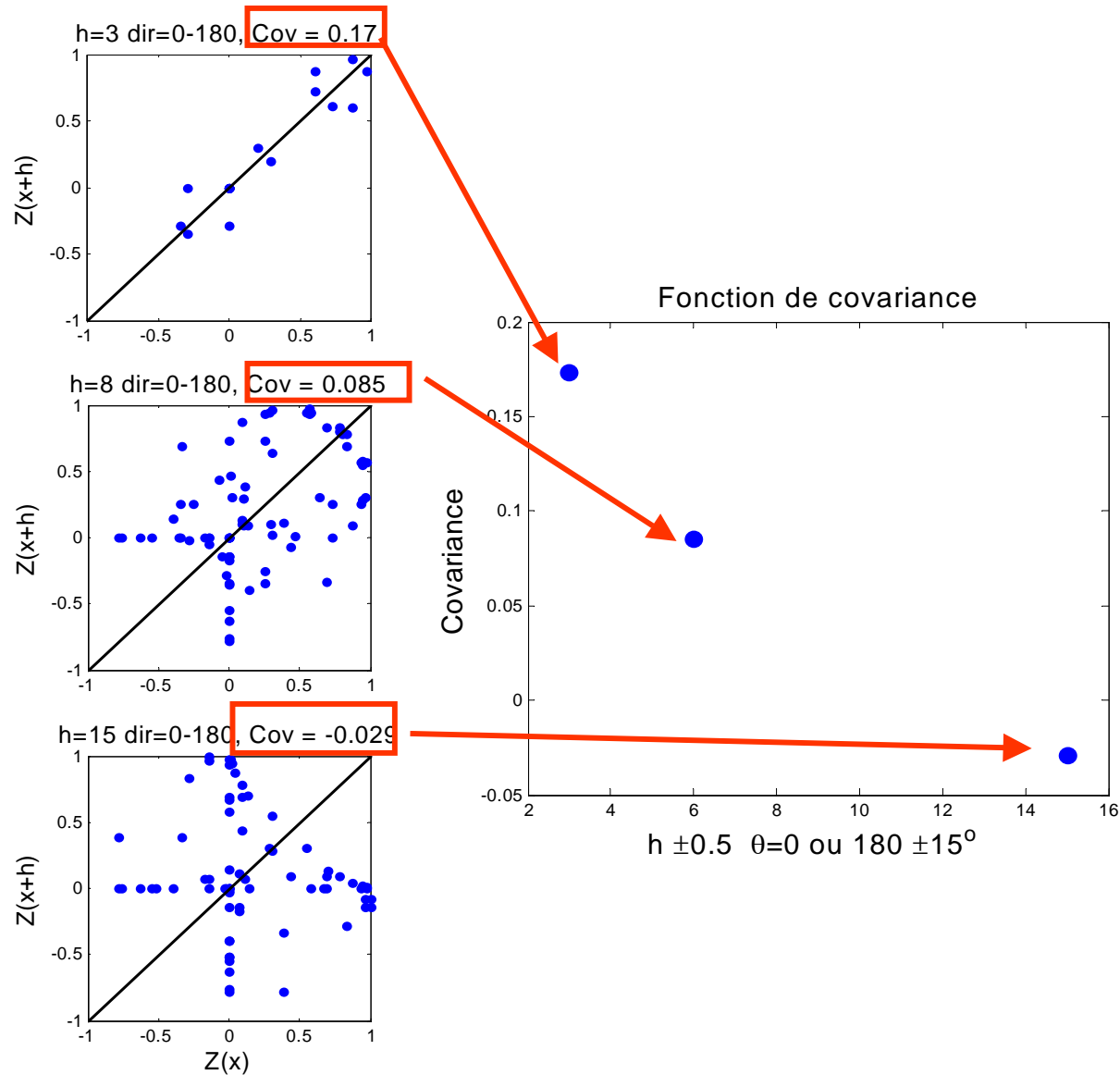
# Example



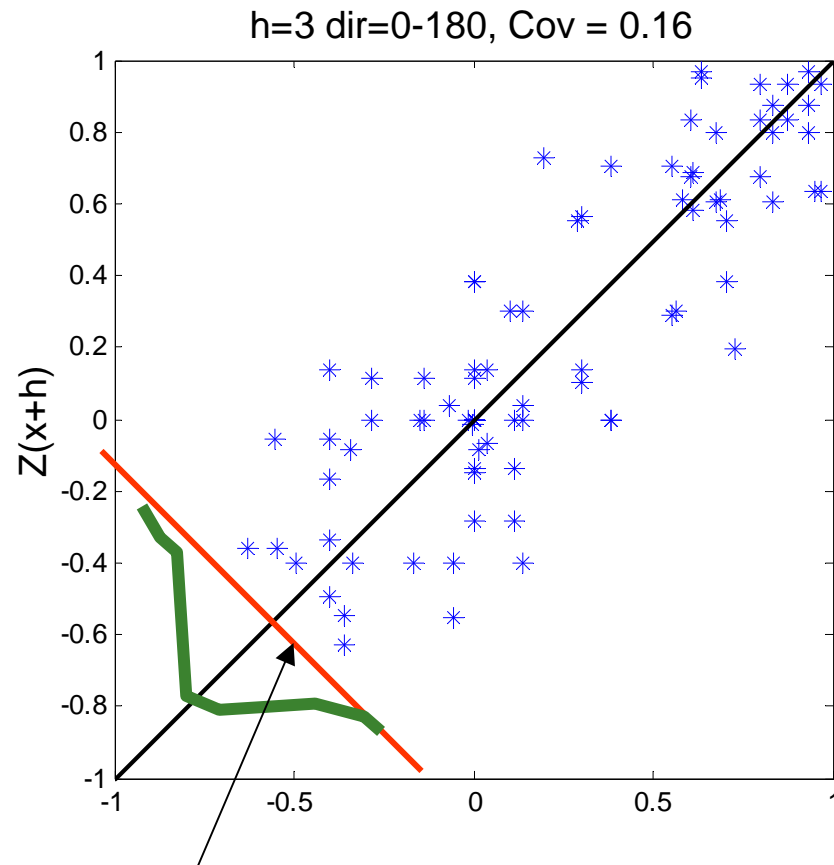
Avec tolérance de 0.5 sur  $|h|$  et  $15^\circ$  sur direction



En pratique, on ne s'intéresse qu'à la covariance (corrélation)



## Le variogramme



Mesure la dispersion sur cette droite

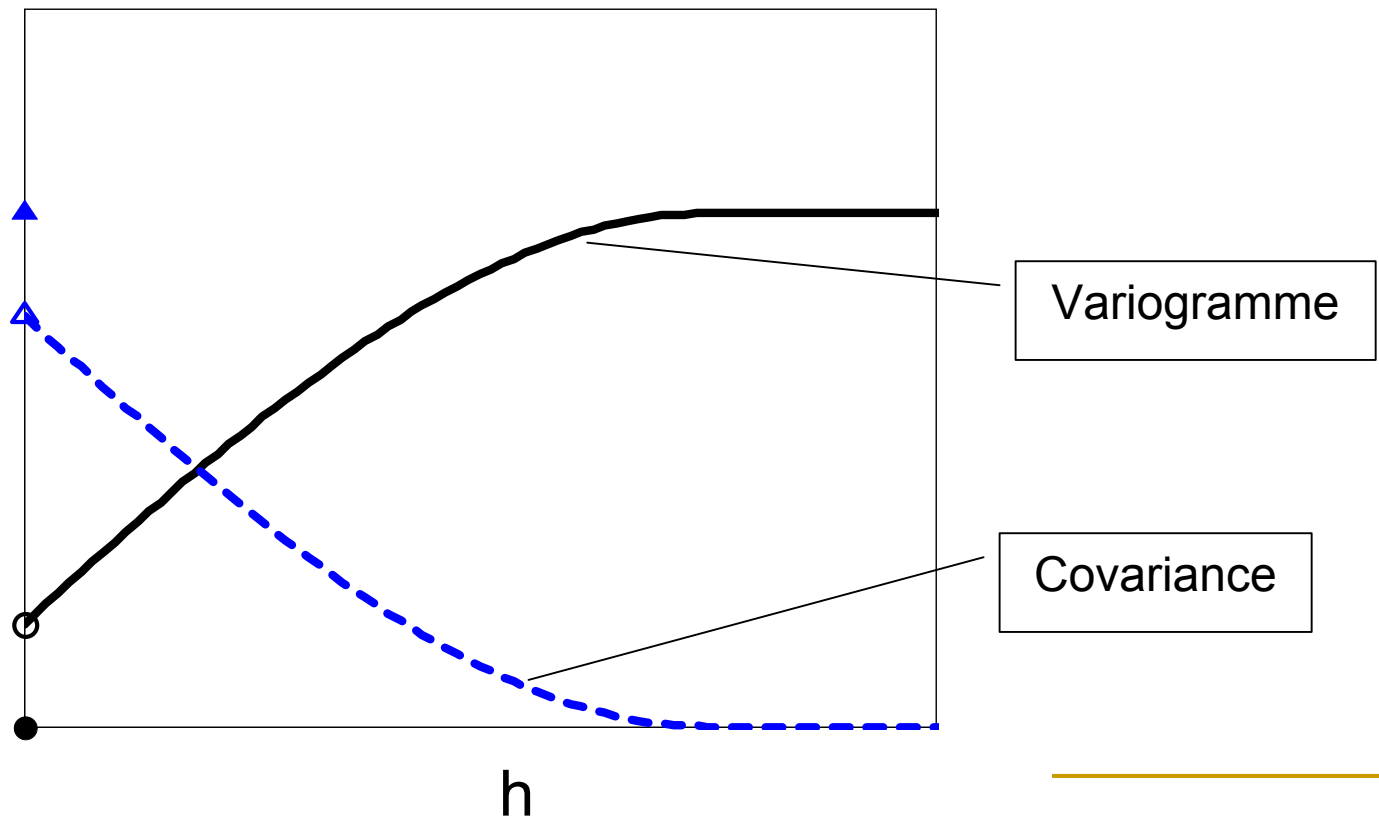
## Variogramme : définition

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x) - Z(x+h)] = \frac{1}{2} E[(Z(x) - Z(x+h))^2]$$

	Variogramme	Covariance
m	Constante	Constante connue
Var	Si existe = palier	Doit exister = Cov(h=0)
h	Croît	Décroît

# Lien entre variogramme et covariance

$$\gamma(h) = \sigma^2 - \text{Cov}(h)$$



# Variogramme expérimental

Choisir une direction + tolérance angulaire

Discrétiser  $|h|$  en classes distinctes

Répartir les paires dans les classes

$$\gamma_e(h) = \frac{1}{2 N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

$N(h)$  : nombre de paires dans la direction considéré et dans la classe de distance  $h$

# Example

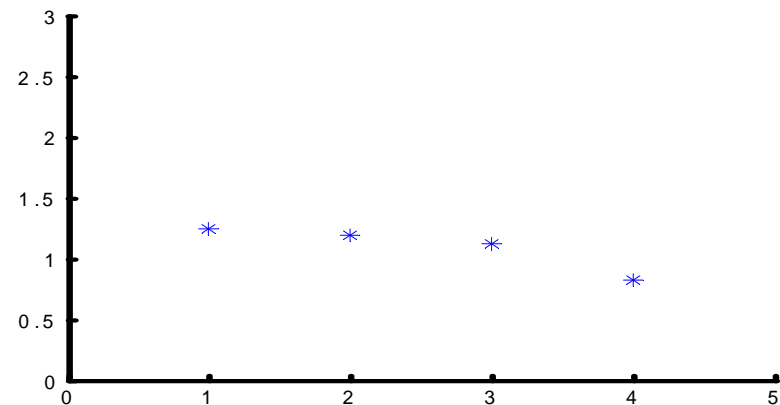
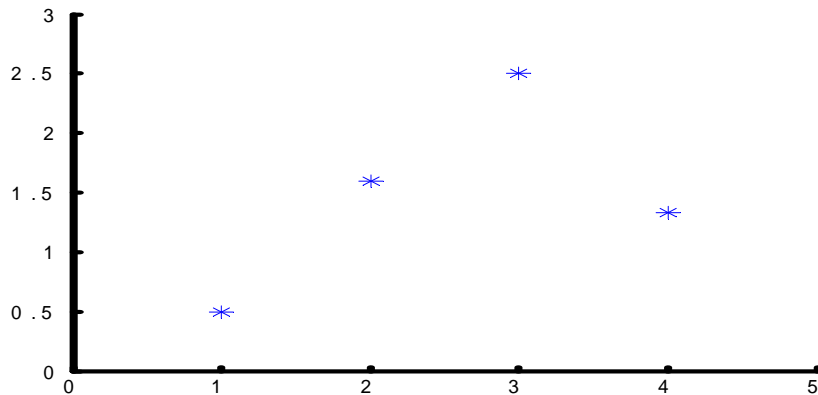
$h=1$

1	2	3	4	3	2	1
---	---	---	---	---	---	---

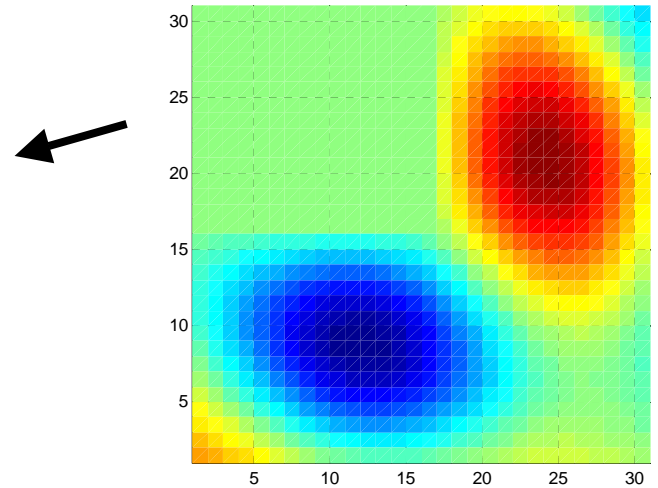
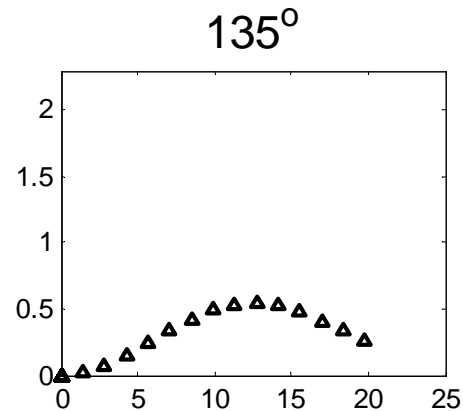
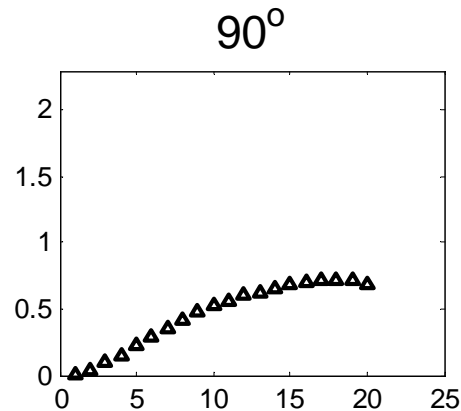
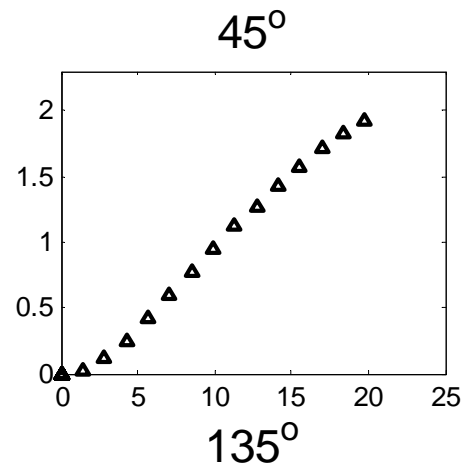
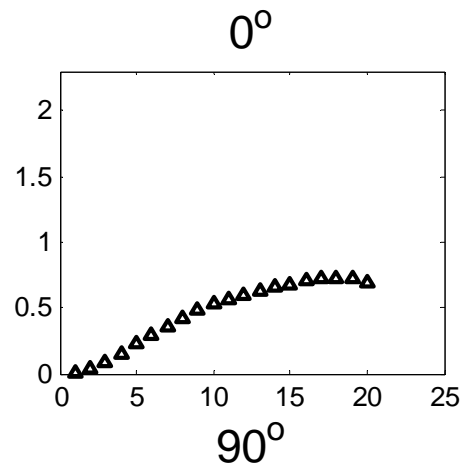
1	3	2	3	1	2	4
---	---	---	---	---	---	---

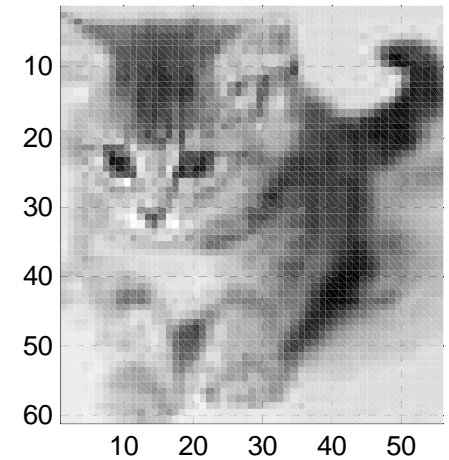
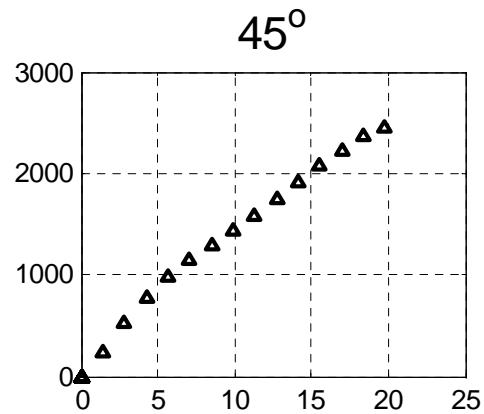
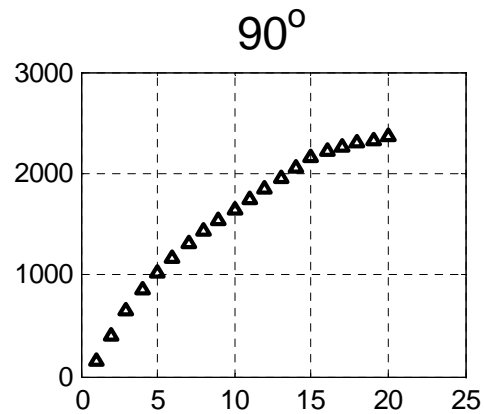
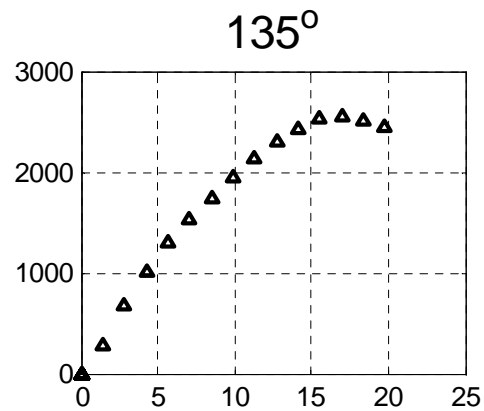
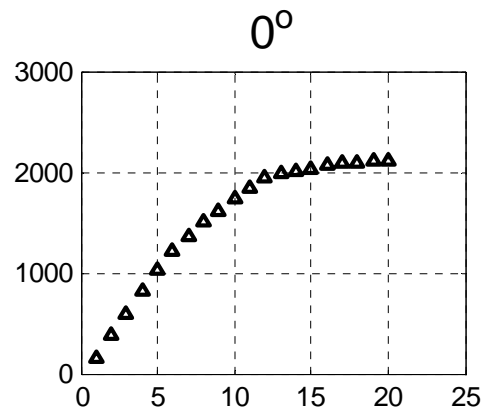
$h$	$N(h)$	$\gamma(h)$
1	6	0.5
2	5	1.6
3	4	2.5
4	3	1.33

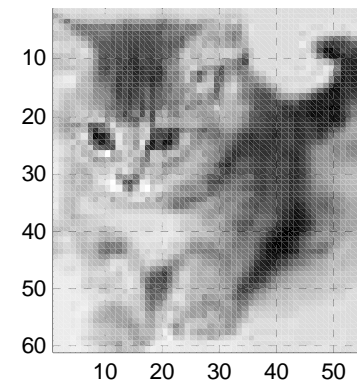
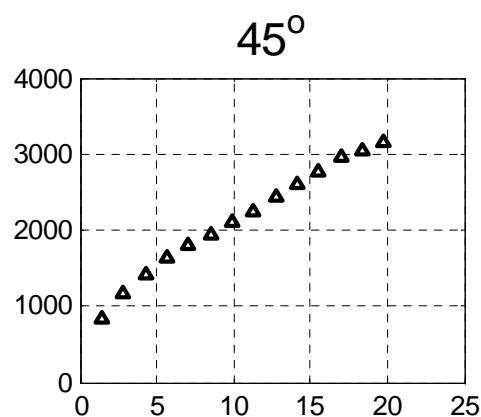
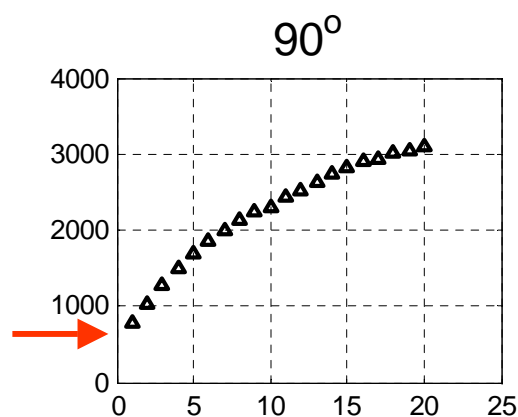
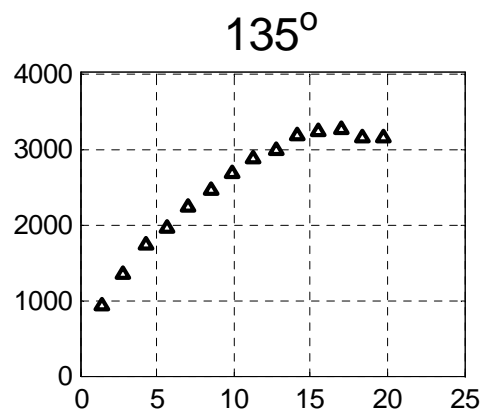
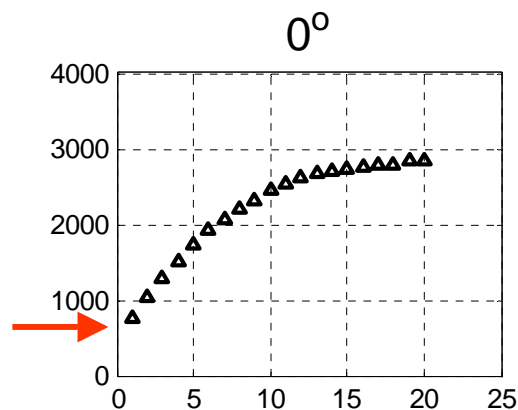
$h$	$N(h)$	$\gamma(h)$
1	6	1.25
2	5	1.25
3	4	1.125
4	3	0.83



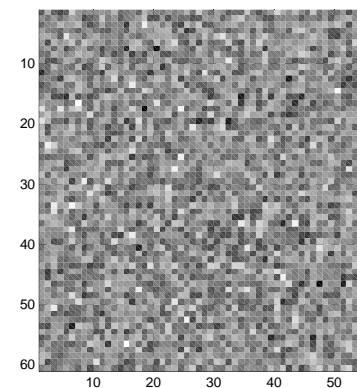
## Le variogramme décrit la continuité spatiale du phénomène



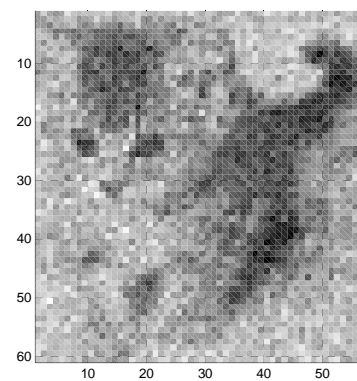




+

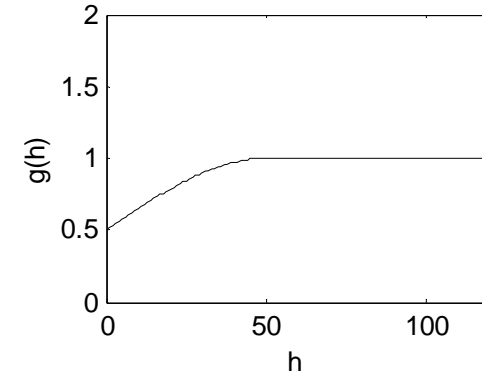
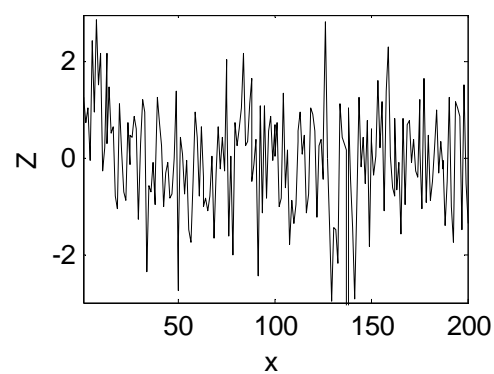
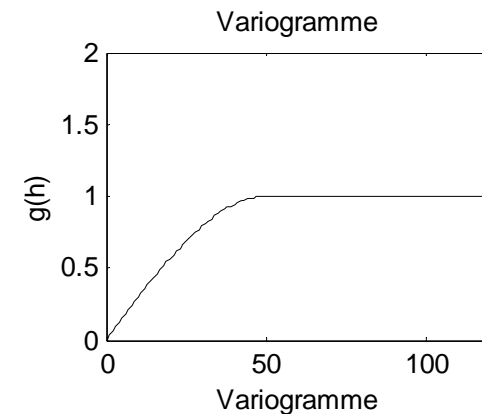
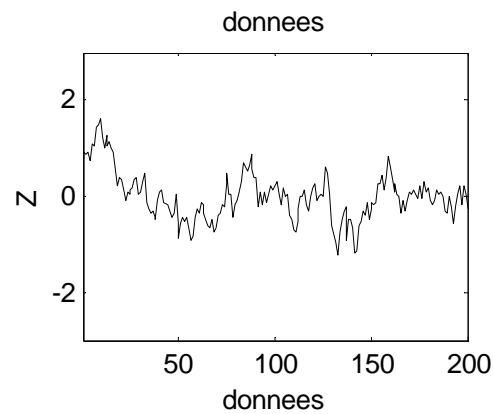
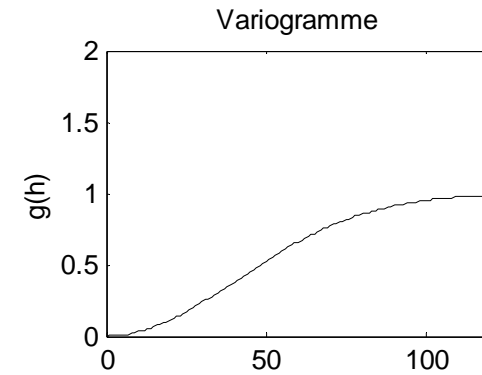
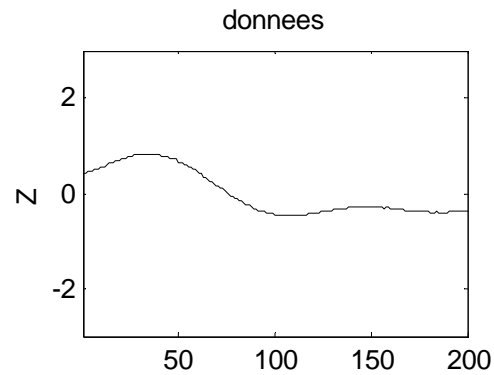


=



Effet pépite causé par le bruit ajouté  
Notez comme la structure sous-jacente  
demeure très visible

## 3 exmples en 1D

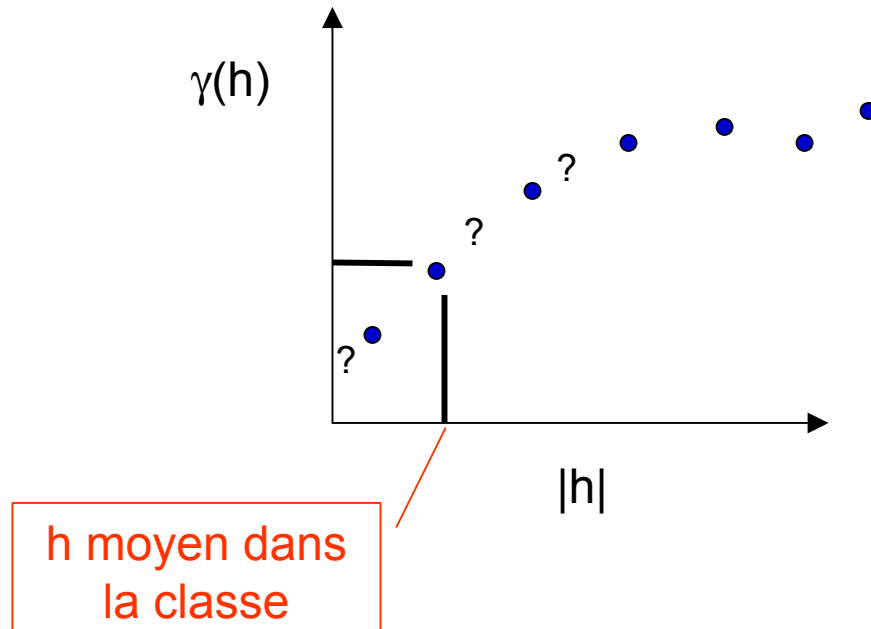


Le variogramme est une statistique d'ordre 2

Ce n'est pas suffisant pour caractériser tous les aspects d'une image ou d'un processus

e.g. on peut créer plusieurs images ayant même « m », même variogramme et présentant pourtant des textures très différentes

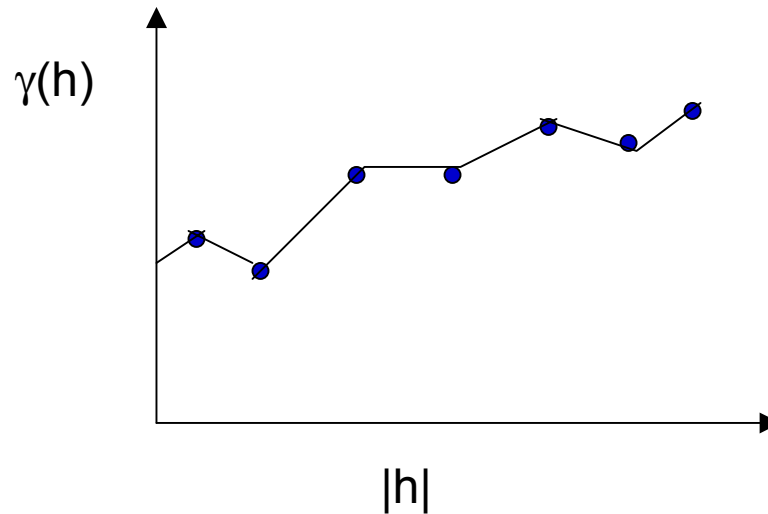
## Variogramme expérimental



Dans les calculs géostat, on doit connaître Cov ou  $\gamma$  pour **tout** «  $h$  »

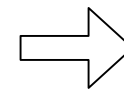


Modèle



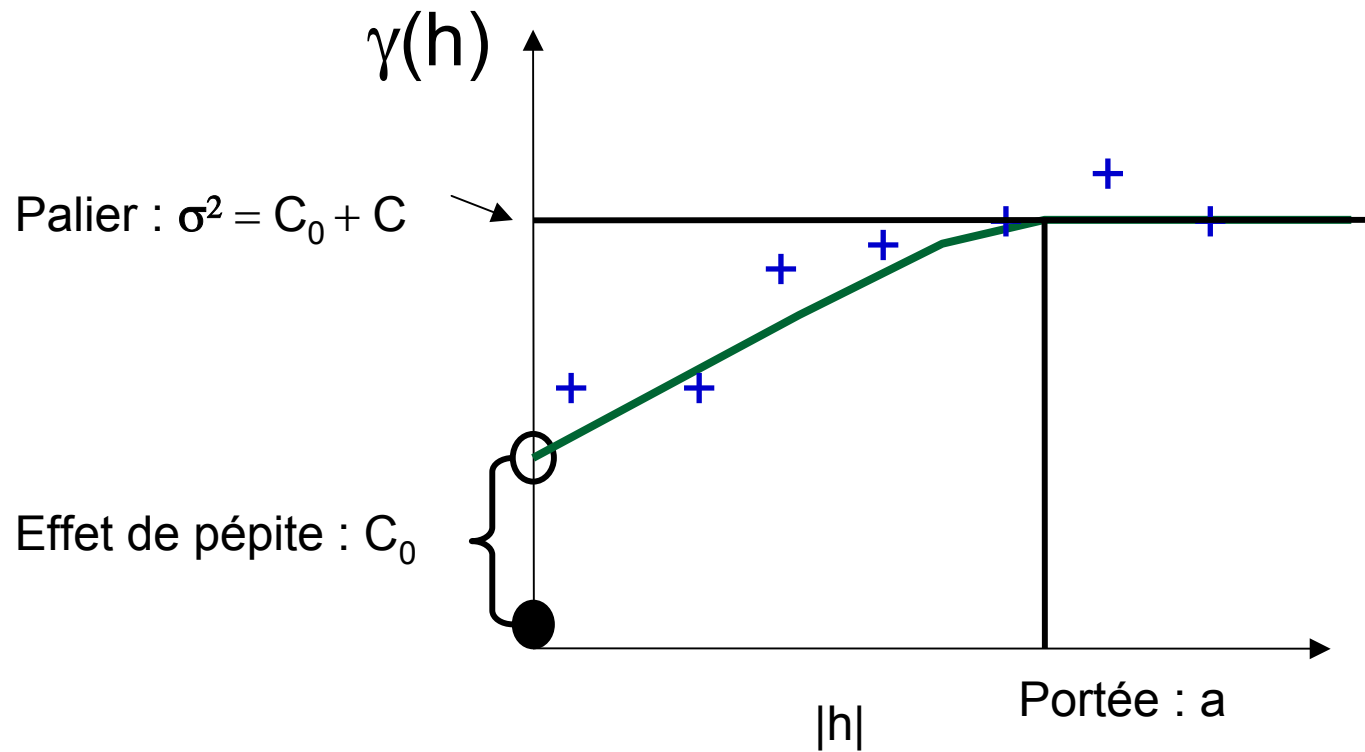
Non, le modèle doit être admissible

Modèle admissible : modèle assurant que toute variance calculée à partir de celui-ci est positive

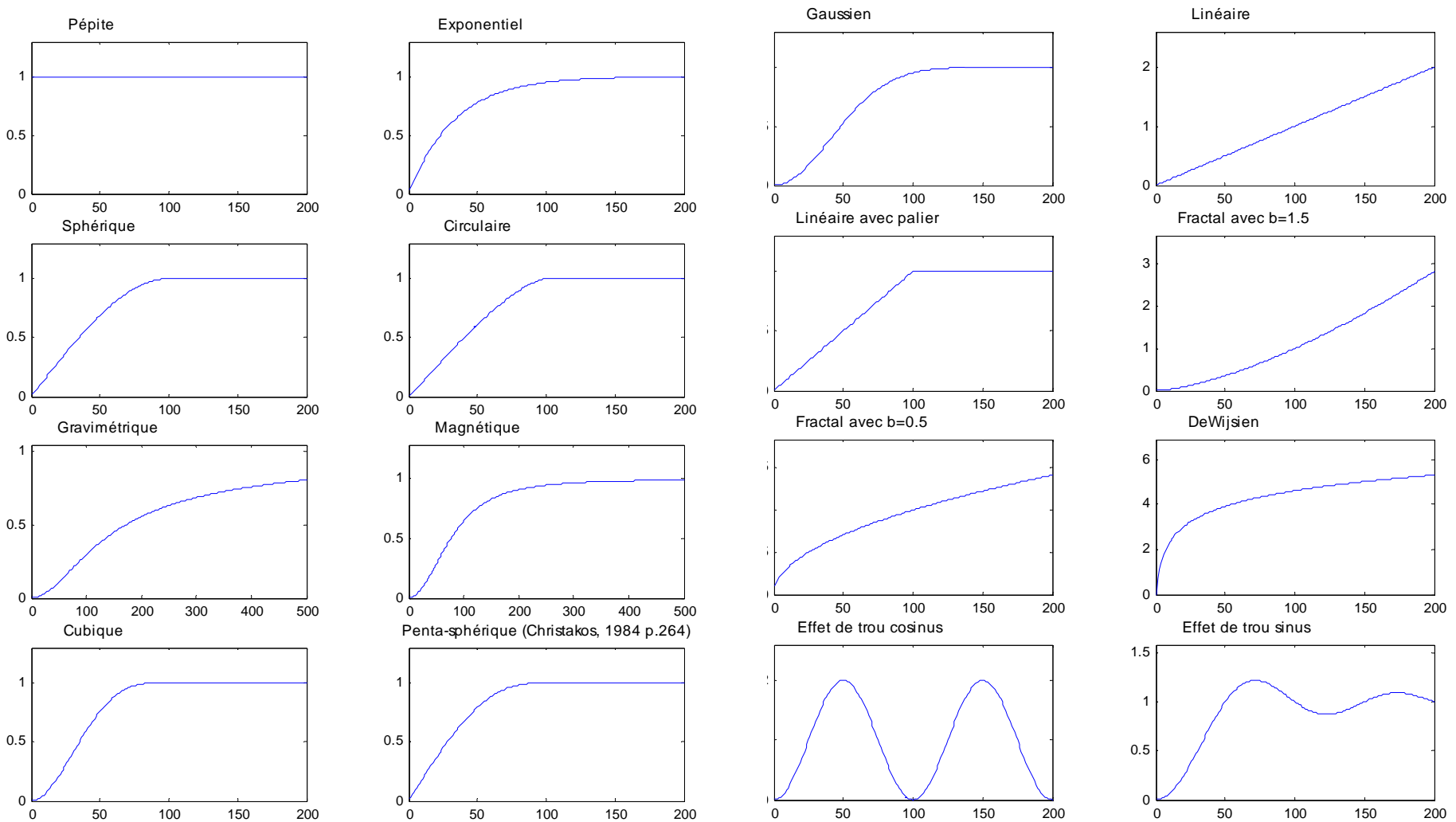


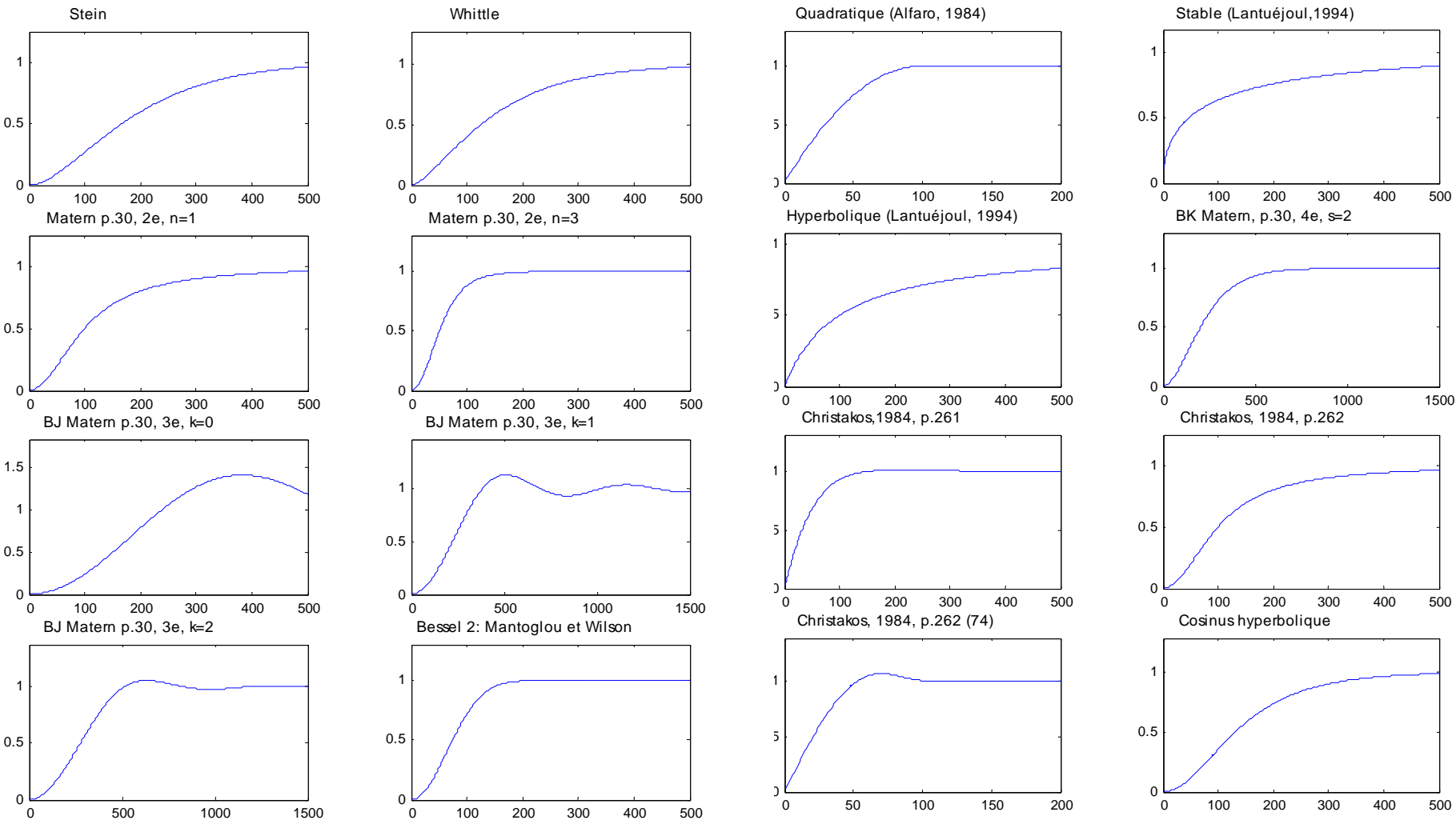
Modèles démontrés  
admissibles

Généralement,



# Exemples de modèle





---

**Toute somme (coefficients positifs) de modèles de variogramme est admissible**

**Toute somme (coefficients positifs) de modèles de covariance est admissible**

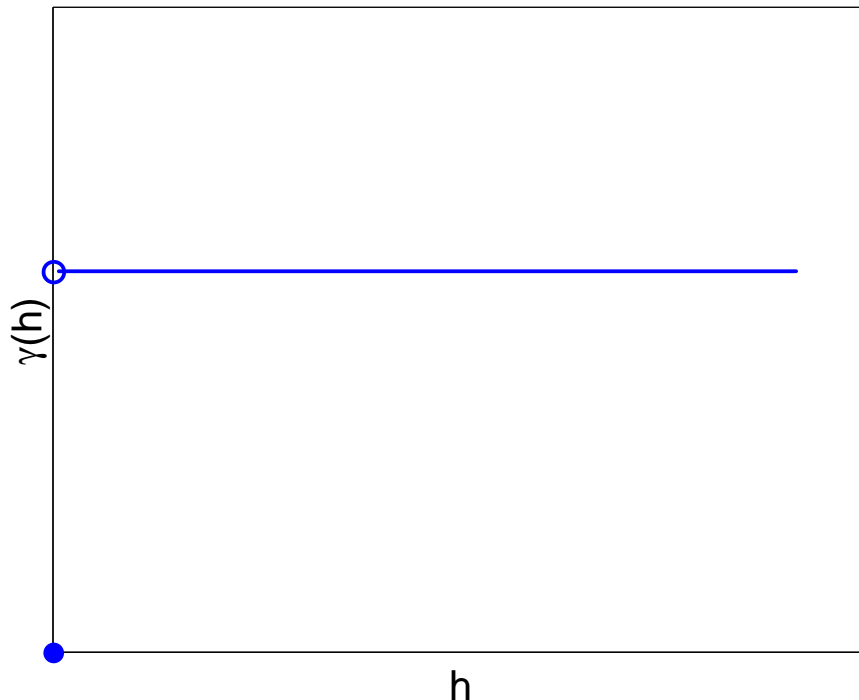
**Tout produit (coefficients positifs) de modèles de covariance est admissible**

**Chaque modèle peut être isotrope ou anisotrope, les directions d'anisotropie peuvent varier d'un modèle à l'autre**

**Un modèle peut être admissible en 1D et non-admissible en 2D, 3D,....**

# Modèles de base en mine

Effet de pépité

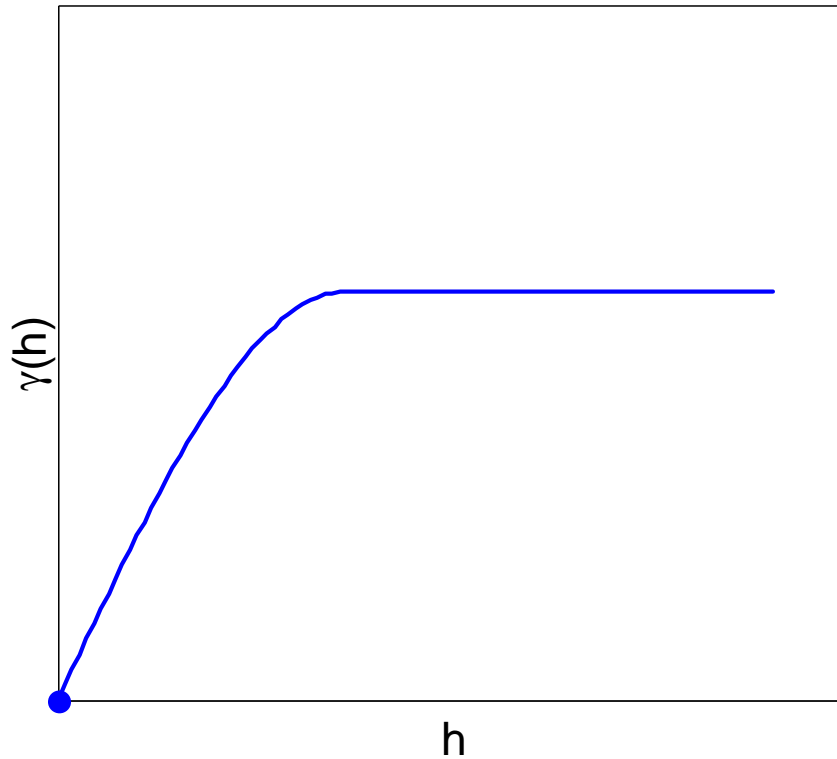


$$\begin{aligned}\gamma(h) &= 0 & \text{si } h &= 0 \\ &C_0 & \text{si } h > 0\end{aligned}$$

- Erreurs de mesure
- Erreurs de localisation
- Erreurs d'analyse (Gy)
- Microstructure non-identifiable dû au manque de données

Presque toujours présent mais  
rarement seul  
Effet de pépité pur =>  
estimation impossible

## Sphérique



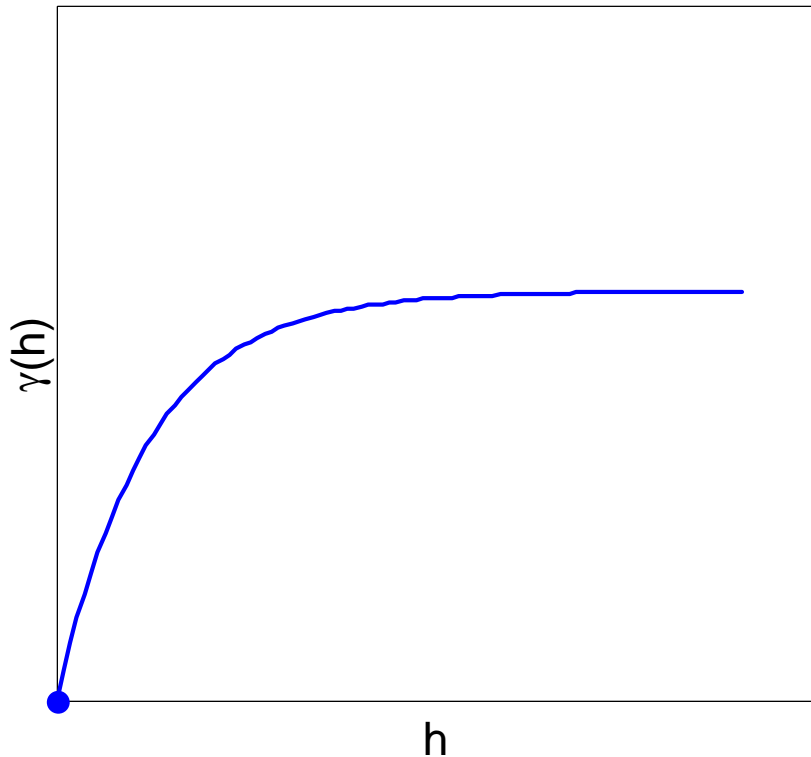
Modèle le + fréquent

e.g. teneur, épaisseur, ...

Combiné avec effet pépité

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C \left( 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \right) & \text{si } 0 < h < a \\ C & \text{si } h \geq a \end{cases}$$

## Exponentiel

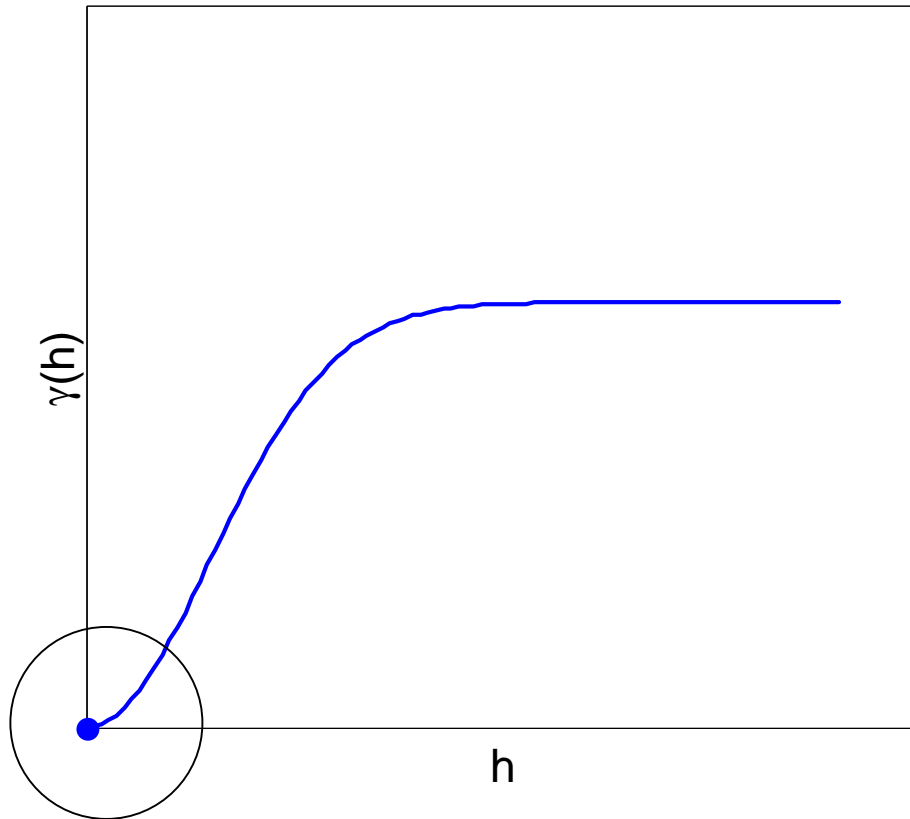


Assez commun

Semblable au sphérique

$$\gamma(h) = C \left( 1 - \exp \left( \frac{-|h|}{a'} \right) \right) \text{ ou } C \left( 1 - \exp \left( \frac{-3|h|}{a} \right) \right) \left. \begin{array}{l} a: \text{ portée effective } \gamma(h)=0.95 \cdot C \\ a'=a/3 \end{array} \right\}$$

## Gaussien

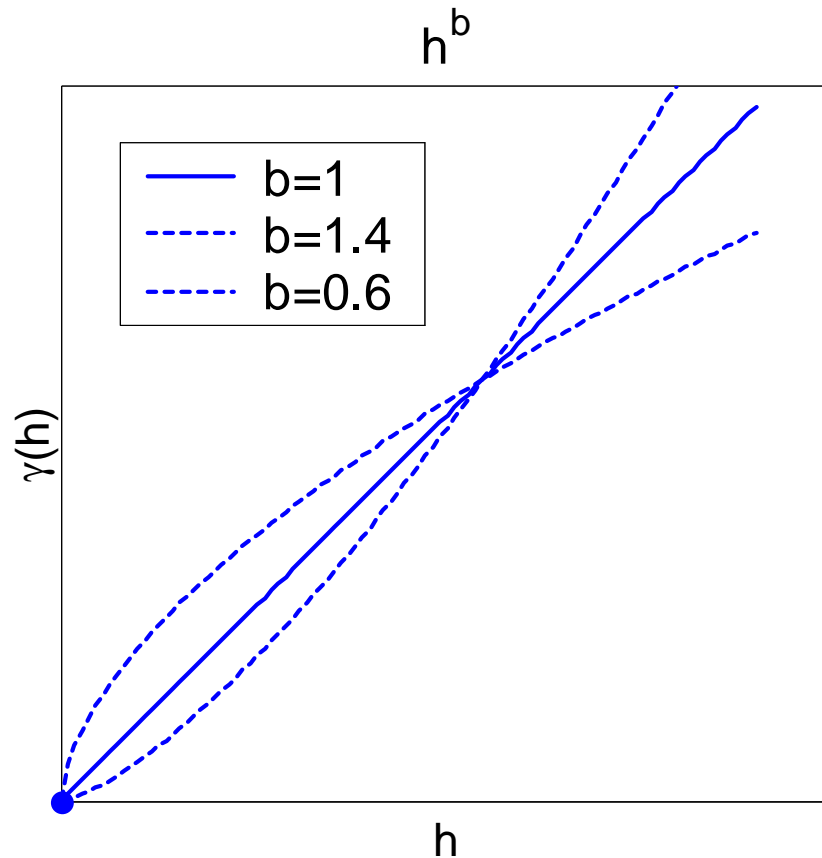


- Peu fréquent en mine
- Variables très continues : e.g. topographie, gravimétrie, magnétisme, épaisseur, ...
- Problèmes numériques si absence d'effet de pépite

$a$ : portée effective  $\gamma(h)=0.95*C$

$$a' = a/3^{0.5}$$

$$\gamma(h) = C \left( 1 - \exp \left( - \left( \frac{|h|}{a'} \right)^2 \right) \right) \text{ ou } C \left( 1 - \exp \left( - 3 \left( \frac{|h|}{a} \right)^2 \right) \right)$$



-Modèles sans palier

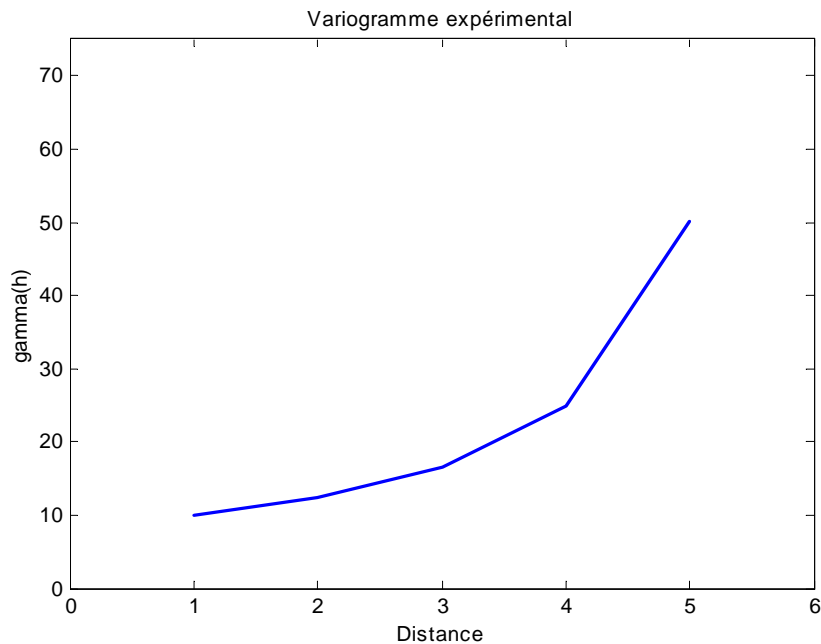
-Moyenne, variance et covariance non définies

$$\gamma(h) = C \left( \frac{|h|}{a} \right)^b \quad |h| > 0, \quad 0 \leq b < 2$$

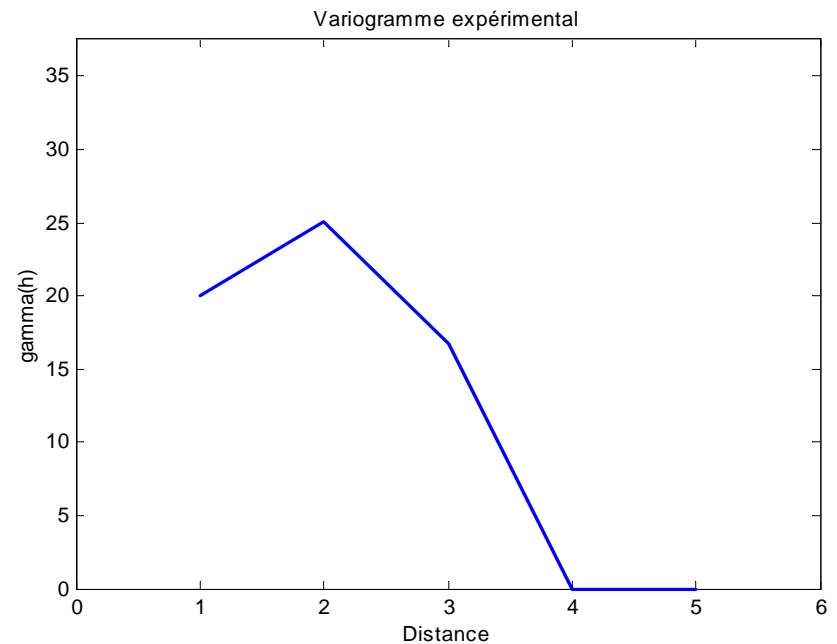
# Problèmes Problèmes Problèmes...

Données extrêmes influencent beaucoup le variogramme

10	0	0	0	0	0
----	---	---	---	---	---



0	0	10	0	0	0
---	---	----	---	---	---

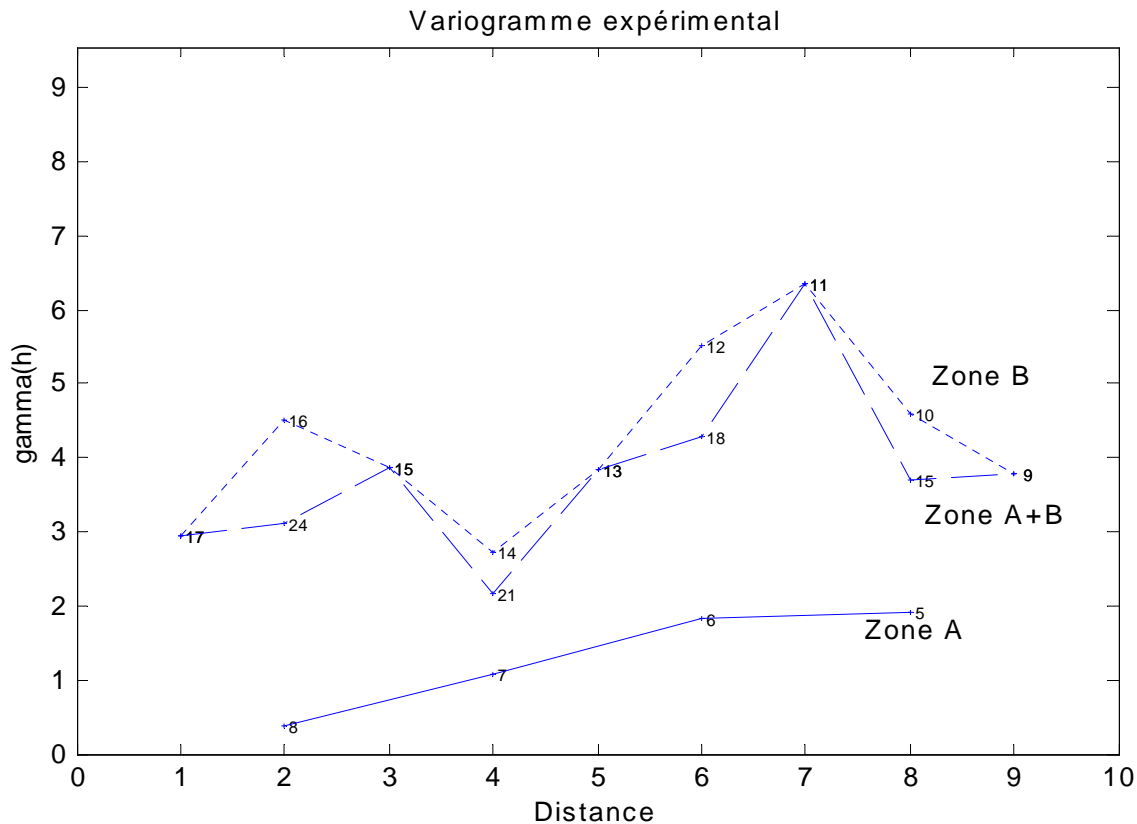


Zone A: (pas de 2m)

4	4	5	6	6	7	6	5	4
---	---	---	---	---	---	---	---	---

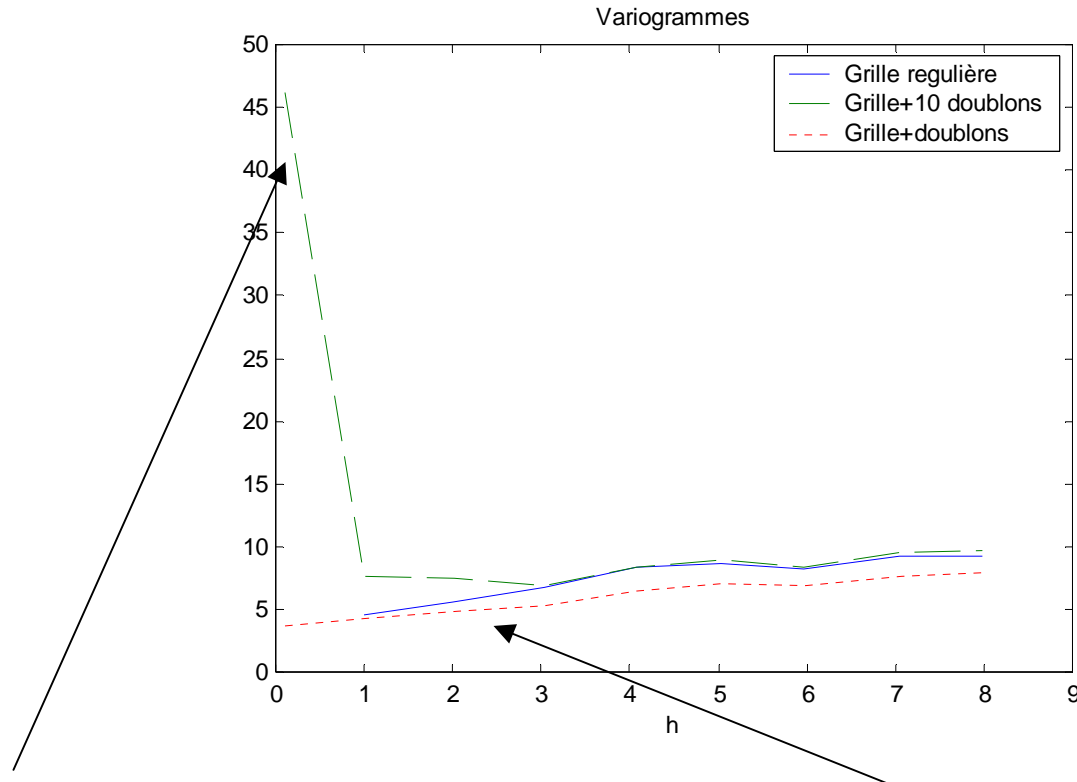
Zone B: (pas de 1m), zone +variable

8	6	8	10	12	8	10	12	14	10	8	6	12	8	10	10	8	10
---	---	---	----	----	---	----	----	----	----	---	---	----	---	----	----	---	----



- Étudier les zones séparément ?
- Sous-échantillonner la zone B

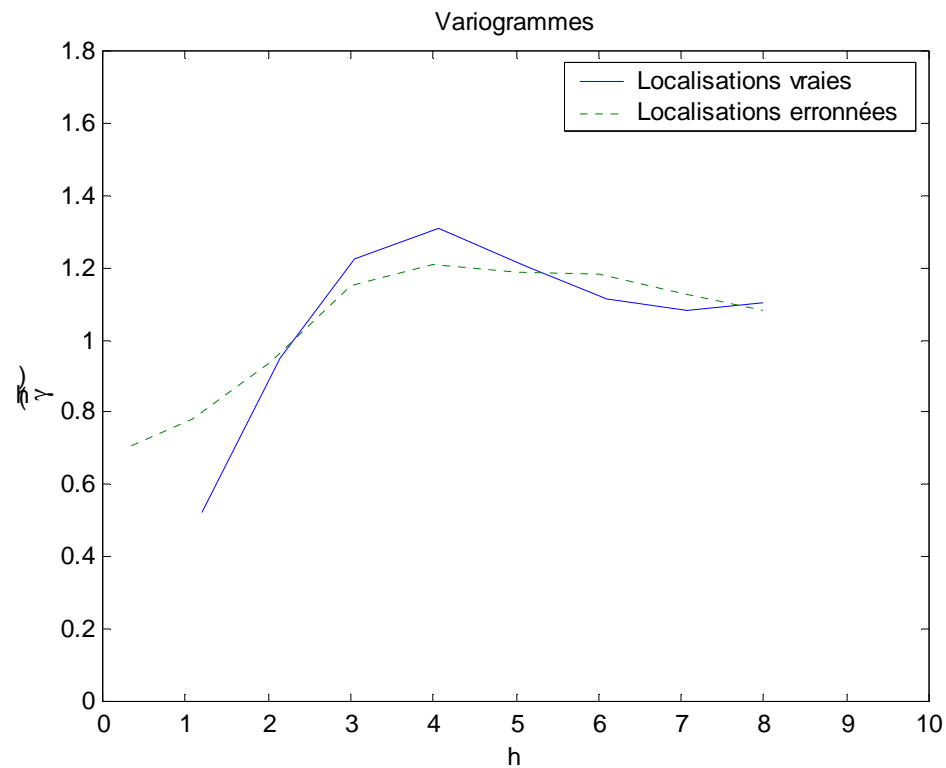
## Pratique de « confirmer » préférentiellement les teneurs riches



A pour effet de fournir beaucoup de paires à petite distance et montrant de très fortes variations

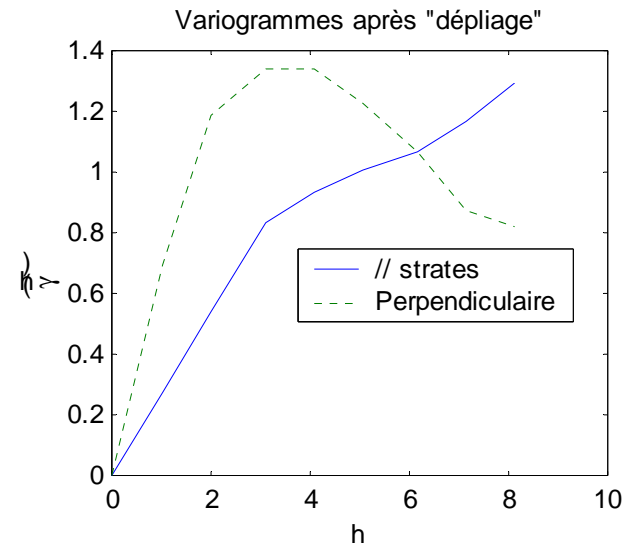
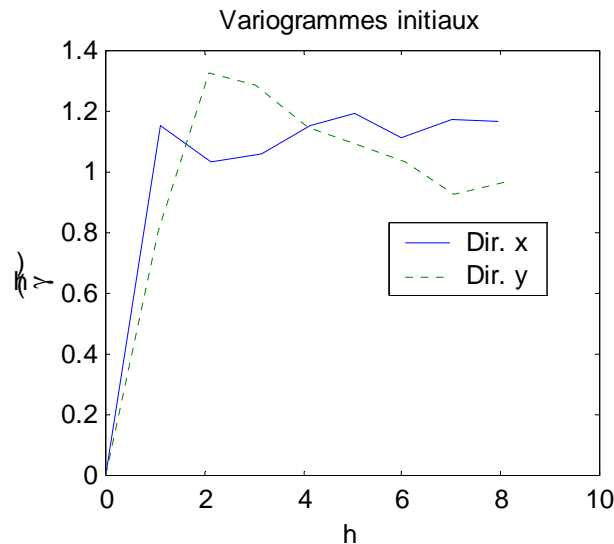
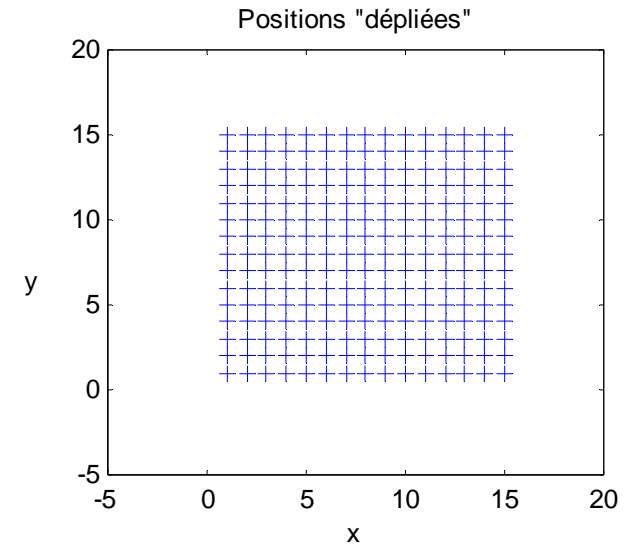
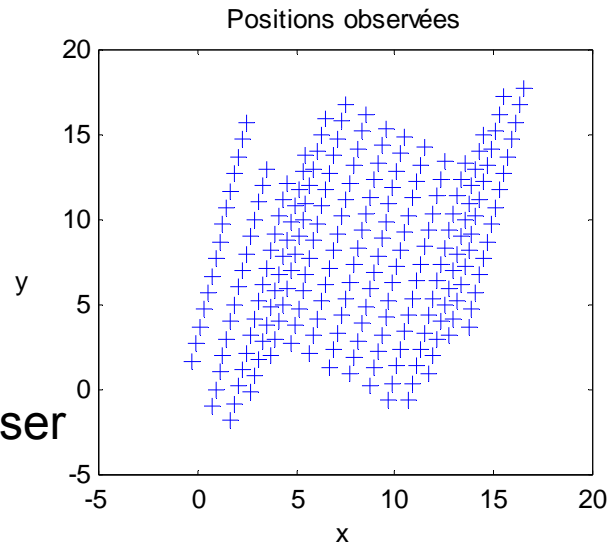
Si on rééchantillonne chaque point, pas de problème

## Erreurs de localisation



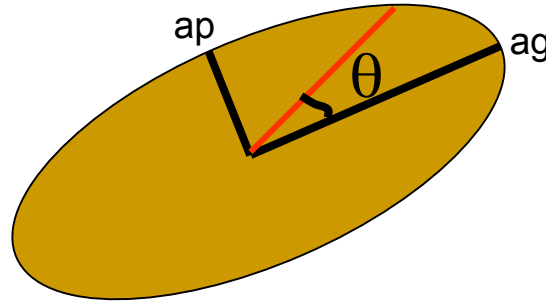
# La géologie ne collabore pas toujours

Certains logiciels permettent de passer à un système de coordonnées géologiques

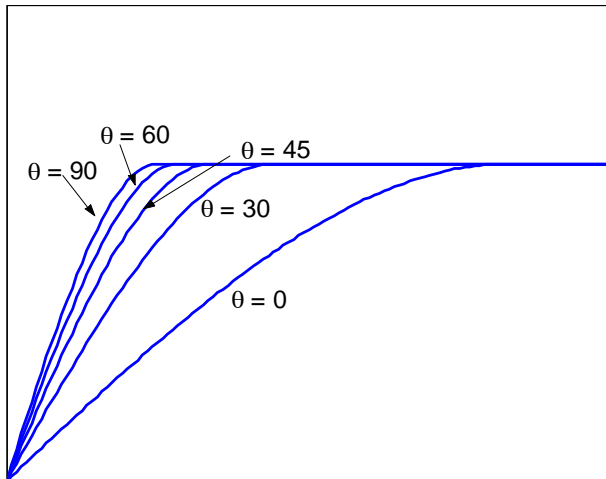


# Anisotropies

1- Géométrie : les portées décrivent une ellipse

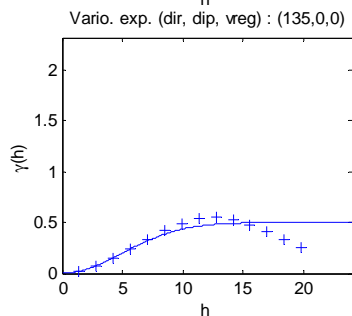
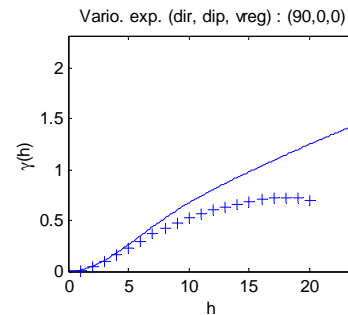
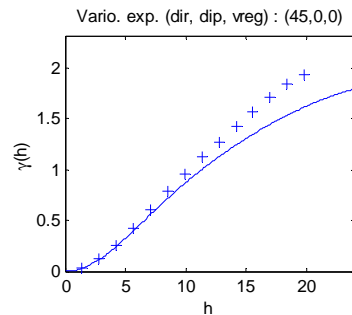
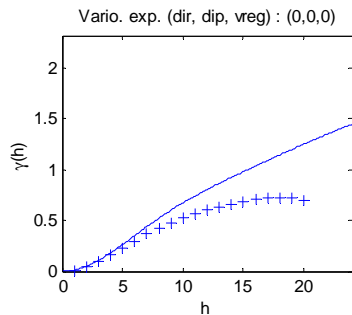


$$a_{\theta} = \frac{a_g a_p}{\left\{ a_p^2 \cos^2 \theta + a_g^2 \sin^2 \theta \right\}^{1/2}}$$



2- zonale : toute anisotropie qui n'est pas géométrique

=> somme de composantes isotropes et avec anisotropies géométriques



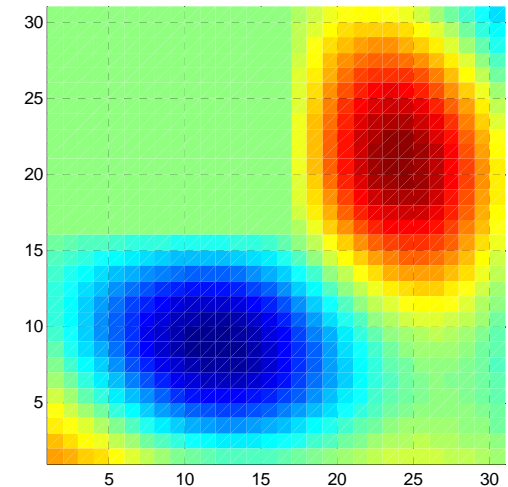
Gaussien isotrope  $a=7$ ,  $C=0.5$

+

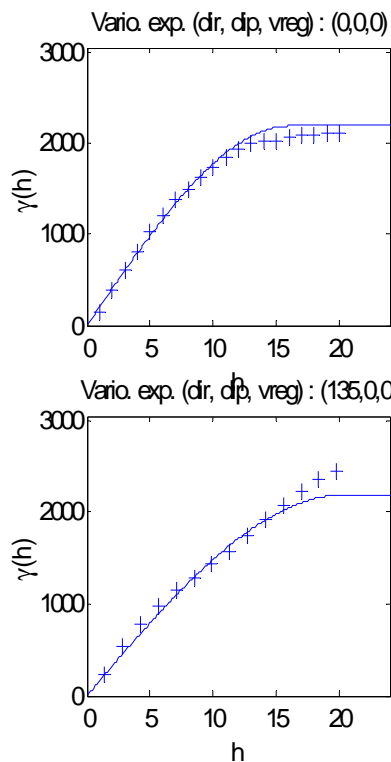
Gaussien aniso géom.

$$a_{45}=17$$

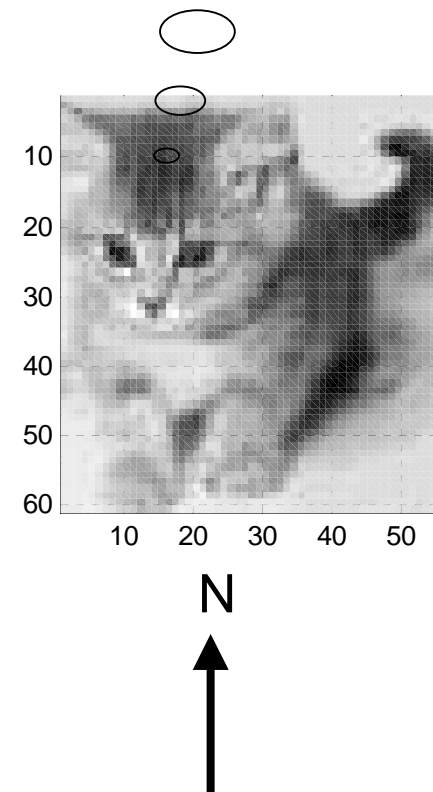
$$a_{135}=\infty$$



**Ajustement assez bon pour 0-5 pixels dans toutes les directions**



Modèle sphérique avec  
anisotropie géométrique  
 $A(135)=20.4$ ,  $a(45)=13.8$



# Stratégie de modélisation

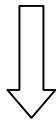
- Définition minutieuse du domaine
- Examen des données, données extrêmes ?
- Au besoin sous-échantillonnage des données pour éviter de sur-représenter des zones particulières
- Variogramme omnidirectionnel => modèle isotrope candidat
- Déterminer directions géologiques principales
- Calculer variogrammes directionnels (au moins 4 directions) *attention aux paramètres de calculs (classes de distance et tolérance)*
- Comparer variogrammes directionnels au modèle isotrope candidat
  - acceptable => terminé
  - Inacceptable => ajuster un modèle anisotrope (géométrique)
    - anisotrope (géométrique) acceptable => terminé
    - Inacceptable => anisotropie zonale ?

Dans tous les cas, il importe surtout d'ajuster les premiers points du variogramme  
Éviter de « surajuster » les données

# Remarques concernant le calcul des variogrammes

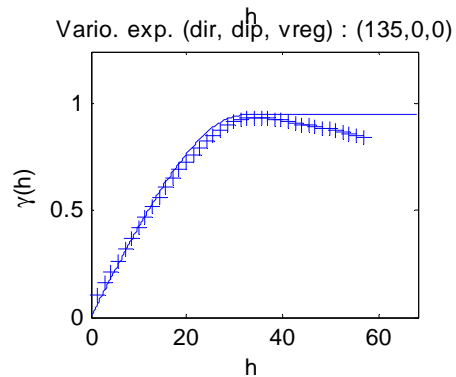
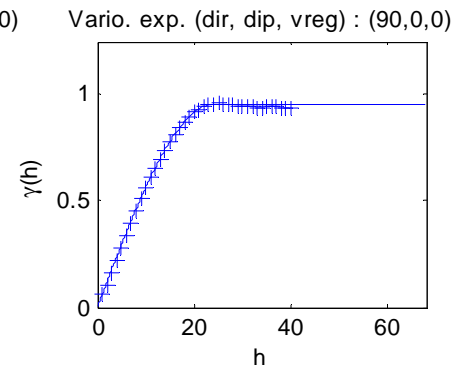
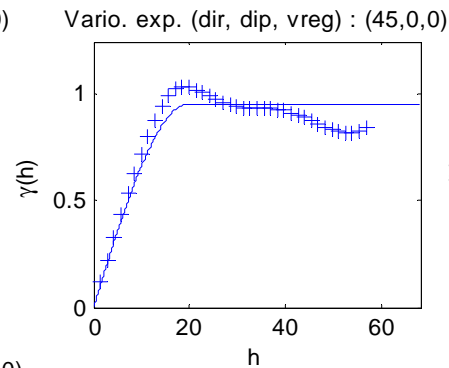
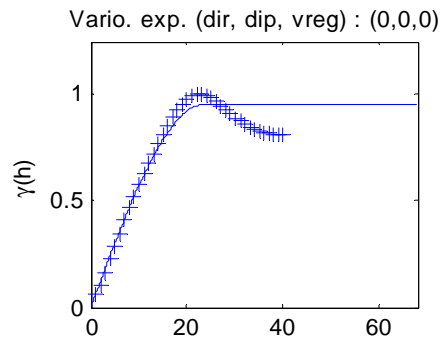
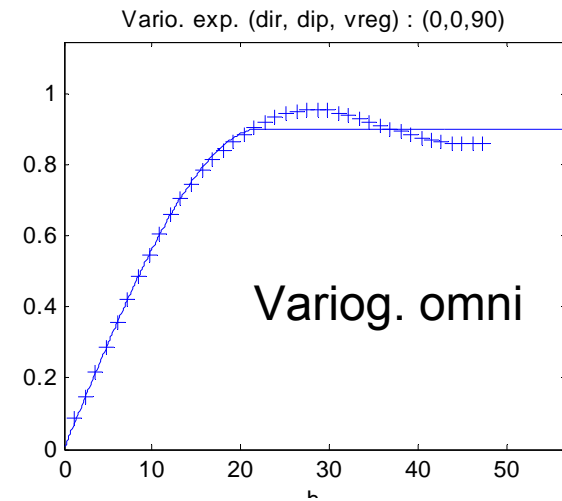
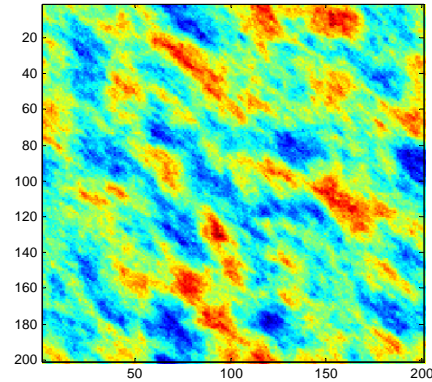
## Objectifs:

- au moins 30 paires pour la plupart des points du variogramme
- 4-6 points avant le palier pour pouvoir ajuster modèle
- $h < h_{\max}/2$



- Doit avoir un minimum de données
  - >30 pour variogramme omnidirectionnel
  - >60 pour variogrammes directionnels
- Influence le choix de largeur des classes

# Exemple

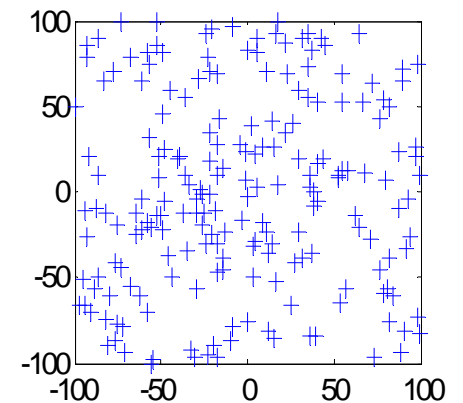
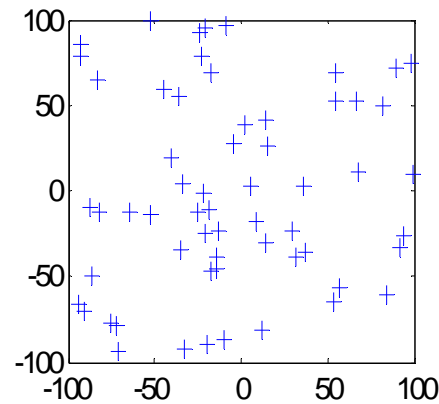
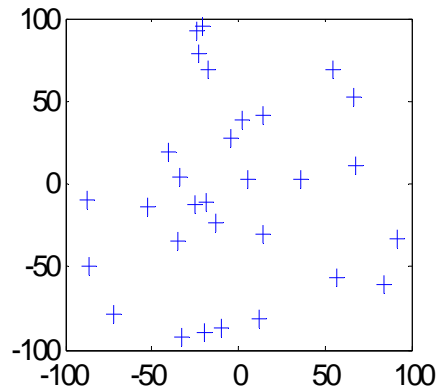
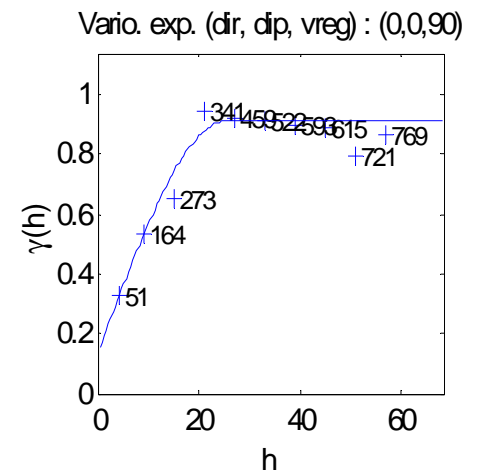
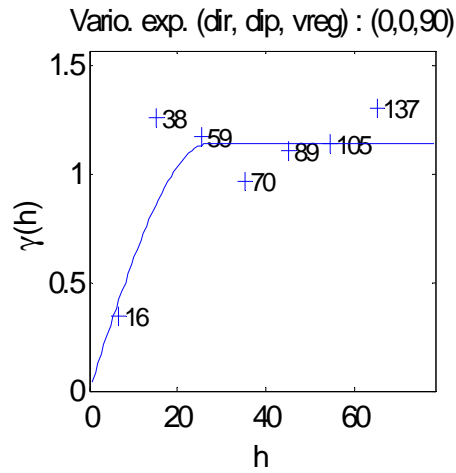
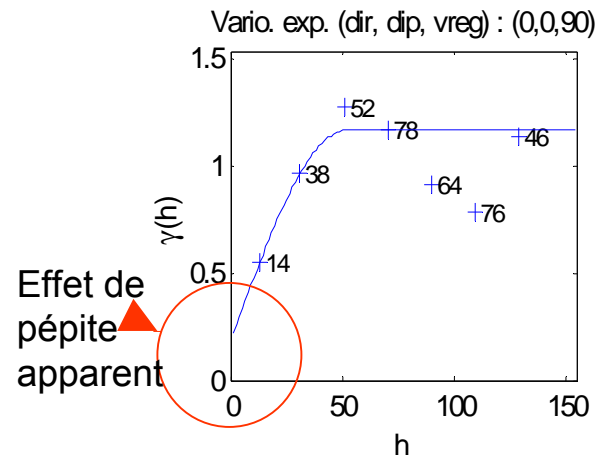


Variog. directionnels

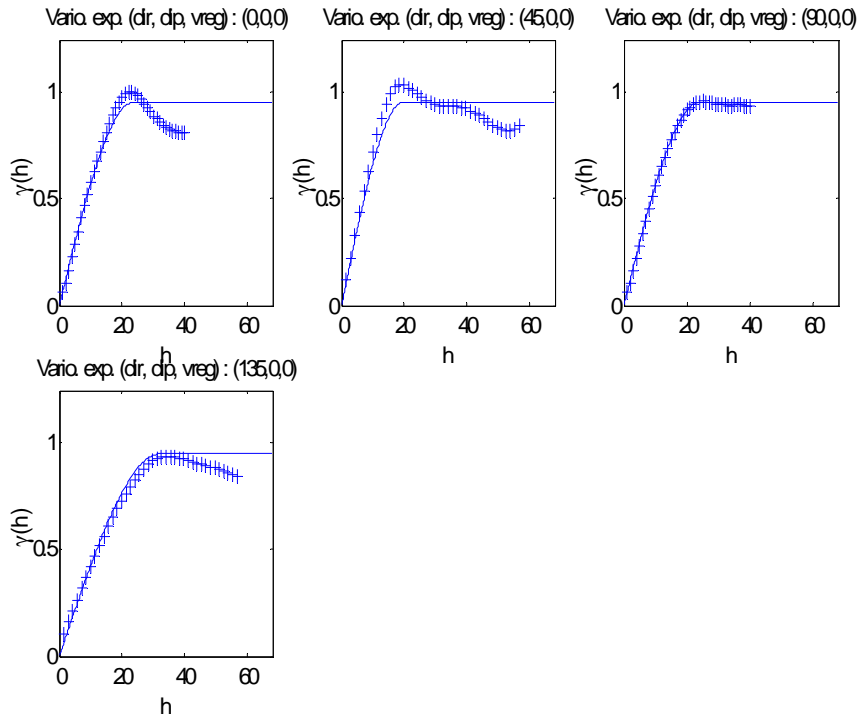
N=30

N=60

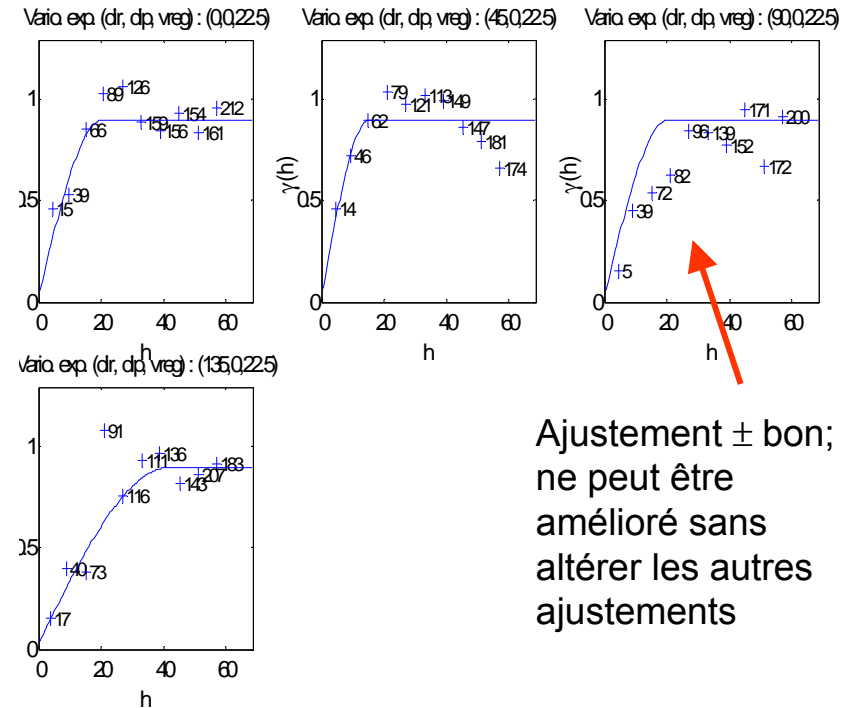
N=200



N=40000

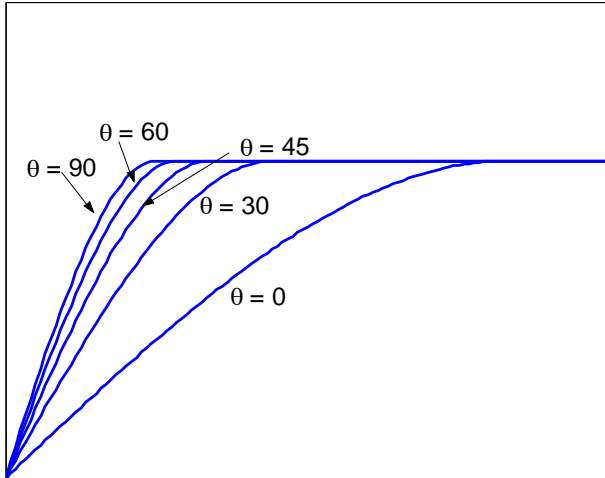


N=200



Ajustement  $\pm$  bon;  
ne peut être  
amélioré sans  
altérer les autres  
ajustements

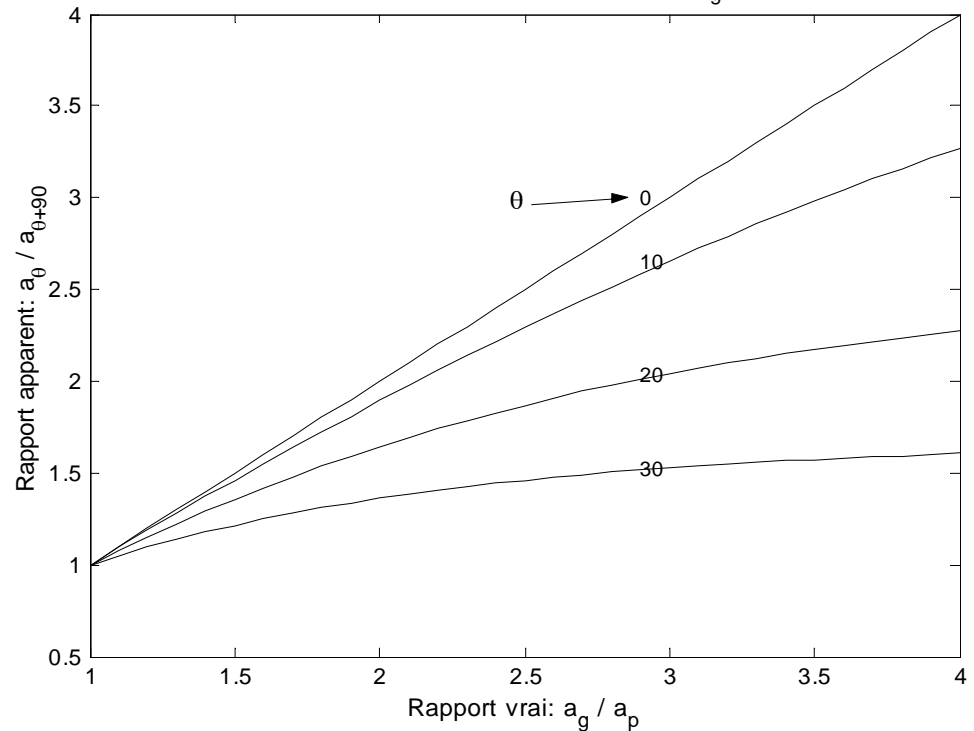
Pour détecter une anisotropie, il faut pouvoir calculer le variogramme dans la direction de meilleure continuité. Celle-ci n'est pas toujours connue.



Tout écart => sous-  
estimation de l'anisotropie

Faible anisotropie peut  
passer inaperçue

Rapport apparent en fonction de la direction ( $\theta$ : angle avec  $a_g$ ) et du rapport d'anisotropie

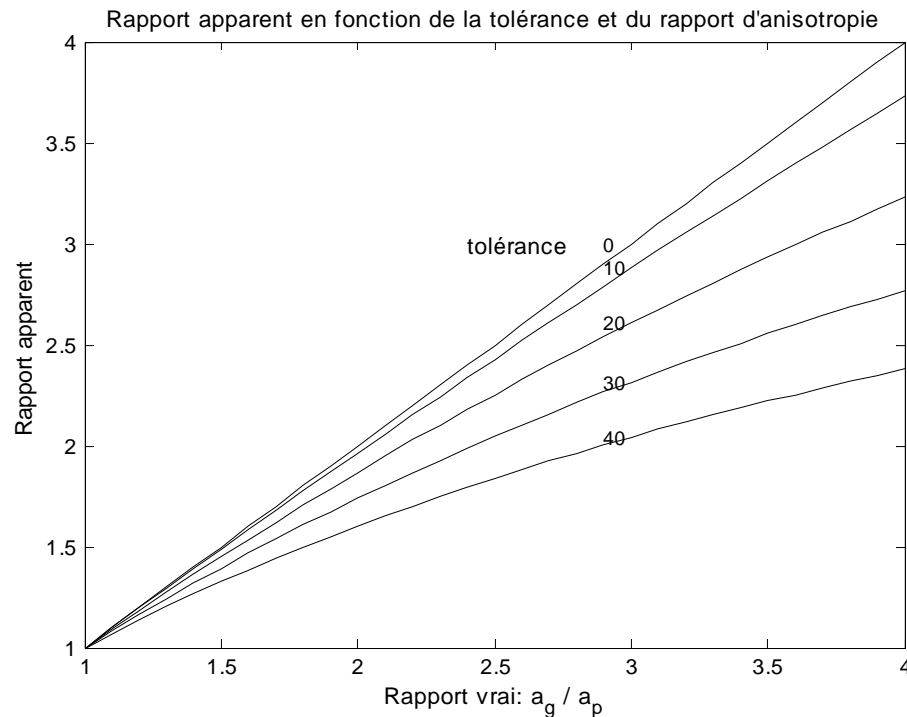


Tolérance angulaire doit être maintenue faible <22.5, idéalement 0-10.

Tolérance trop grande

=> variogrammes peu directionnels

=> sous-estimation de l'anisotropie



Meilleures directions pour le calcul => directions des forages

- Permet de bien estimer le variogramme à petite distance
- Erreurs de localisation et de direction ont moins d'impacts sur le variogramme car les distances inter-carottes demeurent inchangées

Hic: - Nécessite des forages ayant différentes directions pour pouvoir modéliser l'anisotropie

# Autres outils utiles

- Validation croisée de modèles candidats (krigeage); eg. Tester un modèle isotrope vs anisotrope; tester un effet de pépité de 10% vs 30%;...
- Modèle permet de prédire les variances des composites de tailles différentes ?
- Modèle permet de prédire les variances des valeurs krigées ?
- Variogramme des  $\log(\text{teneurs})$  pour identifier les anisotropies possibles, la (les) portée, l'importance approximative de l'effet de pépité
- Variogramme d'une transformation des teneurs (e.g. rang), même chose que les log