

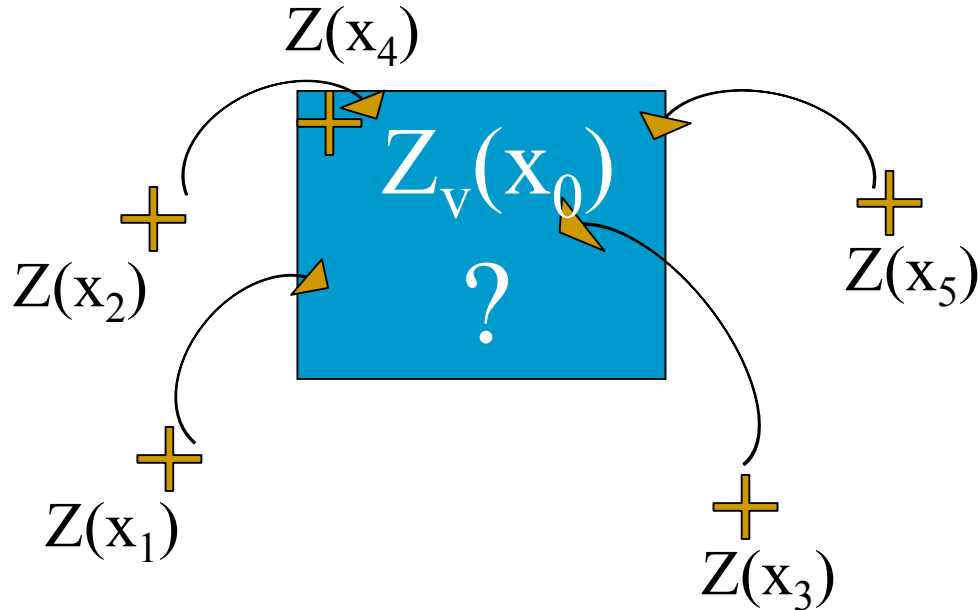
Variance d'estimation

Automne 2003

Plan

- Variance d'estimation : définition
- Les 3 composantes essentielles de $\text{Var}(e)$
- Interprétation
- Calcul
- Abaques
- Principe de combinaison d'erreurs élémentaires
- Précision d'une estimation globale
- Applications

Variance d'estimation



$$Z_v^* = \sum_{i=1}^n \lambda_i Z_i$$

$$e = Z_v - Z_v^*$$

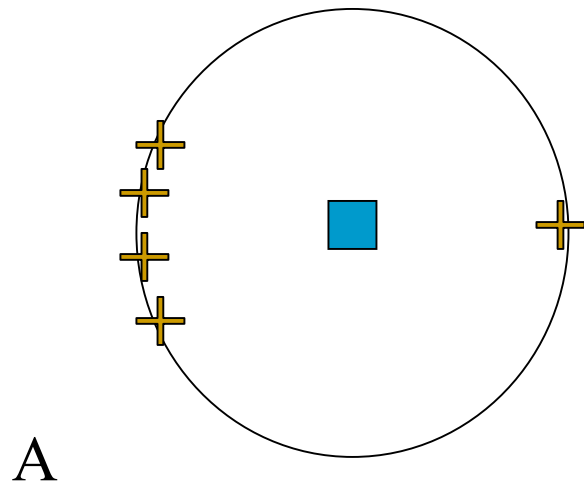
$$\text{Var. d'estimation} = \text{Var}(e)$$

$$Var(e) = Var(Z_v) + Var(Z_v^*) - 2Cov(Z_v, Z_v^*)$$

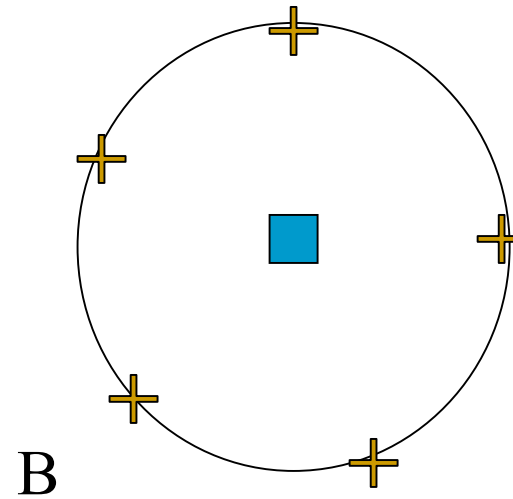
$$\sigma_e^2 = \underbrace{Var(Z_v)}_a + \underbrace{\sum_i \sum_j \lambda_i \lambda_j Cov(Z_i, Z_j)}_b - 2 \underbrace{\sum_i \lambda_i Cov(Z_i, Z_v)}_c$$

- a) Ce que l'on cherche à estimer est-il foncièrement variable ou non ?
- b) Quel est le degré de redondance entre les observations ?
- c) Les observations sont-elles bien placées par rapport à ce que l'on veut estimer ?

Exemple



VS



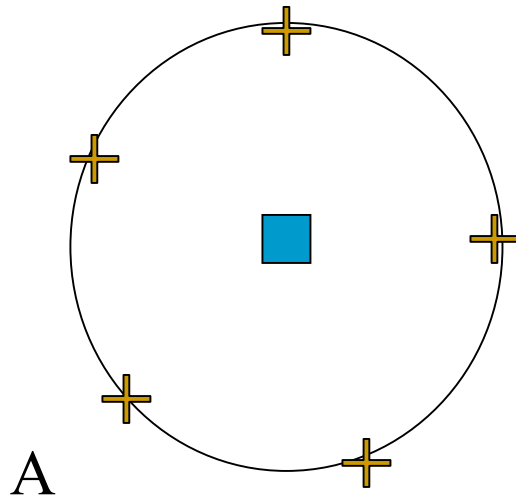
a) Le bloc à estimer est le même

b) La redondance est beaucoup plus forte en A

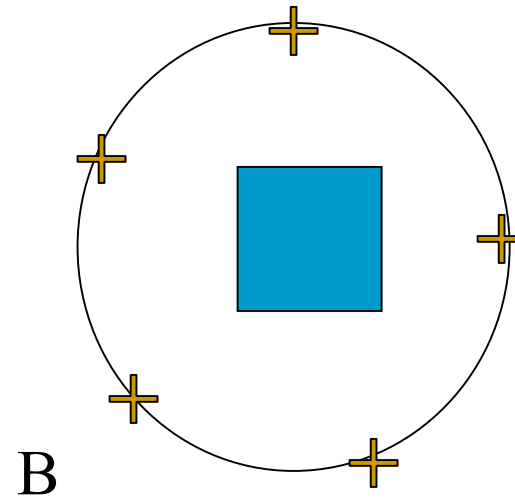
c) La position par rapport au bloc est la même

} $\text{Var}(e)$ + grande en A

Exemple



VS



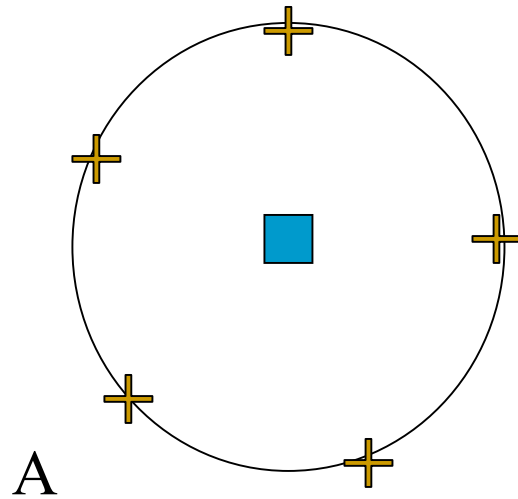
a) **Le bloc à estimer en B est moins variable**

b) La redondance est la même

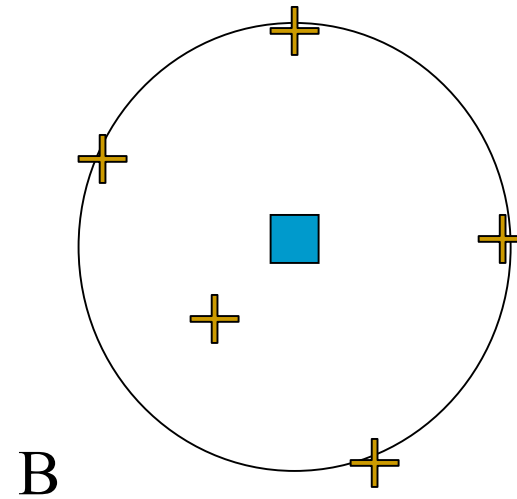
c) La position par rapport au bloc est
légèrement + favorable en B

} $\text{Var}(e)$ + grande en A

Exemple



VS



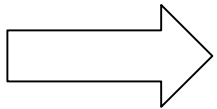
- a) Le bloc à estimer est le même
 - b) La redondance est un peu plus forte en B
 - c) La position par rapport au bloc est plus favorable en B
- } **Var(e) + grande en A**

En terme du variogramme

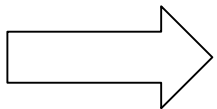
$$\sigma_e^2 = \text{Var}(Z_v) + \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z_i, Z_j) - 2 \sum_i \lambda_i \text{Cov}(Z_i, Z_v)$$

Si $\sum \lambda_i = 1$ (comme c'est le cas pour la plupart des estimateurs)

$$\sigma_e^2 = -\bar{\gamma}(v, v) - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i, x_j) + 2 \sum_i \lambda_i \bar{\gamma}(x_i, v)$$



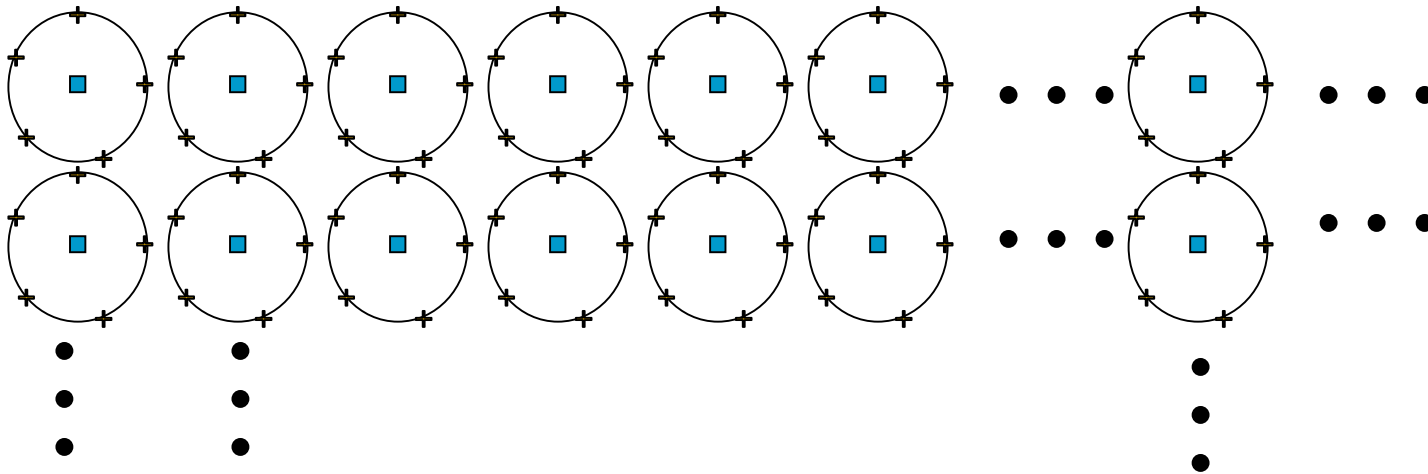
On peut calculer $\text{Var}(e)$ dès que l'on connaît le variogramme
Il n'est pas nécessaire que le variogramme montre un palier



$\text{Var}(e)$ ne dépend pas des valeurs observées, il ne dépend que de la géométrie du problème et du variogramme

Interprétation de $\text{Var}(e)$

- Ne dépend pas des valeurs \Rightarrow mesure non-conditionnelle
- Représente la variance des erreurs que l'on obtiendrait si l'on faisait glisser le bloc et la configuration de données sur un domaine très grand



Calcul de la variance d'estimation

- Numériquement surtout : « v » représenté par une grille de points (régulière ou aléatoire)
- Abaques pour des cas très simples, surtout utiles pour estimation globale (principe de combinaison des erreurs élémentaires)

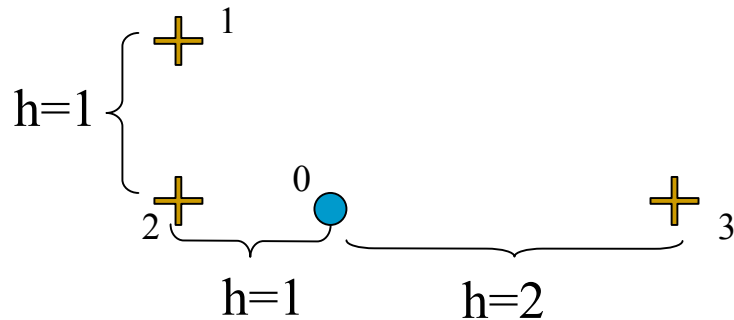
Note importante:

On peut calculer $\text{Var}(e)$ pour tout estimateur linéaire :

- méthode polygonale
- triangles
- inverse de la distance (pour tout coefficient « b »)
- krigeage

=> On a un outil pour juger de la précision d'une méthode vs une autre pour un gisement donné !

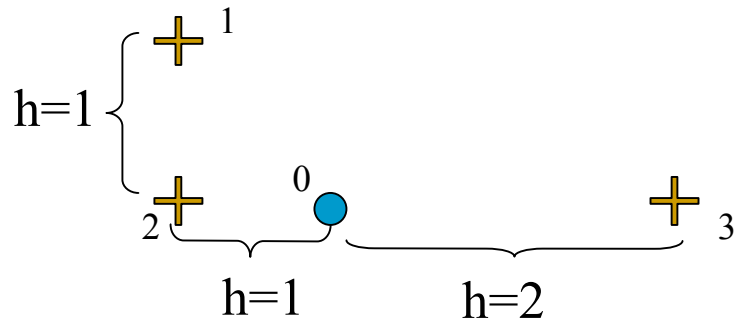
Exemple



Variogramme linéaire, pente=1, $C_0=1$

Méthode	Poids	Var(e)
Polygone	$l_1=0; l_2=1; l_3=0$	4.0
Inverse de la distance (b=1)	$l_1=0.32; l_2=0.45; l_3=0.23$	2.72
Inverse de la distance (b=2)	$l_1=0.29; l_2=0.57; l_3=0.14$	2.88
Triangle (méthode des %)	$l_1=0; l_2=2/3; l_3=1/3$	2.89
Krigeage	$l_1=0.25; l_2=0.43; l_3=0.32$	2.65

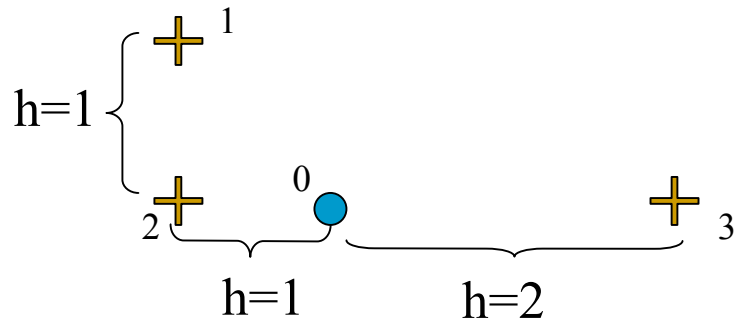
Exemple



Variogr. sphérique $a=10$, $C=1$ $C_0=0$

Méthode	Poids	Var(e)
Polygone	$l_1=0; l_2=1; l_3=0$	0.30
Inverse de la distance ($b=1$)	$l_1=0.32; l_2=0.45; l_3=0.23$	0.21
Inverse de la distance ($b=2$)	$l_1=0.29; l_2=0.57; l_3=0.14$	0.22
Triangle (méthode des %)	$l_1=0; l_2=2/3; l_3=1/3$	0.20
Krigeage	$l_1=0.16; l_2=0.53; l_3=0.31$	0.20

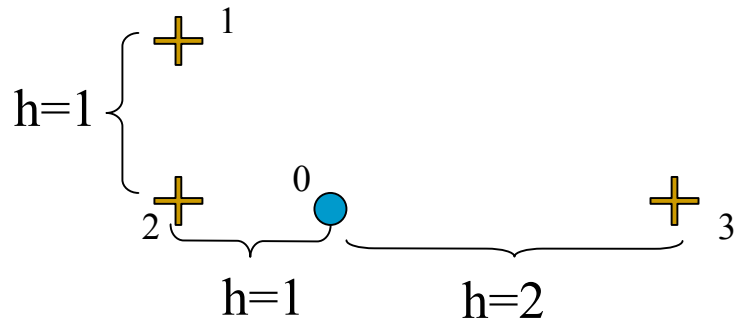
Exemple



Effet de pépité pur, $C_0=1$

Méthode	Poids	Var(e)
Polygone	$l_1=0; l_2=1; l_3=0$	2
Inverse de la distance (b=1)	$l_1=0.32; l_2=0.45; l_3=0.23$	1.36
Inverse de la distance (b=2)	$l_1=0.29; l_2=0.57; l_3=0.14$	1.43
Triangle (méthode des %)	$l_1=0; l_2=2/3; l_3=1/3$	1.56
Krigeage	$l_1=1/3; l_2=1/3; l_3=1/3$	1.33

Exemple



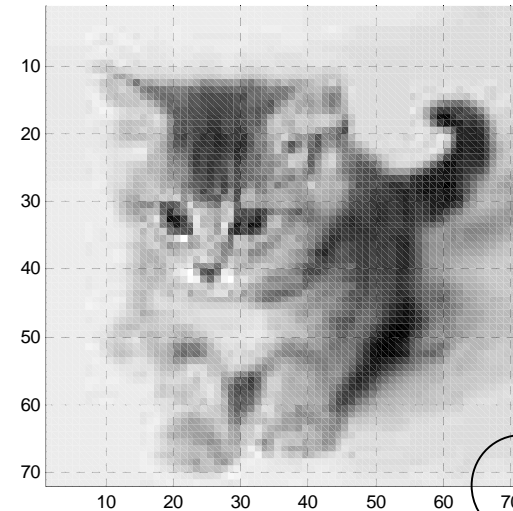
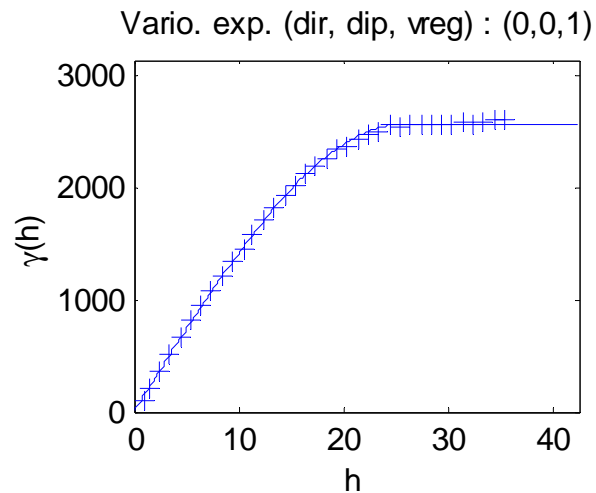
Gaussien $a=2$, $C=1$;

Méthode	Poids	Var(e)
Polygone	$l_1=0$; $l_2=1$; $l_3=0$	0.44
Inverse de la distance ($b=1$)	$l_1=0.32$; $l_2=0.45$; $l_3=0.23$	0.36
Inverse de la distance ($b=2$)	$l_1=0.29$; $l_2=0.57$; $l_3=0.14$	0.37
Triangle (méthode des %)	$l_1=0$; $l_2=2/3$; $l_3=1/3$	0.32
Krigeage	$l_1=-0.01$; $l_2=0.74$; $l_3=0.27$	0.31

L'exemple montre:

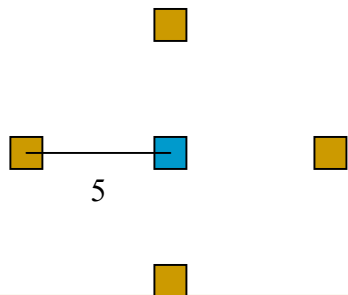
- Le krigeage est toujours la méthode présentant $\text{Var}(e)$ minimale
- Le krigeage est la seule méthode où les poids varient selon le variogramme
- La méthode polygonale présente toujours $\text{Var}(e)$ maximale
- Les autres méthodes ont un rang qui varie selon le variogramme postulé

Exemple



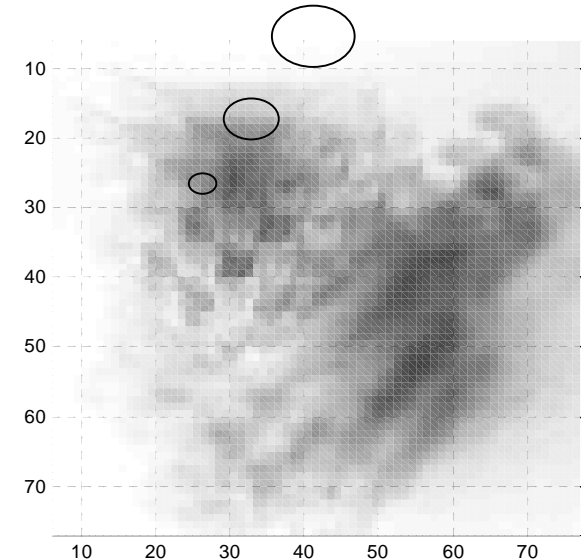
Je me
sens tout
chose

Sphérique isotrope, $C_0=0$, $C=2568$ $a=25.95$



$\text{Var}(e)$ théorique = 601

$\text{Var}(e)$ expérimental = 639



Utilisation des abaques

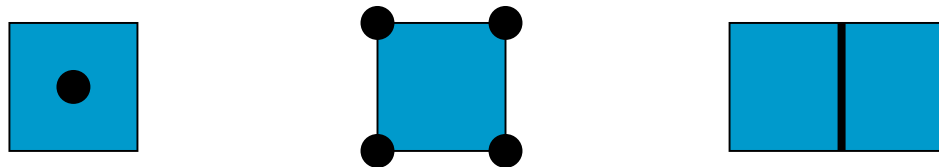
Seules des configurations très simples sont répertoriées

Supposent le variogramme ponctuel connu

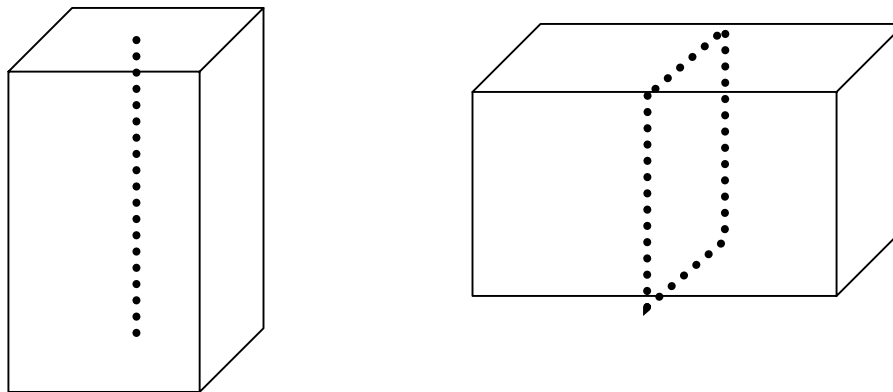
1D



2D



3D



Variances
d'extension

Effet de pépité

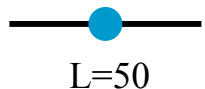
Si le support des données est quasi-ponctuel :

remplacer C_0 par une composante sphérique de portée « epsilon »
utiliser l'abaque avec cette composante

Si le support des données n'est pas quasi-ponctuel

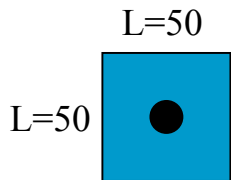
la contribution à $\text{Var}(e)$ est C_0 / n où « n » est le nombre de fois le support d'observation entre dans le support d'extension

Exemple : variogramme sphérique ponctuel $C_0=20$, $C=40$, $a=100$



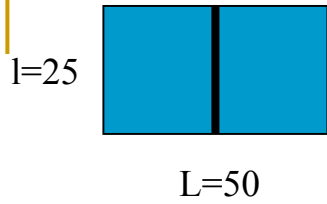
Étendre un point à un segment (abaque 7, courbe E1)

$$\text{Var}(e) = 40 * E1(0.5) + 20 * E1(\infty) = 40 * 0.13 + 20 * 1 = 25.2$$



Étendre un point à un carré (abaque 7, courbe E3)

$$\text{Var}(e) = 40 * E3(0.5) + 20 * E3(\infty) = 40 * 0.19 + 20 * 1 = 27.6$$



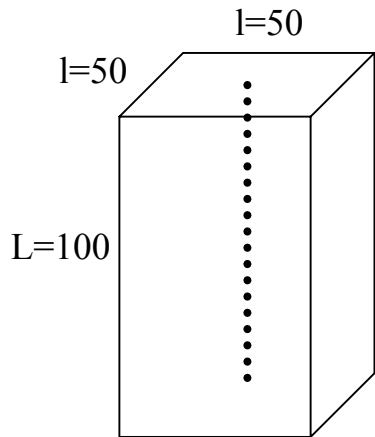
Étendre un segment à un rectangle (abaque 8)

Si ponctuel:

$$\text{Var}(e) = 40 \cdot E(0.5, 0.25) + 20 \cdot E(\infty, \infty) = 40 \cdot 0.065 + 20 \cdot 0 = 2.6$$

Si carottes de 5m

$$\text{Var}(e) = 40 \cdot E(0.5, 0.25) + 20/5 = 40 \cdot 0.065 + 4 = 6.6$$



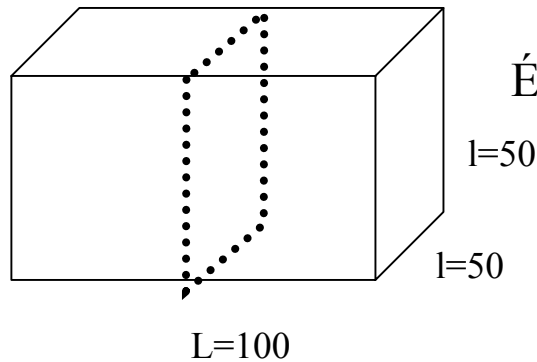
Étendre un segment à un parallélépipède à section carrée (abaque 10)

Si ponctuel:

$$\text{Var}(e) = 40 \cdot E(0.5, 1) + 20 \cdot E(\infty, \infty) = 40 \cdot 0.047 + 20 \cdot 0 = 1.88$$

Si carottes de 5m

$$\text{Var}(e) = 40 \cdot E(0.5, 1) + 20/20 = 40 \cdot 0.047 + 1 = 2.88$$



Étendre un carré à un parallélépipède (abaque 9)

Si ponctuel:

$$\text{Var}(e) = 40 \cdot E(1, 0.5) + 20 \cdot E(\infty, \infty) = 40 \cdot 0.1 + 20 \cdot 0 = 4.0$$

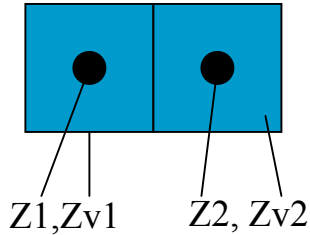
Si carottes de 5m

$$\text{Var}(e) = 40 \cdot E(1, 0.5) + 20 / \infty = 4.0$$

- Attention ! Toutes les abaques 2D et 3D sont non-symétriques; il faut entrer les bonnes valeurs en abscisse et en ordonnée.

- Deux côtés égaux, => en fait 2 rapports $L_x/a_x, L_y/a_y$ ou L_z/a_z égaux.

Principe de combinaison d'erreurs élémentaires



On estime la teneur pour la zone couverte par les deux carrés par $Z_v^* = 0.5 * (Z_1 + Z_2)$

La vraie teneur est $Z_v = 0.5 * (Z_{v1} + Z_{v2})$

L'erreur est :

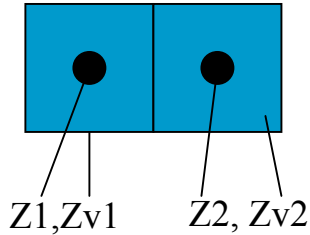
$$\begin{aligned} e &= (Z_v - Z_v^*) = 0.5 * [(Z_{v1} - Z_1) + (Z_{v2} - Z_2)] \\ &= 0.5 * (e_1 + e_2) \end{aligned}$$

$$\begin{aligned} \text{et } \text{Var}(e) &= 0.25 * (\text{Var}(e_1) + \text{Var}(e_2) + 2 * \text{Cov}(e_1, e_2)) \\ &= \text{Var}(e_1)/2 + \text{Cov}(e_1, e_2)/2 \end{aligned}$$

Le terme de variance d'estimation peut être obtenu d'une abaque

Le terme de covariance est en général très faible par rapport au terme de variance et peut être négligé $\Rightarrow \text{Var}(e) \approx \text{Var}(e_1)/2$

Exemple



Blocs 50 x 50

Variogramme sphérique avec $a=100$, $C=40$, $C_0=20$

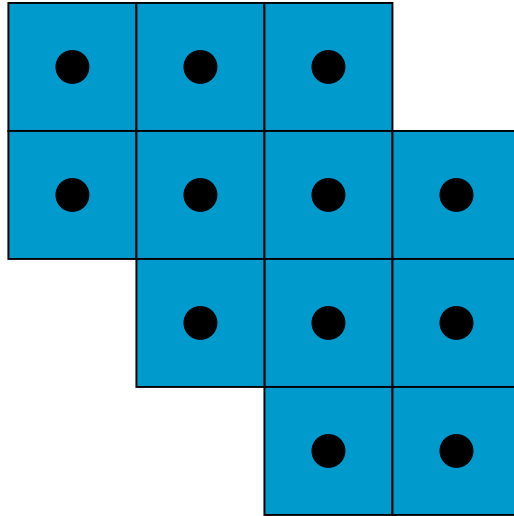
$$\text{Var}(e) \approx [40 * E3(0.5) + 20 * E3(\infty)] / 2 = [40 * 0.19 + 20 * 1] / 2 = 13.8$$

Note : avec l'abaque on obtient $\text{Var}(e1)=27.6$

avec un programme, on calcule: $\text{Cov}(e1, e2) \approx 0.02$

=> la covariance est négligeable

On peut généraliser cette approche à autant de cellules que nécessaire

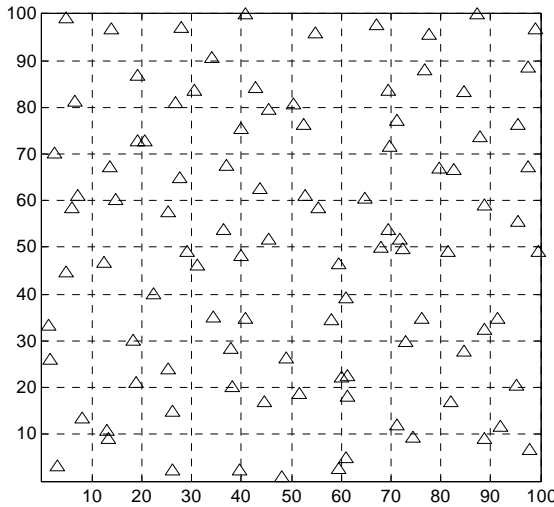


Même contexte,

$$\text{Var}(e) = \text{Var}(e_1)/n = 27.6/12 = 2.3$$

Que faire si les données ne sont pas sur une grille régulière ?

Précision globale pour un échantillonnage aléatoire stratifié



Plusieurs patrons d'échantillonnage peuvent être approchés par celui-ci

Physique

$$Z_g = \frac{1}{n} \sum_i Z_{vi}$$

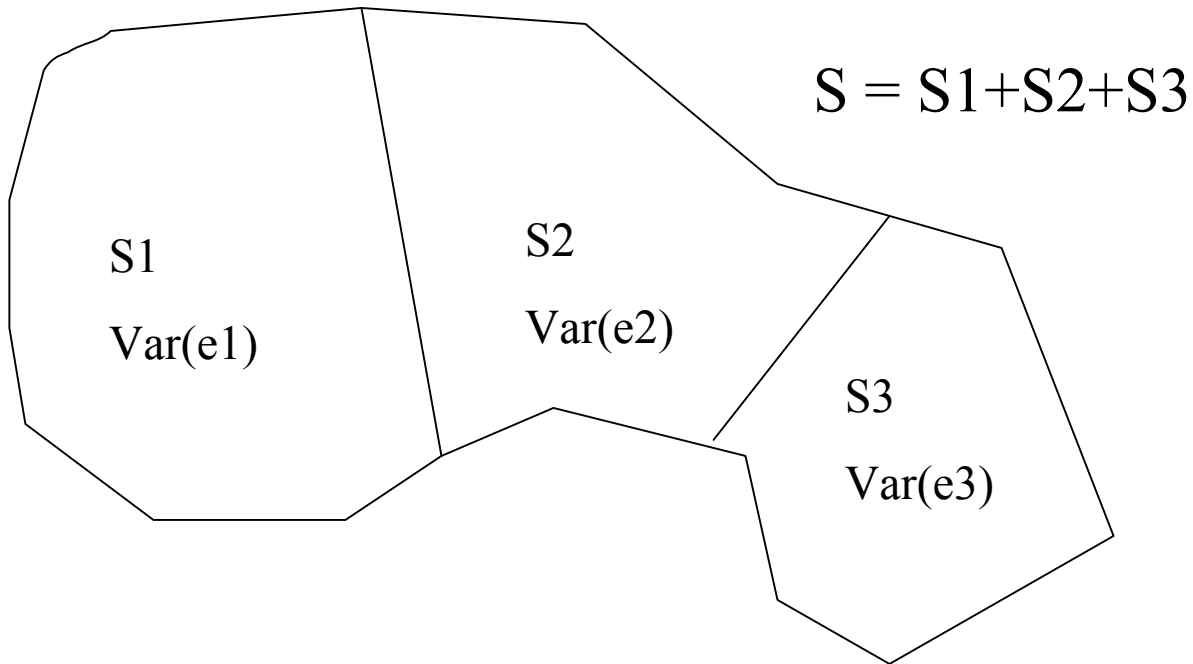
La plupart des estimateurs vont approcher la moyenne échantillonnale

$$Z_g^* \approx \frac{1}{n} \sum_i Z_i$$

En moyenne, la variance de l'erreur d'estimation d'une cellule par un point pris au hasard dans la cellule est la variance de dispersion d'une teneur ponctuelle dans la cellule

$$\left\{ \begin{aligned} Var(e_g) &\approx \frac{1}{n^2} \sum_i Var(e_i) \\ &\approx \frac{n}{n^2} D^2(\bullet | v) = \frac{1}{n} \bar{\gamma}(v, v) \end{aligned} \right.$$

Généralisation de la règle de composition des erreurs élémentaires

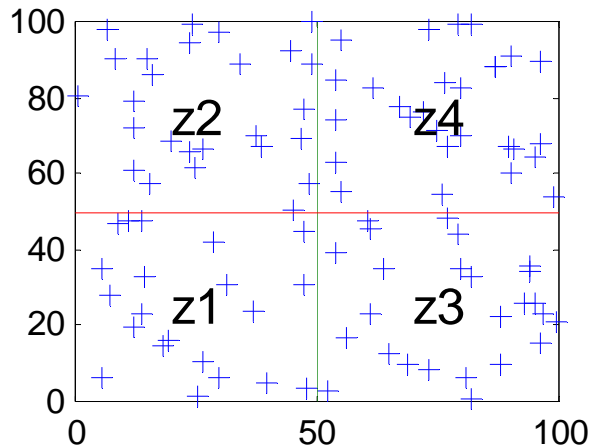


$$Var(e_s) \approx \frac{[S1^2 Var(e1) + S2^2 Var(e2) + S3^2 Var(e3)]}{S^2}$$

Condition d'application : chaque zone (S1, S2, S3) est estimée uniquement avec les données contenues dans la zone

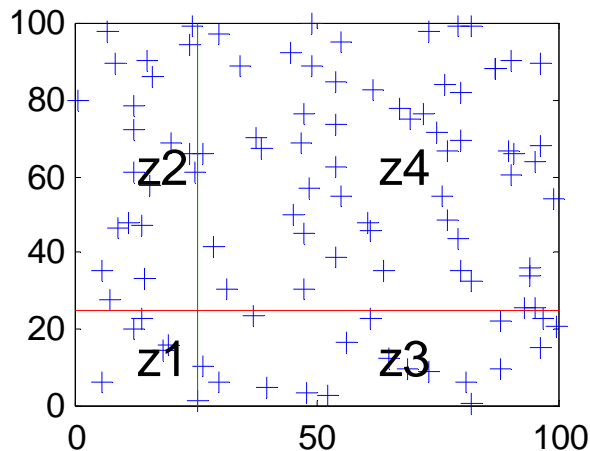
- Peut se généraliser à autant de zones que nécessaire
- $\text{Var}(e_i)$ peut être obtenu soit directement soit par composition des erreurs élémentaires (échantillonnage régulier ou aléatoire stratifié)
- On peut toujours recourir au calcul exact de $\text{Var}(e)$, mais l'approche par composition des erreurs élémentaires allège les calculs

Exemple



σ^2 zone`globale: 0.36
 σ^2 pour z1 : 1.6
 σ^2 pour z2 : 1.4
 σ^2 pour z3 : 1.4
 σ^2 pour z4 : 1.3

σ^2 combiné : 0.35



σ^2 pour z1 : 7.8
 σ^2 pour z2 : 1.7
 σ^2 pour z3 : 1.9
 σ^2 pour z4 : 0.63

σ^2 combiné : 0.36

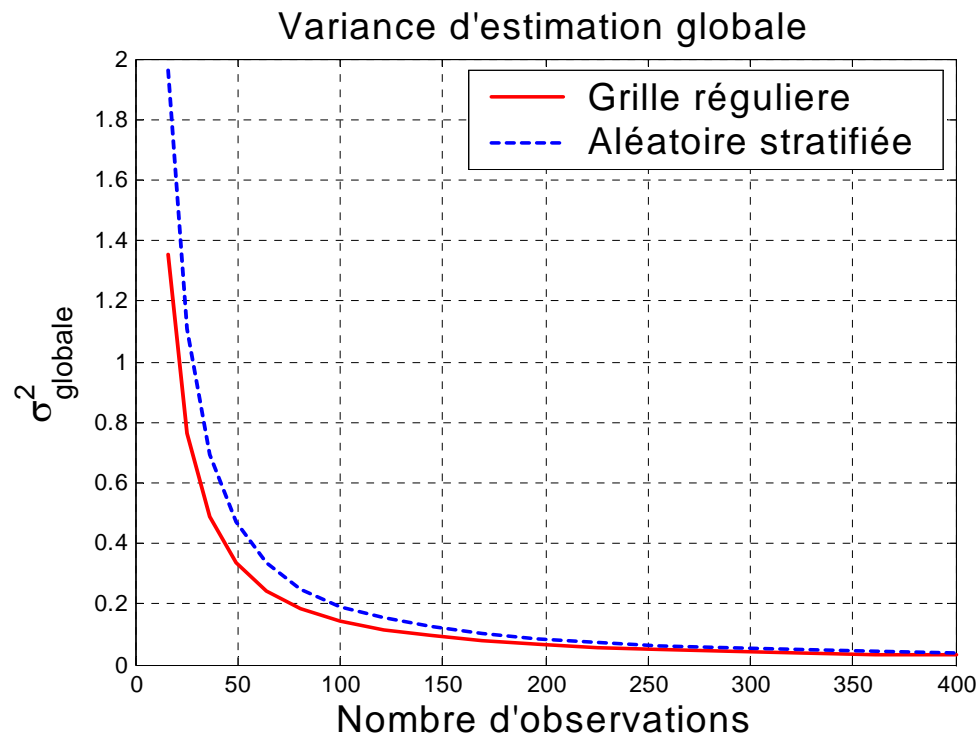
Applications

- Comparer la précision obtenue par différentes méthodes d'interpolation
- Déterminer le nombre d'observations requises pour atteindre une précision donnée
- Déterminer un patron d'échantillonnage optimal

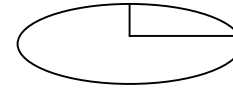
Exemple

Variogramme sphérique, $a=100$, $C=40$, $C_0=10$;

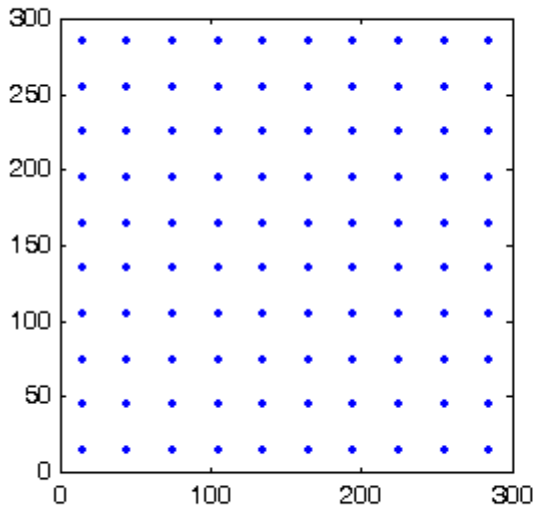
domaine à estimer : 300×300



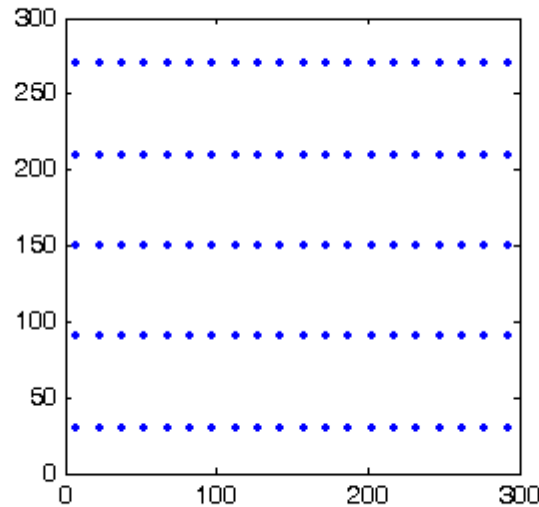
Variogramme sphérique, $C_0=10$, $C=40$, $a_x=100$ $a_y=25$; zone 300 x 300



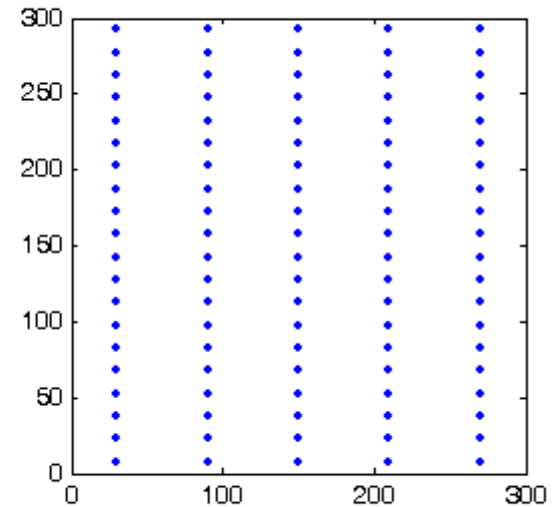
100 observations, 3 patrons, lequel procure la plus grande précision pour l'estimation globale ?



$$\sigma^2 = 0.24$$



$$\sigma^2 = 0.36$$



$$\sigma^2 = 0.19$$