Krigeage d'indicatrice

Le krigeage d'indicatrice est utilisé pour estimer les ressources dans un certain nombre de rapports NI43-101 sous le vocable "multiple indicator kriging" (MIK). Vous verrez un peu plus tard d'où vient le terme "multiple".

Plan

- Introduction: contexte et problématique
- Cas d'une variable binaire
- Cas d'une distribution seuillée
- Interprétation de la valeur estimée
- Équations du krigeage d'indicatrices
- Corrections d'ordre
- Changements de support
- « Soft kriging »
- Exemples

Contexte et problématique

Krigeage d'indicatrices : méthode géostatistique non-linéaire =>

cherche à estimer la fonction de distribution conditionnelle en tout point;

par contraste, le krigeage ordinaire n'estime que la moyenne conditionnelle (sauf dans le cas gaussien, *la variance de krigeage n'est pas une variance conditionnelle*).

Exemples :

- Estimation du volume in-situ de sols contaminés; sur un site donné, quel est le volume de sols excédant le critère « C » du MENV ? Quelle quantité totale de contaminants y retrouve-t-on ?

- Dans une mine, quel est le tonnage de minerai (in-situ); quelle est la teneur moyenne ou la quantité de métal contenue dans le minerai?

Exemples:

- Dans une mine, la sélection finale des blocs se fait à partir d'estimations qui seront obtenues une fois les données des blocs voisins connues. Peut-on prédire *maintenant* la proportion des blocs qui seront identifiés plus tard comme du minerai ? Quelle devrait être la teneur de ces blocs sélectionnés sur des *estimations futures* ?

Note: il y a 3 grandes catégories de méthodes pour répondre à ce genre de questions :

- le krigeage d'indicatrices (et ses variantes multivariables)
- les méthodes gaussiennes (multigaussien)
- le krigeage disjonctif (lois bivariables isofactorielles)

(note: dans le cours, on ne voit que les 2 premières méthodes)

en fait que la première, la seconde indirectement par les méthodes de simulation

Variable binaire

a) Variable binaire (0-1) (ex. I(x)=1 si on a le faciès A au point x et I(x)=0 si l'on a un autre faciès)

Quelle interprétation donner au résultat du krigeage dans ce contexte ?

Soit I(x) la v.a. binaire au point x.

$$E[I(x)] = P(I(x)=0)*0 + P(I(x)=1)*1 = P(I(x)=1)$$

Donc l'espérance d'une variable indicatrice est une probabilité. Or on sait que le krigeage estime (assez bien) l'espérance conditionnelle d'une variable, d'où l'idée de kriger la variable indicatrice et d'interpréter le résultat du krigeage comme une probabilité.

Si l'on tient compte des observations disponibles:

$$E[I(x)|I(x_1),I(x_2)\ldots I(x_n)] = P(I(x)=1|I(x_1),I(x_2)\ldots I(x_n))$$

Pour une variable continue, le krigeage est un bon estimateur de l'espérance conditionnelle, on *suppose* que ceci demeure vrai pour une indicatrice.

$$I^*(x) = \sum_{i=1}^{n} \lambda_i I(x_i)$$

où les poids sont obtenus par krigeage ordinaire de la variable indicatrice est donc une estimation de $P(I(x)=1|I(x_1),I(x_2),\ldots I(x_n))$

$$I^*(x) \equiv P^*(I(x) \mid I(x_1), I(x_2), \ldots I(x_n))$$

Généralisation: variable continue

Soit Z(x) une v.a. continue définie au point x et F(x,c), la fonction de répartition de la v.a. au point x pour la valeur « c ».

Par définition: $F(x,c) = P(Z(x) \leq c) = E[I(x,c)]$

où $I(x,c) = 1$ si $Z(x) \leq c$

0 si $Z(x) > c$

Par krigeage ordinaire d'indicatrices, on aura :

$$I^*(x,c) = \sum_{i=1}^{n} \lambda_i I(x_i,c)$$

Cette valeur *devrait* être un bon (?) estimateur de :

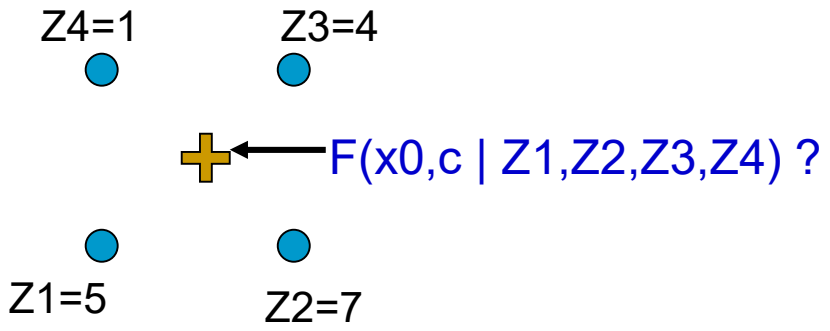
$$P(I(x,c)=1 \mid I(x_1,c), I(x_2,c),...I(x_n,c)) =$$
$$P(Z(x)<c \mid I(x_1,c), I(x_2,c),...I(x_n,c)) =$$
$$F(x,c \mid I(x_1,c), I(x_2,c),...I(x_n,c))$$

Cette dernière fonction est la fonction de réparttion conditionnelle aux indicatrices observées dans le voisinage

Si l'on choisit une infinité de « c » différents, on aura une estimation de la fonction de répartition au point « x » conditionnelle aux indicatrices obtenues aux points échantillons.

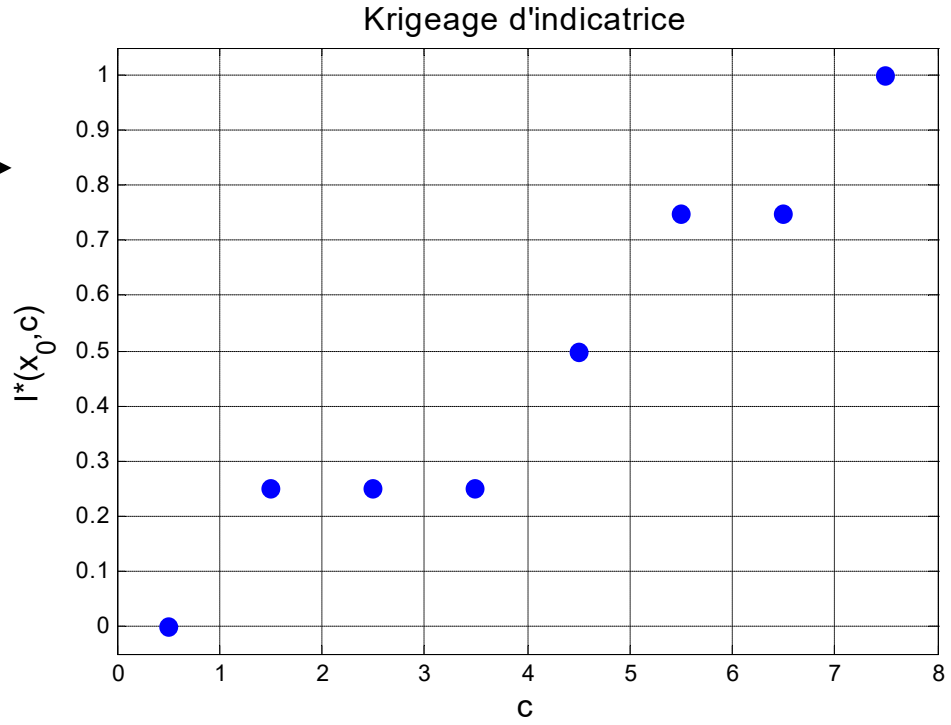
d'où l'appellation "multiple" utilisée par certains (plusieurs seuils)

Exemple



Ici, par symétrie, les poids valent tous 1/4

c	F*(x0,c,|(n))
0.5	0
1.5	0.25
2.5	0.25
3.5	0.25
4.5	0.50
5.5	0.75
6.5	0.75
7.5	1.0



Que gagne-t-on par rapport à un krigeage ordinaire ?

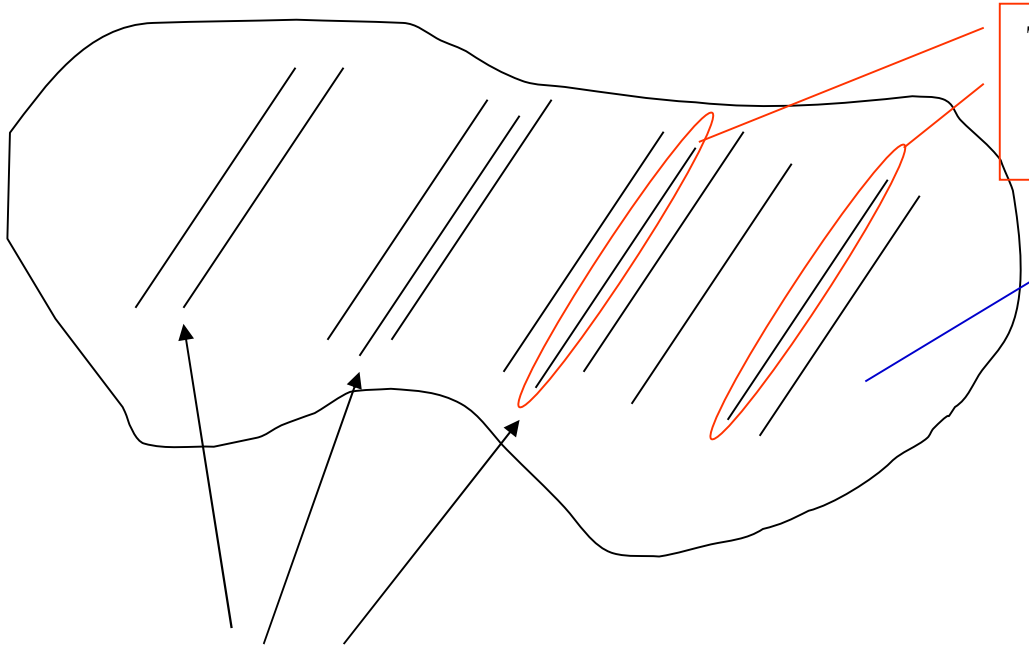
À quel prix ?

Quels sont les problèmes qui se posent ?

Que gagne-t-on par rapport à un krigeage ordinaire ?

- **Estimer une probabilité d'excéder un seuil (e.g. en environnement);**
- **estimer un quantile (e.g. valeur ayant 5% de chances d'être dépassée au point x0);**
- **calculer une variance conditionnelle, i.e. qui dépend des valeurs locales;**
- **fournir un estimateur qui minimise l'espérance d'une fonction de coût;**
- **+ de flexibilité :**
  - **utiliser des informations du type Z(xi)>t, Z(xi)<t, t2>Z(xi)>t1; des données semi-quantitatives fournies par le géologue (e.g. « dans ce type de roche, la teneur n'excède jamais « t »)**
  - **les variogrammes peuvent varier d'une indicatrice à l'autre**

Exemple: les variogrammes peuvent varier d'une indicatrice à l'autre

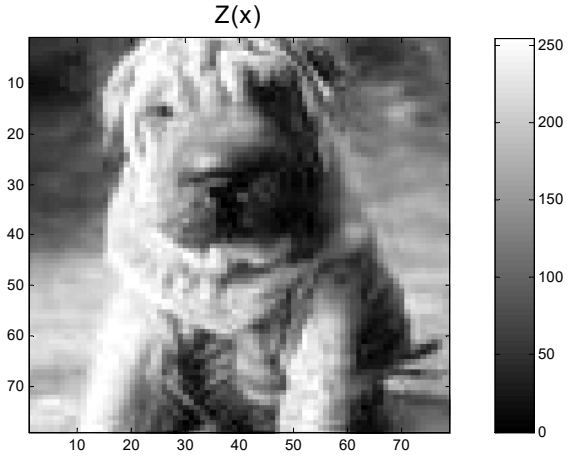


Teneurs riches, indicatrice pour seuil élevé montre une forte anisotropie

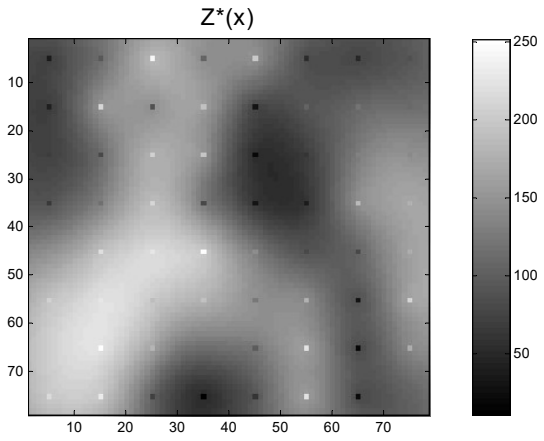
Teneurs faibles et moyennes, indicatrice pour seuils faibles et moyens montre une structure isotrope

Fractures tardives ayant permis un sur-enrichissement en Au

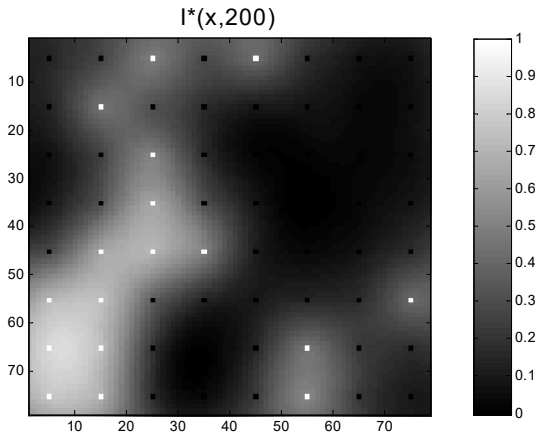
Exemple: le krigeage d'indicatrices permet de mieux estimer les quantiles, probabilités, etc. que le krigeage ordinaire



Réalité



Krigeage ordinaire



Krigeage indicatrice

Nb(Z(x)>200) = 1453

Nb(Z(x)*>200) = 561

sous-estimation de 61% !

Σ_x I*(x,200) = 1655

sur-estimation de 12%

Avec le KI on estime mieux la proportion de points, sur toute l'image, dépassant la valeur 200. Cependant on ne sait pas exactement lesquels parmi ces points sont au-dessus de ce seuil. Tout ce que l'on a, en chaque point, c'est une probabilité qu'il soit au-dessus du seuil (soit 1-I*(x)). Il serait évidemment logique de sélectionner les 1655 points ayant la plus forte probabilité de dépasser le seuil si l'on devait les identifier.

À quel prix ?

- **Fonction de répartition représentée sous forme discrète : combien de seuils ? (en pratique souvent de 5 à 10 seuils)**
- **Chaque seuil => v. indicatrice différente => variogramme => krigeage d'indicatrice => effort ++ important; possibilité d'incohérences dans la modélisation.**
- **Chaque indicatrice doit être stationnaire => fonction de répartition stationnaire, (hypothèse + forte que pour le krigeage).**
- **On a réduit l'ensemble conditionnant à { $I(x_1,c)$, $I(x_2,c)$…$I(x_n,c)$ } au lieu de { $Z(x_1)$, $Z(x_2)$…$Z(x_n)$ } => certaine perte d'information ?**

donc assez grossier

donc assez lourd

commun à toutes les méthodes non-linéaires et aux simulations

Un point Z2 situé tout près d'un point Z1 aura la même probabilité de dépasser disons 10 que l'on ait observé Z1=1000 ou Z1=11. D'autres méthodes non-linéaires reconnaitraient que cette probabilité devrait être supérieure avec Z1=1000 qu'avec Z1=11.

Quels sont les problèmes qui se posent ?

- **Problème de relation d'ordre:**

  -valeurs de $I^*(x_0,c_i)$ peuvent être >1 ou <0;

  - $I^*(x_0,c_i) > I^*(x_0,c_j)$ quand $c_i < c_j$

Une fonction de répartition doit être strictement non-décroissante, ce qui peut n'être pas le cas avec I*(x,c)

- **Variogrammes d'indicatrices sont souvent + faciles à modéliser (il n'y a pas de données extrêmes, que des 0 ou des 1) mais souvent la structure spatiale est faible => manque de précision dans les estimations de I***

voir diapos 18-19

- **Comment interpoler entre les valeurs de $I^*(x_0,c_i)$? Comment extrapoler au-delà de c_{min} et c_{max} ?**

- **Que faire si l'estimation doit porter sur des blocs ? Ex.: $P^*(Zv(x)>c|(n))$**

Problème de relation d'ordre

- **Problème de relation d'ordre:**
  - **$I^{**}(x0,c_i) = \max(0, I^*(x0,c_i))$**
  - **$I^{**}(x0,c_i) = \min(1, I^*(x0,c_i))$**
  - **Correction avant,**

$$-I^*_{avant}(x0,c_{i+1}) = \max(I^*(x0,c_i),I^*(x0,c_{i+1}))$$

  - **Correction arrière,**

$$- I^*_{arr}(x0,c_i) = \min(I^*(x0,c_i),I^*(x0,c_{i+1}))$$

- **$I^*(x0,c_i) = 0.5*[I^*_{avant}(x0,c_{i+1}) + I^*_{arr}(x0,c_i)]$**

Partant de la gauche, on refuse toute décroissance (correction avant).
Partant de la droite, on refuse toute croissance (correction arrière)

On prend la moyenne des deux fonctions précédentes.

Exemple

seuil c	$F_{KI}(x_0,c)$	$F_{KI,avant}(x_0,c)$	$F_{KI,arr}(x_0,c)$	$F_{KI,corr}(X_0,c)$
1	-.01 --> 0	0	0	0
2	0.13	0.13	0.13	0.13
3	0.24	0.24	0.234	0.237
4	0.238	0.24	0.234	0.237
5	0.234	0.24	0.234	0.237
6	0.237	0.24	0.237	0.2385
7	0.53	0.53	0.53	0.53
8	0.79	0.79	0.77	0.78
9	0.77	0.79	0.77	0.78
10	1.02 -> 1.0	1	1	1

Structure spatiale plus faible

Exemple : cas gaussien

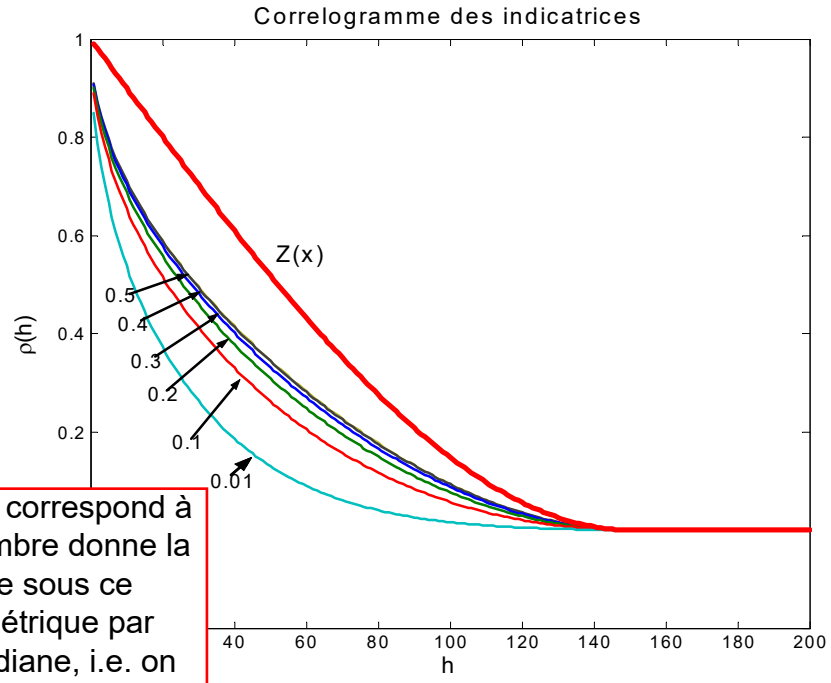
Si Z(x) est bigaussien (0,1), on connaît la relation entre $\gamma_z(h)$ et $\gamma_I(h,c)$

de moyenne 0 et de variance 1

$$C_I(h,c) = \frac{1}{2\pi} \int_0^{\rho(h)} \exp\left(-\frac{c^2}{1+u}\right) \frac{1}{\sqrt{1-u^2}}\, du$$

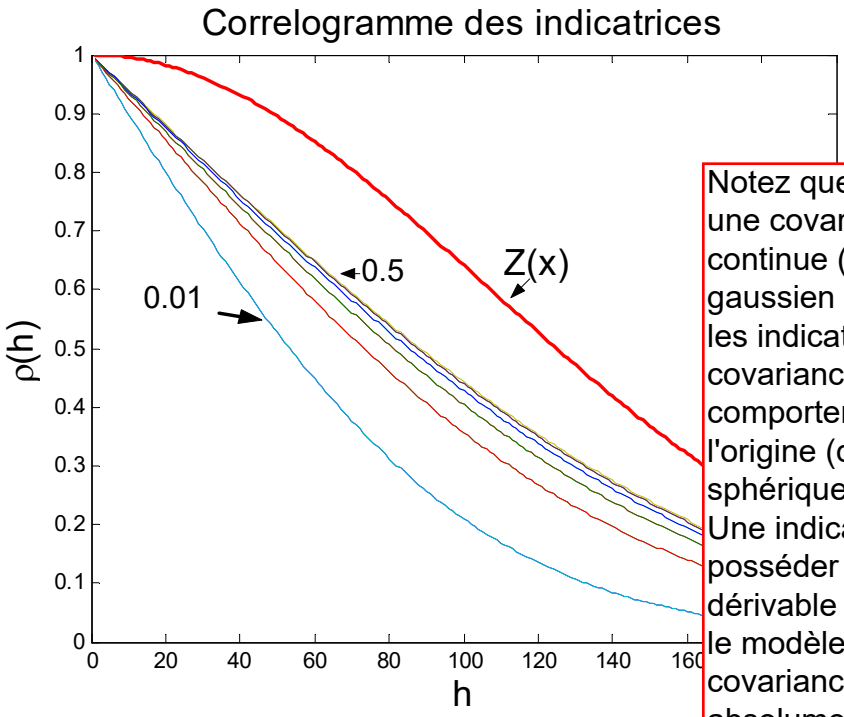
la covariance d'indicatrice (et donc le variogramme d'indicatrice) peut se calculer théoriquement à partir du corrélogramme (covariance) de la teneur.

ou ρ(h) est le corrélogramme de Z(x)



Chaque courbe correspond à un seuil. Le nombre donne la probabilité d'être sous ce seuil. C'est symétrique par rapport à la médiane, i.e. on aurait la même courbe avec p=0.95 qu'avec p=0.05, avec p=0.9 qu'avec p=0.1 et ainsi de suite.

x) : modèle sphérique



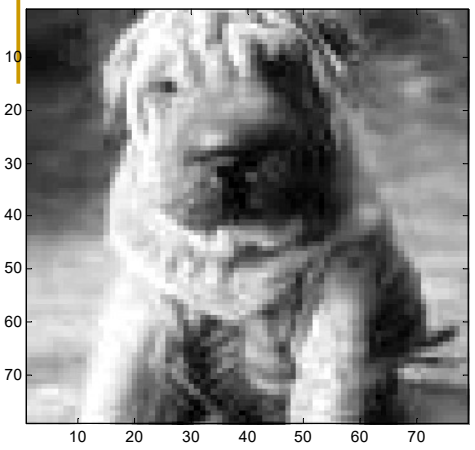
Notez que même si Z(x) a une covariance très continue (modèle gaussien de covariance), les indicatrices ont une covariance avec un comportement linéaire à l'origine (donc plutôt sphérique ou exponentiel) Une indicatrice ne peut posséder une covariance dérivable à l'origine, donc le modèle gaussien de covariance est à proscrire absolument pour les indicatrices.

Z(x) : modèle gaussien

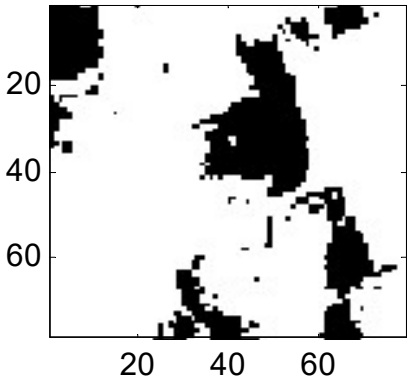
- Plus le seuil « c » est éloigné de la médiane moins il y a de structure

- Variogrammes des indicatrices est linéaire à l'origine même si Z(x) est parabolique à l'origine => proscrire le modèle gaussien pour les indicatrices

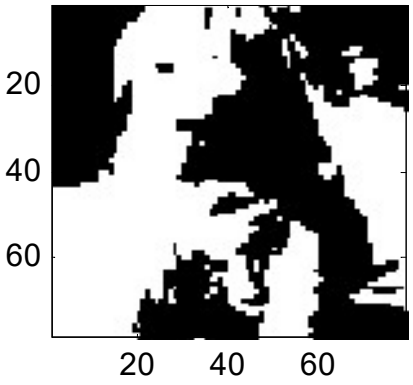
IMPORTANT: distinguer deux usages différents de l'adjectif gaussien qui n'ont rien à voir l'un avec l'autre. Un modèle de variogramme ou de covariance gaussien est un modèele très continu à l'origine (infiniment dérivable).
Un champ (ou domaine) gaussien veut dire que toutes les distributions multivariables suivent une loi multinormale. Ainsi deux points suivent une loi bigaussienne, trois points une loi trigaussienne, etc. On peut avoir un champ gaussien présentant une covariance effet de ppépite, sphérique, exponentielle, gaussienne, etc.



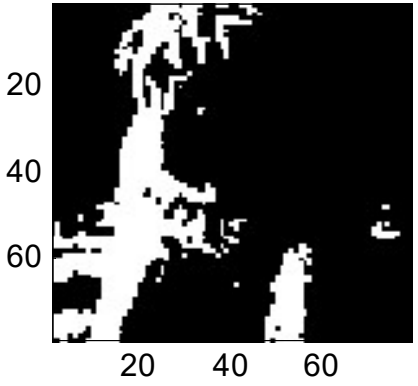
q20



q50

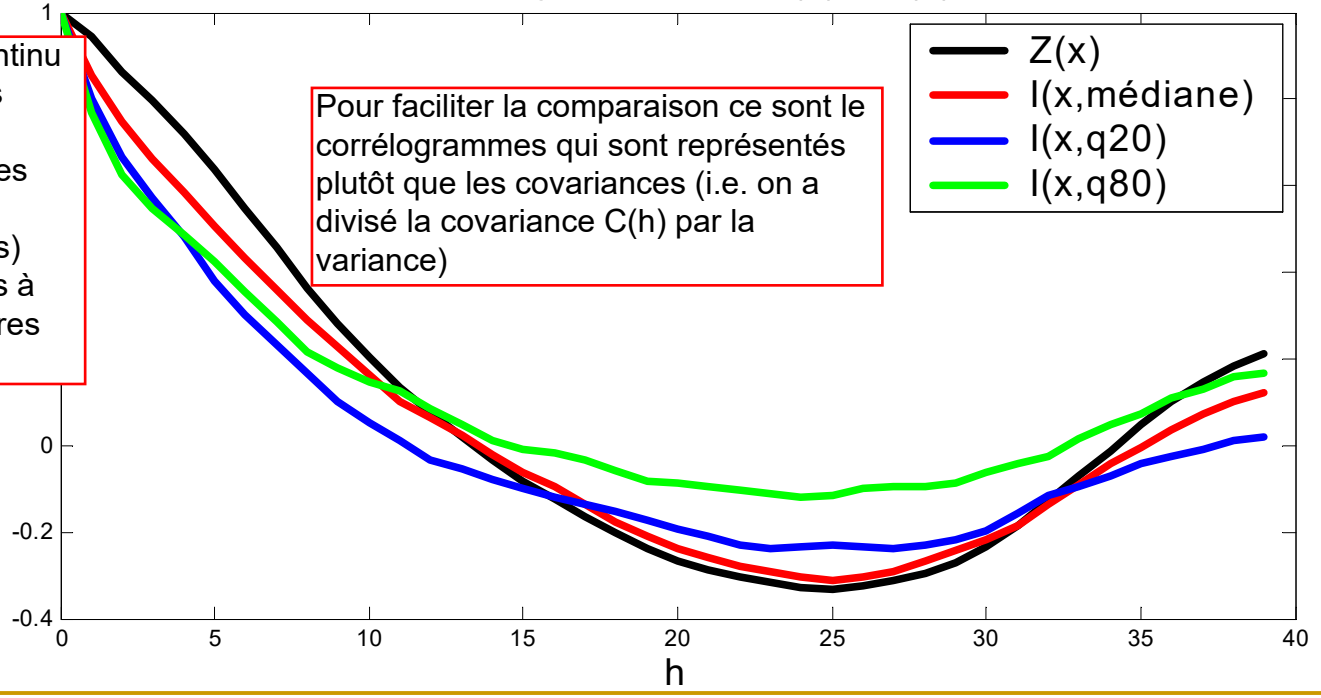


q80



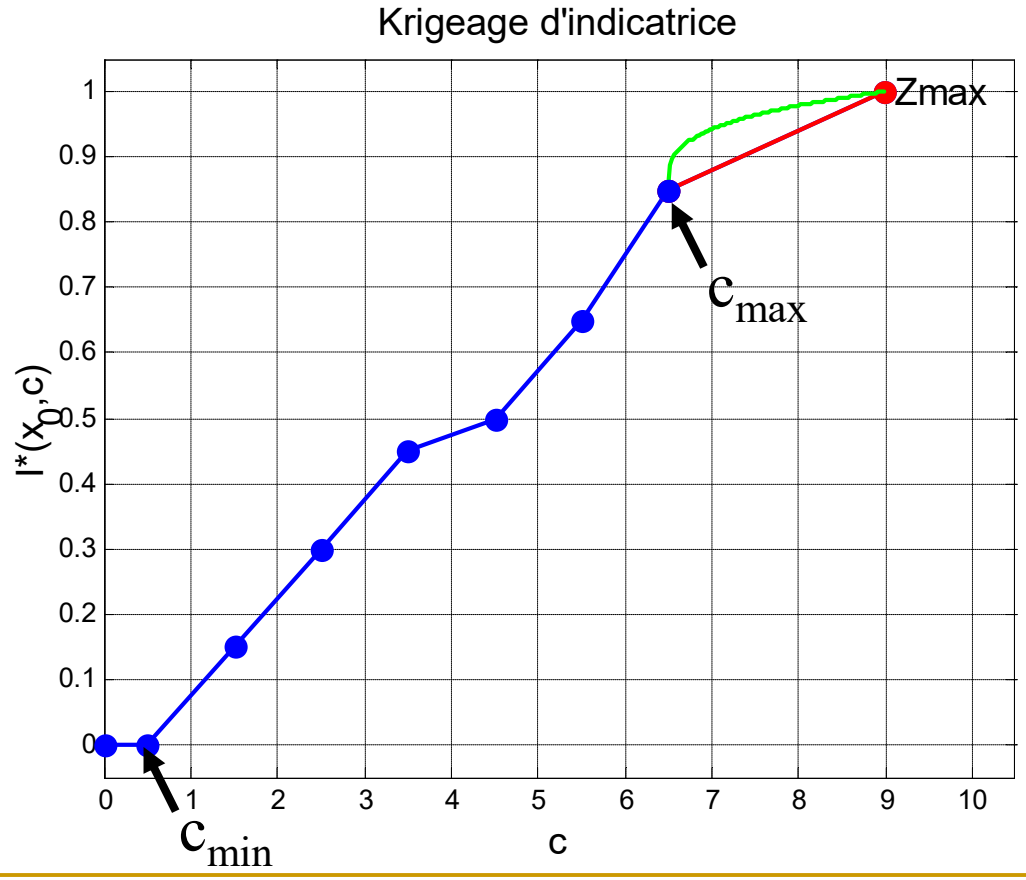
Comme prévu Z(x) est plus continu que les indicatrices à différents seuils.
Comme pour le cas gaussien les covariances d'indicatrices sont semblables (à petites distances) pour des seuils correspondants à des probabilités complémentaires (ici 0.2 et 0.8)

Pour faciliter la comparaison ce sont le corrélogrammes qui sont représentés plutôt que les covariances (i.e. on a divisé la covariance C(h) par la variance)



Interpolation entre les valeurs de $I^*(x_i,c_j)$?

Linéaire, sauf possiblement la dernière classe



Il faut décider du zmax que l'on considère raisonnable. Ce choix n'influence pas les estimations de I* mais influence toutefois assez fortement les calculs de teneur que l'on fait à partir de la fonction de répartition estimée.

Que faire si l'estimation doit porter sur des blocs ?

Reconnaître que:

$$P(Z_v(x)>c) \neq P(Z(x)>c)$$

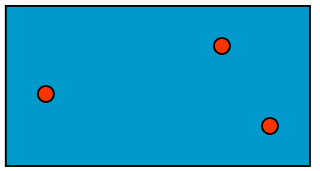
$$P(Z_v(x) > c) \neq \frac{1}{v}\int_{v(x)} P(Z(y) > c)\, dy$$

ce que l'on voudrait

n'est pas la moyenne des valeurs I* que l'on obtiendrait sur les points du bloc (contrairement au KO et KS par exemple où la teneur krigée du bloc est la teneur moyenne des teneurs krigées sur les points du bloc).

La teneur moyenne du bloc dépasse le seuil même si peu de points le dépassent

$$P(Z_v(x) > c) = 1$$



Z(x)=1e6 c

Z(x)= 0.9 c

Peu de points dépassent le seuil donc la moyenne des probabilitésdépasse à peine 0.
Donc ces deux quantités sont différents.

$$\frac{1}{v}\int_{v(x)} P(Z(y) > c)\, dy \approx 0$$

La fonction qui lie teneur à probabilité est non-linéaire. Ce n'est pas différent par exemple de: "la moyenne des logarithmes de teneurs n'est pas égale au logarithme de la teneur moyenne de ces teneurs".

Solutions ?

Plusieurs propositions, dont:

- correction affine
- correction indirecte lognormale

Selon moi ces solutions n'en sont pas vraiment.

aucune n'est entièrement convaincante

Tendance actuelle: recourir à des simulations!

Exemple: correction affine

$$F_v(Z_v) = F\left((Z-m)\left\{ \frac{D^2(v \mid G)}{D^2(\bullet \mid G)} \right\}^{0.5} + m \right)$$

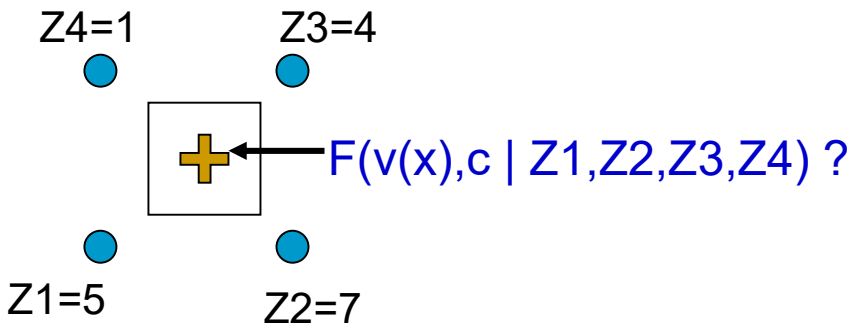
on contracte la distribution d'un facteur fixe déterminé par le ratio des variances de dispersion de blocs sur les variances ponctuelles

m est la moyenne de la distribution locale estimée par KI

F est la fonction de répartition locale estimée par KI (i.e. I*(x,c) après corrections pour relations d'ordre)

F_v est la fonction de répartition « de blocs »

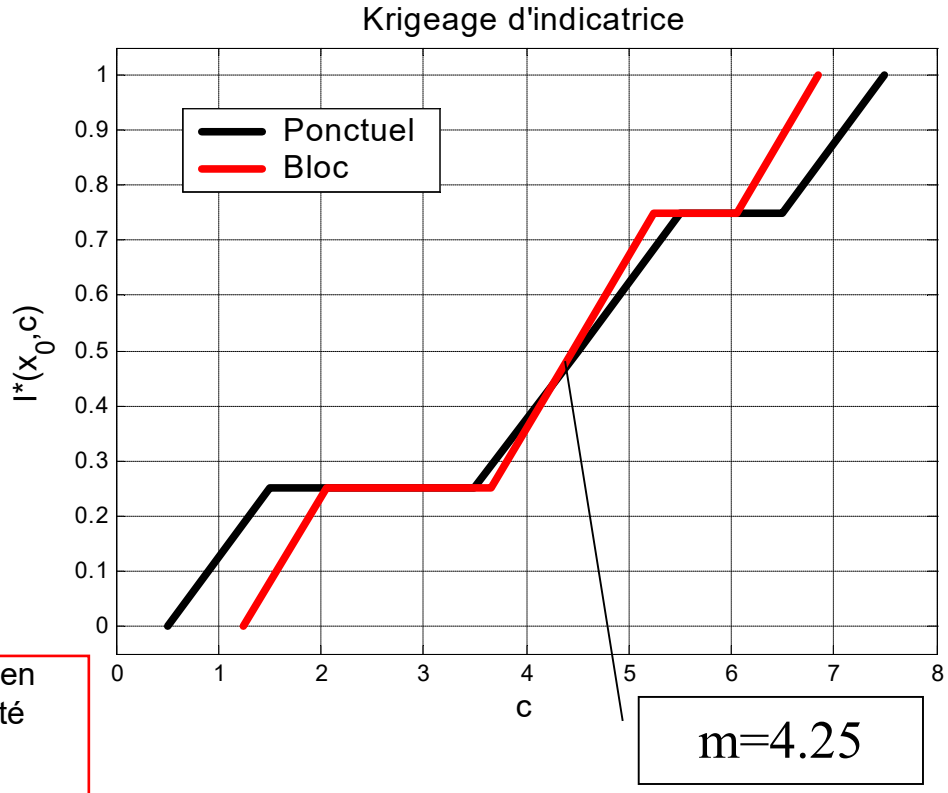
Le facteur de contraction $\left\{ \frac{D^2(v \mid G)}{D^2(\bullet \mid G)} \right\}^{0.5}$ est *global* malgré que la correction soit appliquée à une *distribution locale* !



Variogramme => $D^2(.|G)$

=> $D^2(v|G)$

$$\left\{\frac{D^2(v\,|\,G)}{D^2(\bullet\,|\,G)}\right\}^{0.5} = 0.8$$



Notez comment la courbe en rouge donne une probabilité nulle que la teneur soit inférieure à 1.2 environ ou supérieure à 6.9. Est-ce réaliste?

le point charnière de la correction affine est la moyenne "m" obtenue avec la distribution estimée

Variantes

au krigeage d'indicatrices en vue d'améliorer ses performances.

1- Krigeage simple d'indicatrice :

$$I^*(x_0, c) = \sum_{i=1}^{n} \lambda_i \, I(x_i, c) + \left(1 - \sum_{i=1}^{n} \lambda_i\right) F_Z(c)$$

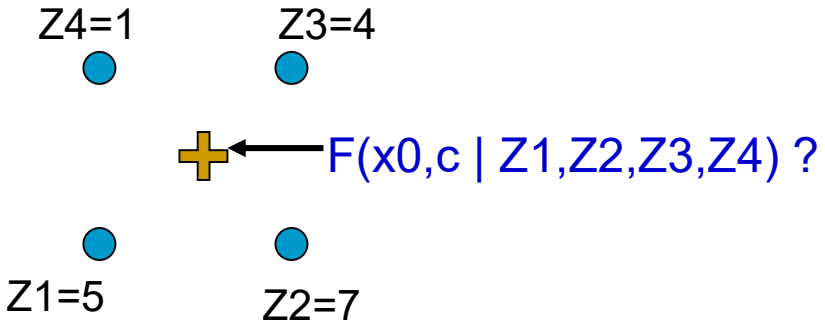
local

global (non-localisé)

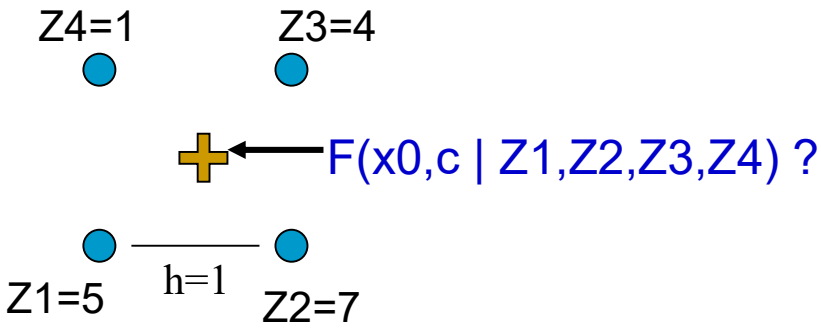
$F_Z(c)$: fonction de répartition (globale)

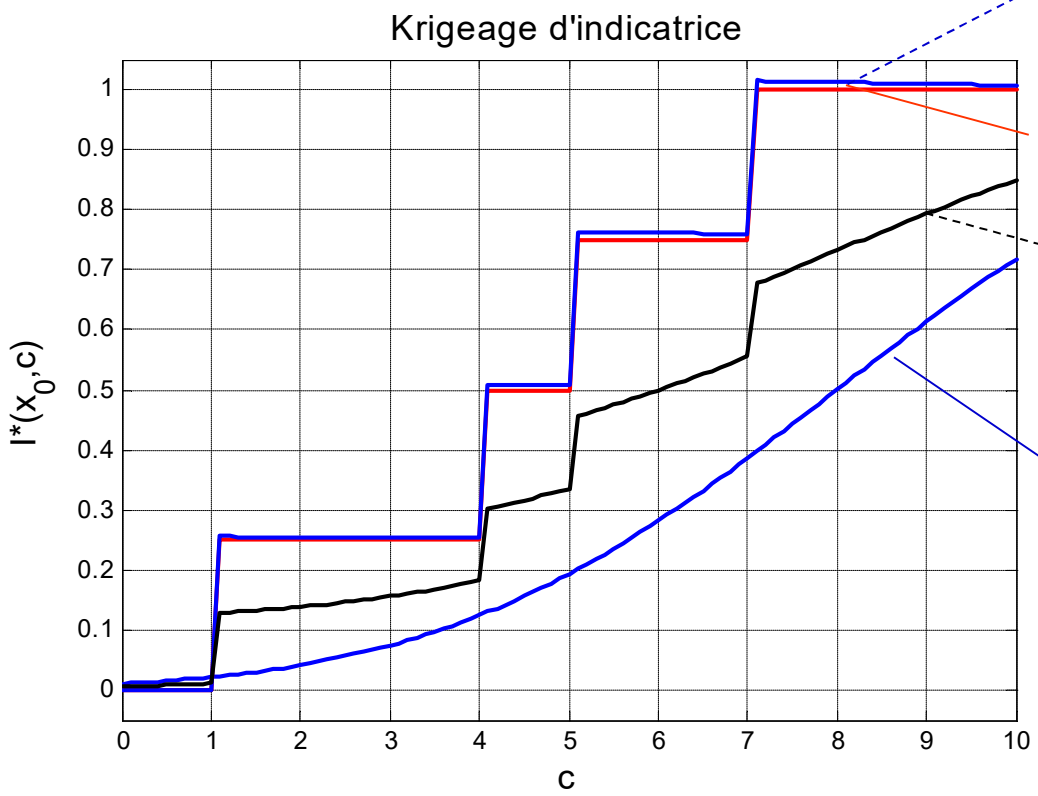
Permet une gradation plus souple de I*(x,c)

Permet de mieux tenir compte du degré de corrélation locale



S'il n'y a pas de corrélation entre les points, la fonction estimée sera simplement $F_z(c)$





I* par KS, sphérique a=10

fortes corrélations, KS rejoint KO

I* par K0

faible corrélation KS est intermédiaire entre KO et information globale

I* par KS, sphérique a=1

Fct. de répartition N(8,12)
I* par KS si $a<1/2^{0.5}$

information globale

2- Cokrigeage :

v. principale : $I(x,c_j)$

v. secondaires : $I(x,c_k)$, $k \neq j$

Très lourd, presque jamais utilisé

ou

v. principale : $I(x,c_j)$

v. secondaire : $Z(x)$ ou mieux $U(x) = \text{rang}(Z(x)) / (n+1)$

connu aussi sous le nom de probability kriging. Un bon compromis entre le KI et le cokrigeage de toutes les indicatrices.

« Soft kriging »

-+ de flexibilité :

on ne sait pas trop comment utiliser ces informations en KO et en KS

- utiliser des informations du type Z(xi)>t, Z(xi)<t, t2>Z(xi)>t1;

des données semi-quantitatives fournies par le géologue (e.g. « dans ce type de roche, la teneur n'excède jamais « t »)

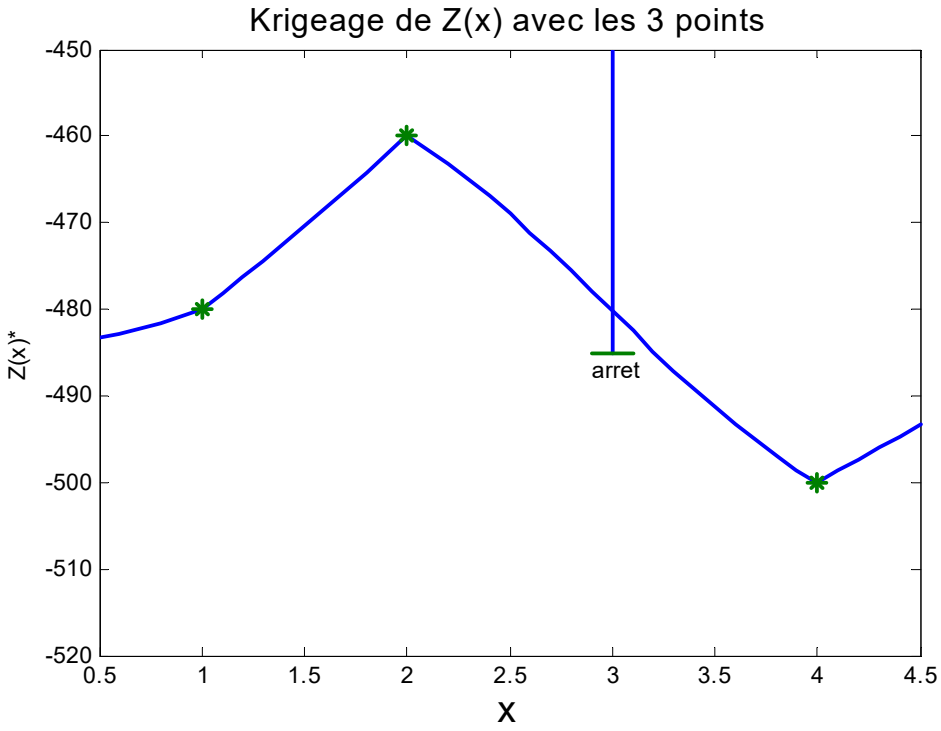
Exemple :

3 forages ont intercepté le sommet d'un réservoir pétrolier

Z(1) = -480, Z(2) = -460, Z(4) = -500

un 4e forage situé en x=3, a dû être arrêté au niveau –485 sans que le sommet n'ait été intercepté !

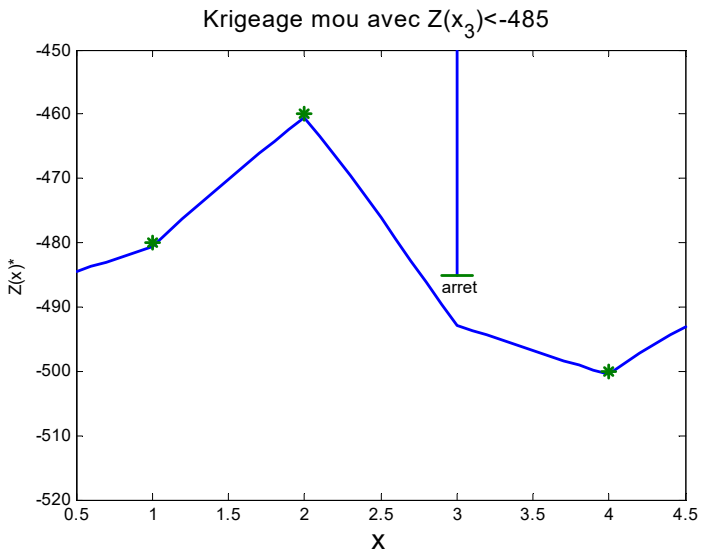
Solution par KO de Z(x) ≈ solution par KO de I(x) avec les 3 forages (1,2,4)

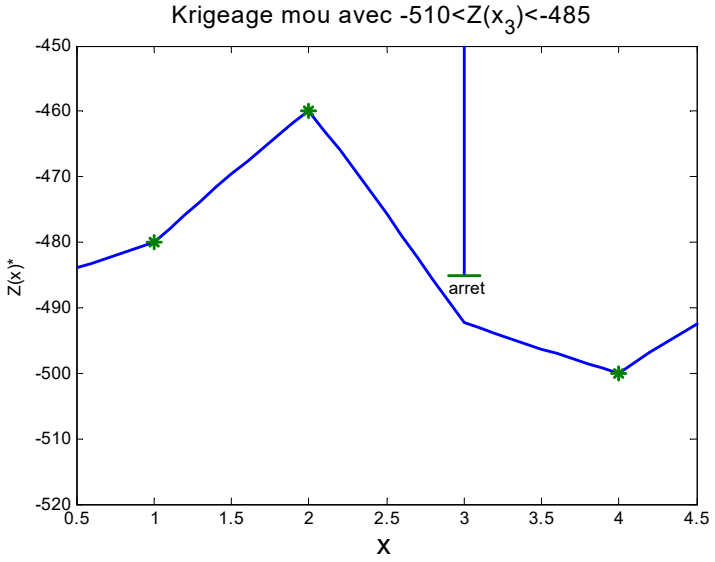


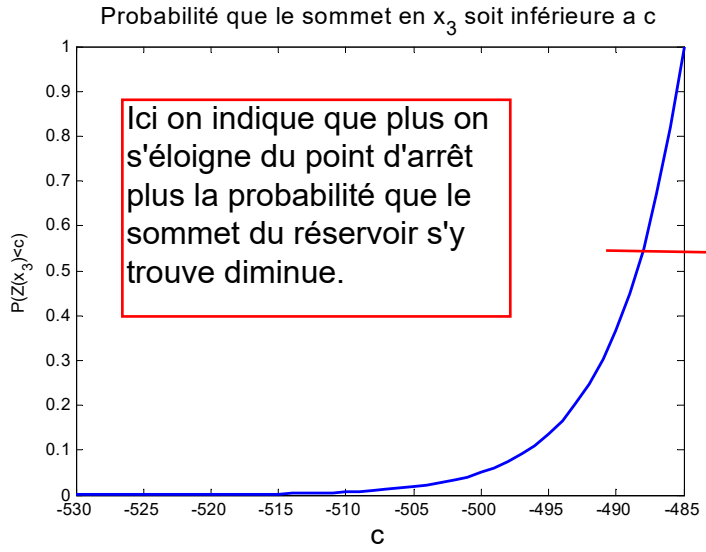
ici la valeur krigée par KO de l'élévation contrdit l'information au forage qui a été arrêté.

La solution n'est pas acceptable ! Elle contredit l'information en x=3.

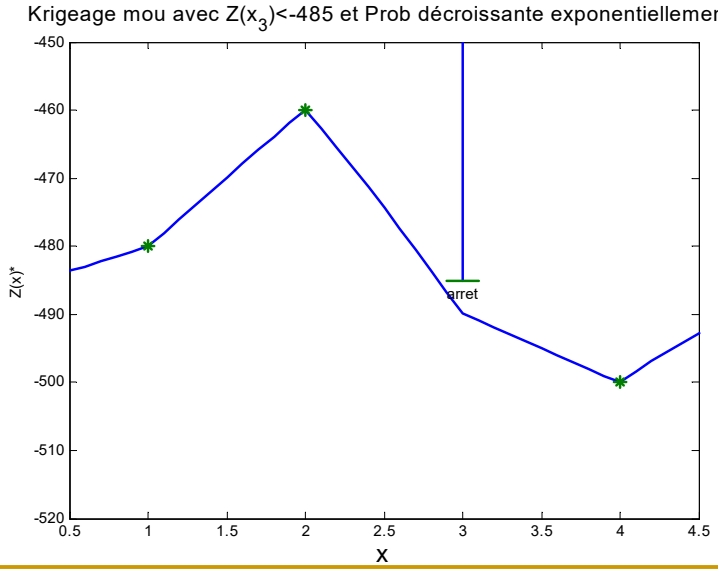
Trois indications concernant l'élévation au forage arrêté. En haut à gauche: le sommet est plus profond que le point d'arrêt. En haut à droite: le sommet est plus profond que le point d'arrêt mais devrait survenir avant -510 m. En bas èa droite, le sommet devrait survenir avec probabilité décroissante (au fur et à mesure que l'on descend) décrite par la fonction en bas à gauche.







Ici on indique que plus on s'éloigne du point d'arrêt plus la probabilité que le sommet du réservoir s'y trouve diminue.



Le résultat est assez insensible à la façon dont on représente l'information au sujet du forage arrêté. Le point important est que le sommet n'a pas été rencontré avant -485

Remarques

Soit le seuil « c » correspondant à un quantile « p » de la distribution de Z(x)

I(x,c) a les propriétés suivantes:

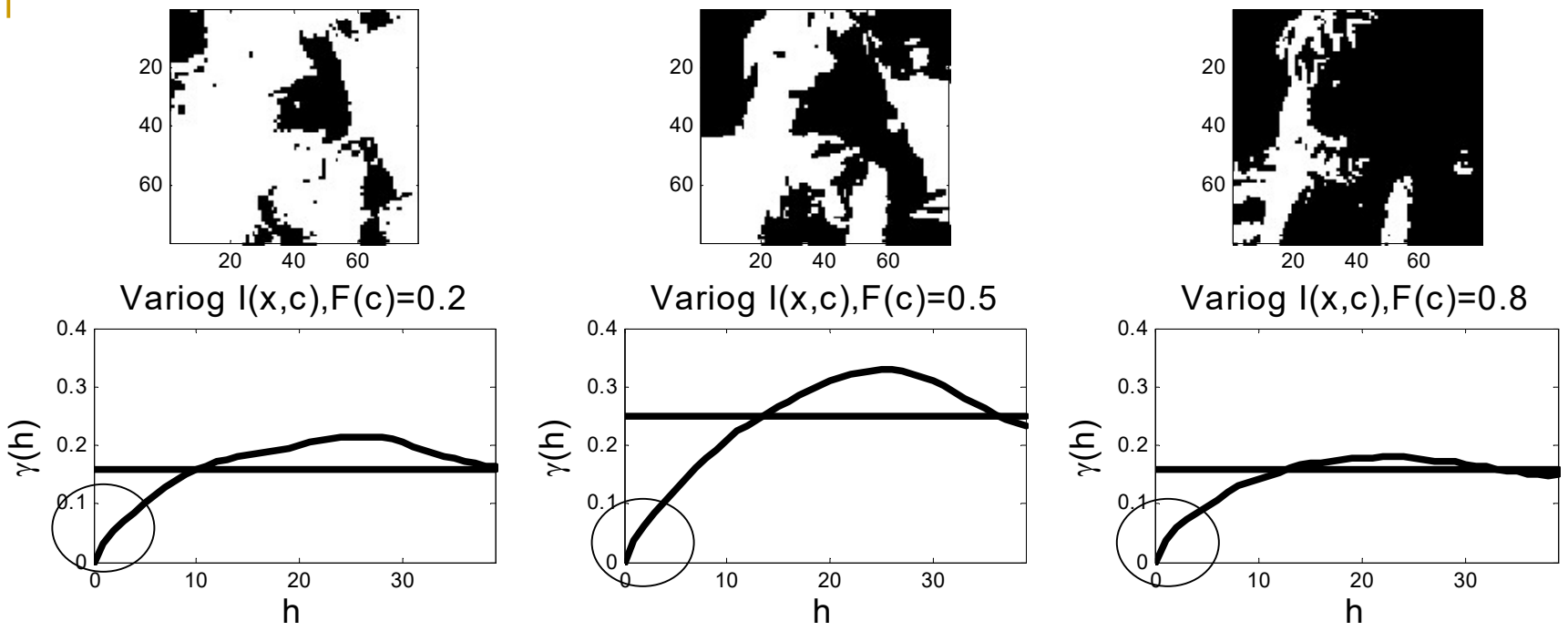
$$E[I(x,c)] = p$$

$$Var(I(x,c)) = p(1-p)$$

Ex.: si l'on choisit un seuil pour lequel 20% des observations sont inférieures, $D^2(I(x,c)|G) \approx$ **0.2*0.8 = 0.16**

normalement, le palier est légèrement supérieur à $D^2(I(x,c)|G)$ **(dépendant de l'importance de la structure spatiale).**

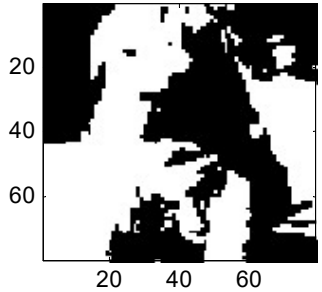
⇒ Les paliers sont presque déterminés par le seuil du codage de l'indicatrice

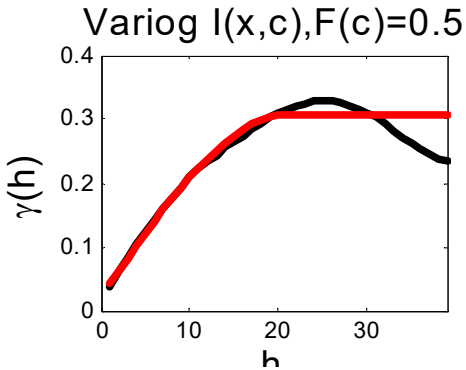


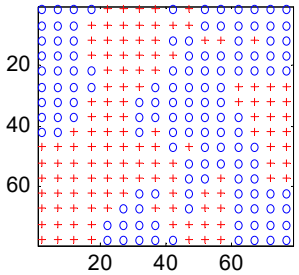
Les variogrammes d'indicatrices ne peuvent montrer un comportement parabolique à l'origine => proscrire le modèle gaussien ! de variogramme

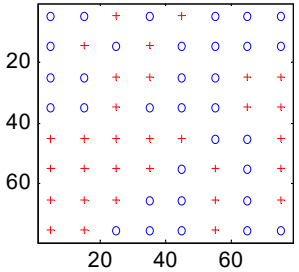
Exemple : déterminer le volume d'un sol contaminé au delà d'une norme

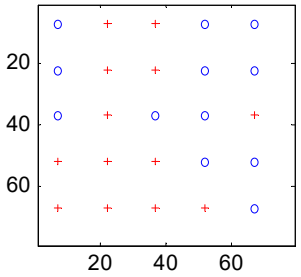
Cet exemple montre le volume estimé au dessus d'un seuil (et les écarts-types de Ki) sur l'ensemble de la zone et pour quatre scénarios d'échantillonnage.
On voit que les volumes estimés varient très peu ici. Ce sont les intervalles de confiance qui se rétrécissent au fur et à mesure que le nombre de données augmente et donc que l'incertitude diminue.

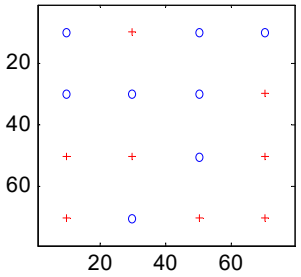












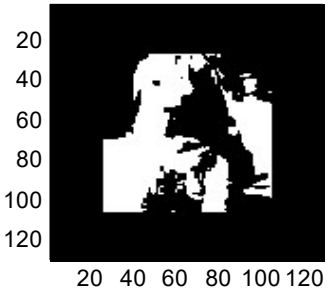
C=130; V=3120

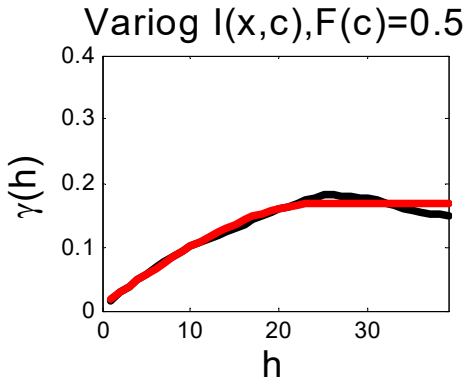
Pas	n	V*	σ	CI (95%)
5	256	3020	93	[2834,3205]
10	64	3023	210	[2603,3443]
15	25	3002	387	[2228,3775]
20	16	3089	527	[2036,4142]

avec n=16, le volume pourrait varier du simple au double (2000 è 4000). Si l'on juge que c,est trop incertain on a intérêt à augmenter l'échantillonnage.

dans la diapo précédente on avait comme idnetifié au départ les limtes de la zone d'intérêt. Souvent on est à la recherche de ces limites (e.g. contamination sur un site mais on ne sait pas exactement où cela a pu se produire).

Que se passe-t-il si l'échantillon déborde de la zone d'intérêt ?





Le variogramme des indicatrices change

	Surface>c		Écart-type	
Pas	Sans	Avec	Sans	Avec
5	3020	3096	93	132
10	3023	3110	210	230
15	3002	4115	387	403
20	3089	3267	527	577

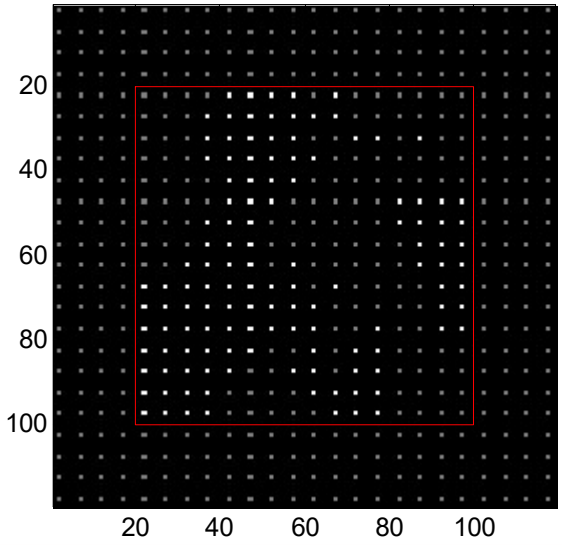
on obtient des volumes estimés semblables (sauf pour le cas du pas de 15) et des écarts-types similaires.

Les estimés sont semblables et l'ordre de grandeur des écarts-types est comparable, même si la zone couverte est 2.2 fois + grande en superficie

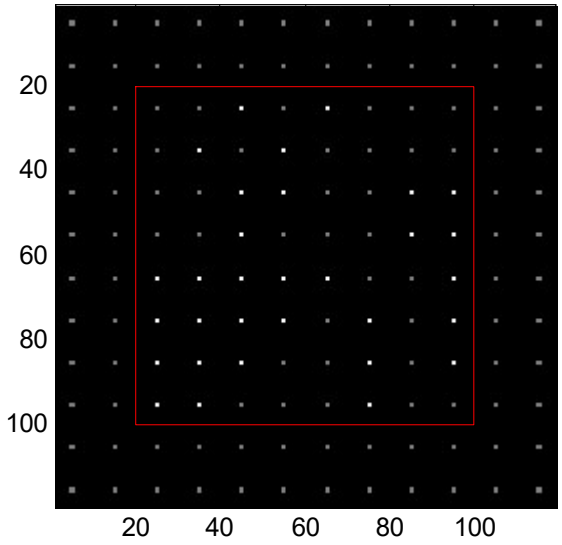
⇩

Bonne robustesse au choix initial de la zone d'étude

as=5

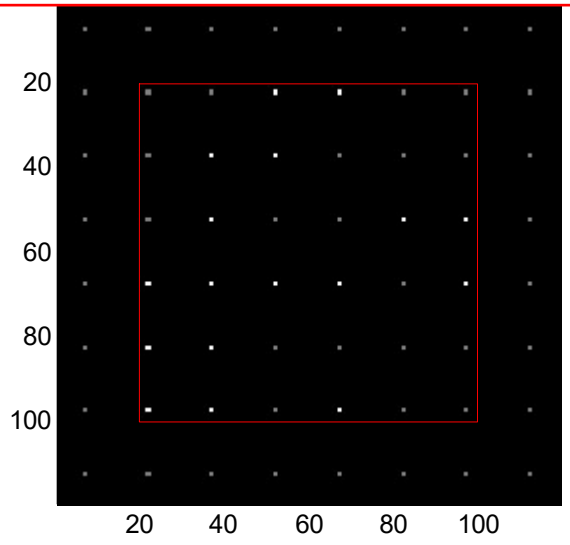


Pas=10



avec le pas de 15 on a proportionnellement plus de points de l'échantillon dans la zone originale, ce qui explique la surestimation du volume. Un échantillonnage aléatoire stratifié aurait permis d'éviter cela.

Pas=15



Pas=20