

### 3. VARIANCES DE BLOCS, DE DISPERSION, D'ESTIMATION

#### 3.1 *Variances de blocs:*

On a vu précédemment l'importance de connaître la variance de la variable aléatoire correspondant au support d'exploitation de la mine. Ce support n'est évidemment pas la carotte mais plutôt un bloc d'une certaine taille (ex. 5m x 5m x 5m). On distingue 2 notions de variances différentes et complémentaires :

Variance de bloc: décrit l'amplitude théorique des variations des teneurs de bloc pour un domaine infini. C'est l'analogue de la variance ponctuelle (palier du variogramme) pour des blocs. Cette notion n'est définie que pour les modèles de variogramme avec palier.

Variance de dispersion: décrit l'amplitude théorique des variations des teneurs de bloc à l'intérieur d'un domaine fini. La variance de dispersion peut s'obtenir à partir de la variance de bloc comme on le verra plus loin. Cette notion est définie même pour les variogrammes sans palier.

Ces 2 notions interviennent dans le calcul des réserves récupérables d'un gisement, dans des calculs pour déterminer l'efficacité et le rendement de piles d'homogénéisation, dans le calcul de la variabilité de la production minière pour tout intervalle de temps désiré et dans le calcul de la variance de l'erreur d'estimation (variance d'estimation).

On peut calculer la variance des blocs si l'on connaît le variogramme des informations ponctuelles ou quasi-ponctuelles (carottes). De fait, on peut même calculer le variogramme (et le covariogramme) de blocs.

Soit  $Z(x)$  la v.a. correspondant à l'information ponctuelle.

Soit  $Z_v(x)$  la v.a. correspondant à un bloc centré en  $x$ .

On a

$$Z_v(x) = \frac{1}{v} \int_v Z(y) dy$$

Cette relation exprime simplement la réalité physique que la teneur d'un bloc est la moyenne des teneurs des points composant le bloc.

$$E [ Z_v(x) ] = m$$

et la variance de  $Z_v(x)$  s'écrit:

$$\begin{aligned} \text{Var}(Z_v(x)) &= \sigma_v^2 = E [ ( Z_v(x) - m )^2 ] \\ &= E \left[ \left( \frac{1}{v} \int_v Z(y) dy - m \right)^2 \right] \\ &= E \left[ \frac{1}{v^2} \int_v \int_v (Z(y_1) - m) (Z(y_2) - m) dy_1 dy_2 \right] \end{aligned}$$

On peut intervertir E et  $\int$  car ce sont deux opérateurs linéaires.

$$\begin{aligned} &= \frac{1}{v^2} \int_v \int_v E \left[ (Z(y_1) - m) (Z(y_2) - m) \right] dy_1 dy_2 \\ &= \frac{1}{v^2} \int_v \int_v \text{Cov}(Z(y_1), Z(y_2)) dy_1 dy_2 \end{aligned}$$

$$\boxed{\text{Var}(Z_v) = \sigma_v^2 = \bar{C}(v, v)}$$

Cette dernière expression indique que la variance du bloc  $v$  est donnée par la moyenne des covariances entre toutes les paires de points que l'on peut former à l'intérieur du bloc  $v$ .

En termes de variogramme, utilisant la relation

$$\boxed{C(h) = \sigma^2 - \gamma(h)}$$

On obtient

$$\boxed{\sigma_v^2 = \sigma^2 - \bar{\gamma}(v, v)}$$

**Si on connaît le variogramme (ponctuel), et si celui-ci montre un palier, alors on connaît toutes les variances de blocs, peu importe la taille ou la forme des blocs.** les blocs peuvent être constitués de plusieurs portions spatialement non-contiguës.

On vérifie aisément de cette dernière expression que pour tous les modèles croissants de variogramme:

$$\boxed{\begin{array}{ll} \text{si } v \rightarrow 0 & \sigma_v^2 \rightarrow \sigma^2 \\ \text{si } v \rightarrow \infty & \sigma_v^2 \rightarrow 0 \\ \text{si } v \uparrow & \sigma_v^2 \downarrow \end{array}}$$

Complément : variogramme de blocs (peut être omis dans une première lecture)

Considérons maintenant deux blocs séparés d'une distance  $h$ . Il peut être utile de définir la relation entre variogramme de bloc et variogramme ponctuel. Pour deux blocs séparés d'une distance  $h$ , on aura:

$$\gamma_v(h) = \frac{1}{2} E \left[ \left( \frac{1}{v(x)} \int_{v(x)} Z(y) dy - \frac{1}{v(x+h)} \int_{v(x+h)} Z(y') dy' \right)^2 \right] = \bar{\gamma}(v, v_h) - \bar{\gamma}(v, v)$$

où  $\bar{\gamma}(v, v_h)$  représente la valeur moyenne du variogramme ponctuel pour toutes les paires ayant un point dans  $v(x)$  et l'autre point dans  $v(x+h)$ .

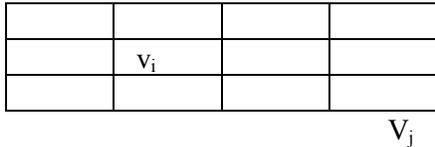
### 3.2 Variance de dispersion

La variance de bloc permet de calculer la variance théorique de la teneur de bloc dans un domaine d'extension infinie. Bien sûr les gisements réels ne sont jamais infinis et il est souhaitable de pouvoir prévoir l'amplitude des variations des teneurs de bloc pour un domaine fini correspondant au gisement ou à une partie du gisement.

Considérons un grand bloc  $V_j$  découpé en petits blocs  $v_i$ .

On a bien sûr:

$$Z(V_j) = \frac{1}{n} \sum_{i=1}^n Z(v_i)$$



On peut vouloir déterminer l'importance de la variation de  $v_i$  dans  $V_j$ , en moyenne pour l'ensemble des blocs  $V$ . C'est ce que l'on appelle la variance de dispersion de  $v$  dans  $V$  que l'on note  $D^2(v|V)$ .

Soit la variance échantillonnale pour un bloc  $V_j$ :

$$s_{v_i | V_j}^2 = \frac{1}{n} \sum_{i=1}^n (Z(v_i) - Z(V_j))^2$$

On définit la variance de dispersion comme l'espérance de cette variance expérimentale lorsqu'on considère tous les blocs possibles  $V_j$ :

$$\begin{aligned} D^2(v|V) &= E \left[ s_{v_i | V_j}^2 \right] = E \left[ \frac{1}{n} \sum_i (Z(v_i) - Z(V_j))^2 \right] \\ &= E \left[ \frac{1}{n} \sum_i \{ (Z(v_i) - m) - (Z(V_j) - m) \}^2 \right] \\ &= \sigma_V^2 + \sigma_v^2 - \frac{2}{n} \sum_{i=1}^n \text{Cov}(Z(V_j), Z(v_i)) \\ &= \sigma_V^2 + \sigma_v^2 - 2\text{Cov}(Z(V_j), Z(V_j)) \\ &= \sigma_v^2 - \sigma_V^2 \end{aligned}$$

$$D^2(v|V) = \sigma_v^2 - \sigma_V^2$$

i.e., la variance de dispersion n'est autre qu'une différence de variabilité de teneurs mesurées sur 2 volumes différents. Utilisant les résultats précédents concernant les variances de blocs, on peut obtenir les formulations équivalentes suivantes:

$$\begin{aligned} D^2(v|V) &= \bar{C}(v,v) - \bar{C}(V,V) \\ &= \bar{\gamma}(V,V) - \bar{\gamma}(v,v) \end{aligned}$$

On notera en particulier que, pour les modèles croissants de variogramme, on aura:

$$\begin{array}{ll} v \rightarrow 0 & D^2(v|V) \rightarrow \bar{\gamma}(V,V) \\ v \rightarrow V & D^2(v|V) \rightarrow 0 \\ V \rightarrow \infty & D^2(v|V) \rightarrow \sigma_v^2 \end{array}$$

Dans une mine, "v" pourrait correspondre à la production quotidienne et "V" à la production hebdomadaire ou mensuelle. Le rendement du concentrateur pourrait être relié à l'importance des fluctuations journalières sur une période mensuelle, i.e. à  $D^2(v|V)$ .

Les relations précédentes se généralisent aisément et permettent de définir une règle d'additivité très générale pour plusieurs blocs de taille différentes:

$$D^2(v_1|v_n) = D^2(v_1|v_2) + D^2(v_2|v_3) + \dots + D^2(v_{n-1}|v_n)$$

avec les tailles des supports ordonnés par ordre croissant:  $v_1 < v_2 < \dots < v_{n-1} < v_n$

Notes importantes concernant l'effet de pépité :

- i. Dans les relations précédentes,  $\gamma(h)$  représente le variogramme ponctuel. Dans la pratique, ce variogramme n'est pas accessible, seul le variogramme défini sur un certain support « s » existe (par exemple les carottes de forage). Les relations précédentes peuvent toutefois être utilisées tant que le support des données ayant conduit à la modélisation du variogramme est petit devant « v ». Dans ce cas, on constate que l'effet de pépité n'intervient pas dans le calcul de la variance de bloc et de la variance de dispersion.
- ii. Lorsque « v » n'est pas beaucoup plus grand que le support des données, alors un terme de variance lié à l'effet de pépité doit être inclus dans la variance de bloc. Ce terme a généralement pour valeur  $C_{0v} = sC_0/v$  où s est le volume du support des données et v le volume du bloc. Cependant cette interprétation n'est valide que si l'effet de pépité représente réellement une micro-structure (ex. pépites d'or) mais non s'il représente des erreurs de localisation ou une précision analytique.

Remarque : Le calcul des variances de bloc et de variance d'estimation ne nécessite pas de connaître explicitement les données, seul le variogramme des données est requis. Ces notions ne peuvent donc pas rendre compte de la connaissance accrue du gisement, que l'on a, localement, dû à l'acquisition de données.

### 3.3 Variance d'estimation

Dans cette section, on cherche à établir les résultats permettant de fournir une mesure de la précision des estimés effectués par une méthode d'estimation quelconque (linéaire).

Soit une v.a.  $Z_v$  que l'on veut estimer, d'une façon ou d'une autre, en formant une combinaison linéaire des valeurs observées en différents endroits, i.e.

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

où:

$Z_i$  : valeur observée au point  $x_i$  (v.a.)

$Z_v^*$  : estimateur de  $Z_v$

On définit l'erreur d'estimation :

$$e = Z_v - Z_v^*$$

La variance de cette erreur est la **variance d'estimation** :

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

Substituant  $Z_v^*$  par son expression, en fonction des  $Z_i$ , donnée en (1), on obtient:

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

Qui peut être réécrit en fonction du variogramme:

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

Puis finalement, puisqu'on a habituellement  $\sum \lambda_i = 1$ ,

$$\boxed{\sigma_e^2 = 2 \sum_i \lambda_i \bar{\gamma}(x_i, v) - \bar{\gamma}(v, v) - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i, x_j)} \quad (3)$$

On peut calculer la variance d'estimation soit en utilisant le covariogramme (2) ou le variogramme (3). Dans les cas de modèles sans palier, seul le variogramme peut être utilisé pourvu que  $\sum \lambda_i = 1$  dans (1).

- i. Dans les formules précédentes (2 et 3), on reconnaît 3 termes : 1 terme lié au bloc à estimer ( $\text{Var}(Z_v)$  ou  $\bar{\gamma}(v, v)$ ), 1 terme lié aux points servant à l'estimation ( $\text{Cov}(Z_i, Z_j)$  ou  $\gamma(x_i, x_j)$ ) et 1 terme croisé entre les points servant à l'estimation et le bloc à estimer ( $\text{Cov}(Z_i, Z_v)$  ou  $\bar{\gamma}(x_i, v)$ ).
- ii. La variance de l'erreur d'estimation est une mesure de la précision de l'estimation. On pourrait vouloir choisir les  $\lambda_i$  de façon à ce que  $\sigma_e^2$  soit minimale. C'est ce que nous ferons plus tard avec le krigeage.
- iii. La variance d'estimation est une mesure de précision obtenue, en moyenne, sur l'ensemble du gisement pour une même configuration points-bloc. On constate en effet que la variance d'estimation peut être calculée dès que l'on connaît le variogramme (ou le covariogramme), l'emplacement des points de données et le bloc à estimer. La variance d'estimation ne permet donc pas de tenir compte du fait que certaines portions pauvres du gisement sont peut-être plus faciles à estimer que des zones à haute teneur (effet proportionnel rencontré par exemple avec des distributions lognormales). On peut en tenir compte par l'emploi d'un variogramme relatif (normé par la moyenne locale au carré). Toutefois l'estimation du variogramme relatif est souvent délicate. Les méthodes non-linéaires et les simulations permettent de mieux tenir compte de ce facteur.

Ex. Soit  $x_1, x_2$  et  $x_3$  trois points échantillonnés que l'on veut utiliser pour estimer la valeur inconnue au point  $x_0$ .

$x_1$

$x_2 \quad x_0 \quad x_3$

Soit la matrice des distances séparant ces points:

	$x_0$	$x_1$	$x_2$	$x_3$
$x_0$	0	1.4	1	2
$x_1$		0	1	3.2
$x_2$			0	3

Supposons que le variogramme est un modèle linéaire avec pente unitaire et effet de pépite égal à 1, i.e.:

$$\gamma(h) = 1 + h$$

Calculons la variance d'estimation obtenue par trois méthodes différentes d'estimation:

a) Estimation polygonale (plus proche voisin)

$$\lambda_1 = \lambda_3 = 0 \quad \lambda_2 = 1$$

$$Z_0^* = Z_2$$

$$\begin{aligned} \sigma_{\varepsilon}^2 &= 2 \gamma(x_0, x_2) - \gamma(x_0, x_0) - \gamma(x_2, x_2) \\ &= 2(1 + 1) - 0 - 0 = 4 \end{aligned}$$

b) Inverse de la distance

$$1/d_1 = 1/1.4 \quad 1/d_2 = 1 \quad 1/d_3 = 1/2$$

$$\sum_i 1/d_i = 2.21$$

$$\lambda_1 = .32 \quad \lambda_2 = .45 \quad \lambda_3 = .23$$

$$\begin{aligned} \sigma_{\varepsilon}^2 &= 2 \sum_i \lambda_i \gamma(x_i, x_0) - \gamma(x_0, x_0) - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i, x_j) \\ &= 2( .32(1+1.4) + .45(1+1) + .23(1+2) ) - 0 \\ &\quad - 2( .32 \cdot .45(1+1) + .32 \cdot .23(1+3.2) + .45 \cdot .23(1+3) ) \\ &\approx 2.7 \end{aligned}$$

Ici l'inverse de la distance est très supérieur à la méthode polygonale en terme de variance d'estimation.

c) krigage

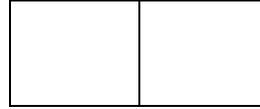
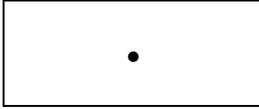
Par krigage, on aurait obtenu  $\lambda_1=.25$ ,  $\lambda_2=.43$ , et  $\lambda_3=.32$ . Le calcul de la variance d'estimation donne alors  $\sigma^2 \cong 2.65$ , une amélioration négligeable par rapport à l'inverse de la distance.

Variance d'extension :

C'est la variance d'estimation obtenue lorsqu'on étend la valeur d'un point à une surface ou un volume; la valeur d'un segment à une surface ou un volume; la valeur d'une surface à un volume, etc. Bref, il s'agit de

variances d'estimation correspondant à des situations particulières qui, par leur simplicité, se prêtent bien à la construction d'abaques.

ex.



etc.

### 3.4 Calcul des quantités $\bar{\gamma}(v,v)$ ou $\bar{\gamma}(x_i,v)$

Les termes  $\bar{\gamma}(v,v)$  ou  $\bar{\gamma}(x_i,v)$  sont requis pour obtenir les variances de blocs, les variances de dispersions et les variances d'estimation. Ces valeurs peuvent être calculées de trois façons différentes :

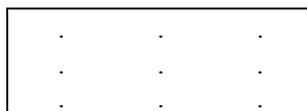
i. Intégration de  $\gamma(h)$  pour obtenir une expression analytique : praticable surtout en 1-D.

$$\begin{aligned} \text{ex. Sphérique : } \bar{\gamma}(v,v)/C &= 0.5 v/a - 0.05 (v/a)^3 && \text{si } v < a \\ &= 1 - 0.75 a/v + 0.2 (a/v)^2 && \text{si } v \geq a \end{aligned}$$

où  $a$  est la portée.

ii. Abaques pour certains modèles (sphérique; exponentiel).

iii. Approximation numérique par ordinateur. On représente le bloc  $v$  par une grille discrète de points suffisamment fine et on calcule la valeur moyenne du variogramme pour toutes les paires que l'on peut former avec cette grille de points (traitement spécial requis pour  $h=0$ ). On peut aussi utiliser une méthode de Monte Carlo pour l'approximation numérique, cette dernière méthode présentant des avantages dans le calcul de  $\bar{\gamma}(v,v)$ . Il s'agit alors de placer aléatoirement  $n$  paires de points à l'intérieur de " $v$ " et de calculer la valeur moyenne du variogramme pour toutes ces paires.



On calcule  $\gamma(x_i-x_j)$  pour toutes les paires possibles sur cette grille et on effectue la moyenne.

(Méthode la plus utilisée de nos jours).

### Utilisation des abaques :

#### ex. Variances de blocs, en 2-D.

Soit le modèle de variogramme sphérique avec  $C_0=5$ ,  $C=15$  et  $a=100$  m.

Quelle est la variance de blocs de 20 m X 20 m?

$$\begin{aligned}
\sigma_v^2 &= \sigma^2 - \bar{\gamma}(v, v) \\
&= (C_0 + C) - \left( C_0 + C \cdot F\left(\frac{L}{a_L}, \frac{l}{a_l}\right) \right) \\
&= C \left( 1 - F\left(\frac{L}{a_L}, \frac{l}{a_l}\right) \right) \\
&= 15 (1 - F(0.2, 0.2)) \\
&= 15 \cdot 0.84 = 12.6
\end{aligned}$$

Pour des blocs 50 m X 50 m, on trouve  $F(0.5, 0.5) \approx 0.38$  et

$$\sigma_v^2 \approx 15 \cdot 0.62 = 9.3$$

Pour des blocs 50 m x 100 m, on trouve  $F(0.5, 1) \approx 0.54$  et

$$\sigma_v^2 \approx 15 \cdot 0.46 = 6.9$$

Si au lieu du variogramme isotrope précédent une anisotropie géométrique avec  $a_x = 100$ ,  $a_y = 50$  était présente, on aurait alors pour:

un bloc 50 m en x par 100 en y;  $F(.5, 2) \approx .73$

$$\sigma_v^2 \approx 15 \cdot .27 \approx 4.1$$

un bloc 100 m en x par 50 en y;  $F(1, 1) \approx .68$

$$\sigma_v^2 \approx 15 \cdot .32 = 4.8$$

On constate que la variance des blocs dépend de leur allongement par rapport à la direction d'anisotropie!

**Note:** Lorsqu'il y a anisotropie, il faut, pour pouvoir utiliser les abaques, que les directions principales d'anisotropie soient parallèles aux côtés du bloc considéré.

### ex. Variance de blocs 3-D

Les abaques sont construites en supposant qu'au moins 2 des 3 rapports  $l_x/a_x$ ,  $l_y/a_y$ ,  $l_z/a_z$  sont égaux. Ces rapports égaux sont placés en ordonnée de l'abaque  $F(L/a_L, l^2/a_l)$ . Attention, cette abaque n'est pas symétrique!

Ex. Bloc 50 m X 50 m X 25 m

Avec le modèle sphérique précédent ayant  $C_0=5$ ,  $C=15$  et  $a=100$ , on trouve  $F(.25, .50) \approx 0.4$

$$\sigma_v^2 \approx 15 \cdot 0.6 = 9.0$$

### ex. Variance de dispersion

On utilise les abaques pour chacun des termes et on obtient la variance de dispersion par différence. (Rappelons que l'abaque F( , ) permet d'obtenir la valeur de  $\bar{\gamma}(v, v)$ ).

### ex. Variances d'estimation

Chaque configuration et chaque estimateur donnent lieu à un calcul différent. Il n'existe donc des abaques que pour les cas simples d'extension.

Ex. Toujours avec le sphérique ayant  $C_0=5$ ,  $C=15$  et  $a=100$  m,

Si on estime une cellule de 10 m X 10 m par son point central, on obtient avec l'abaque #7,  $\sigma_{E3}^2$ , évaluée à  $1/a = 10/100 = 0.1$  :

$$\sigma_e^2 \approx 15 \cdot .038 = 0.57$$

A cette quantité, il faut ajouter un terme relié à  $C_0$ , d'où:

$$\sigma_e^2 \approx .57 + 5 = 5.57$$

Le terme relié à  $C_0$  provient de l'expression générale de la variance d'estimation (pour tout estimateur linéaire):

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

En identifiant les termes en  $C_0$  dans cette dernière expression, on voit que la variance d'estimation est accrue de:

$$2 \sum_i^n \lambda_i C_0 - C_0 - \sum_{\substack{i, j \\ i \neq j}}^n \lambda_i \lambda_j C_0$$

$$\text{i.e. } C_0 \left( 2 \sum_i^n \lambda_i - 1 - \sum_{\substack{i, j \\ i \neq j}}^n \lambda_i \lambda_j \right)$$

On peut simplifier davantage puisqu'on a presque toujours:

$$\sum_i \lambda_i = 1$$

$$C_0 \left( 1 - \sum_{\substack{i, j \\ i \neq j}}^n \lambda_i \lambda_j \right) = C_0 \cdot \sum_i \lambda_i^2$$

Avec un bloc de 100 m X 100 m, on aurait obtenu:

$$\sigma_e^2 \approx 15 \cdot .41 + 5 = 11.15$$

Si plutôt que d'utiliser la valeur de l'intersection centrale, on avait utilisé les quatre coins i.e.  $n=4, \lambda_i = 1/4$ ,  $i = 1..4$ , on aurait obtenu pour le bloc de 10 m X 10 m:

$$\sigma_e^2 \approx 15 \cdot .024 + 5 \cdot 1/4 = 1.61$$

Et, pour le bloc de 100 m X 100 m:

$$\sigma_e^2 \approx 15 \cdot .27 + 5 \cdot 1/4 = 5.3$$

Une réduction très substantielle de la variance d'estimation est obtenue en utilisant les quatre coins de préférence à la valeur centrale. On voit que lorsqu'on a un effet de pépité important, on a intérêt à former un estimateur utilisant plus d'observations de façon à diminuer la variance d'estimation. En effet, lorsque les poids sont égaux ( $1/n$ ) la part de la variance d'estimation reliée à l'effet de pépité est  $C_0/n$ . On voit ici une justification théorique aux meilleures performances observées pour les estimateurs inverse de la distance et méthode des triangles, par rapport à la méthode polygonale. Le simple fait d'utiliser plus d'observations dans l'estimation améliore la précision de celle-ci surtout en présence d'un effet de pépité important.

On peut aussi vouloir estimer une cellule ou un bloc par un segment, un bloc par une surface, etc. Des abaques existent pour certaines de ces situations.

ex. Extension d'un segment à une surface

Toujours avec le même modèle de variogramme, pour une cellule de 10 m X 10 m qu'on estime par son segment central. S'il n'y avait pas d'effet de pépité, on aurait:

$$\sigma_e^2 = 15 \cdot .009 = 0.14$$

Le segment renferme "n" petits segments de la longueur des supports ayant servi à définir le variogramme. Le poids associé à chaque petit segment (support) est donc de 1/n et la part de l'effet de pépité dans la variance d'extension sera donc de  $C_0/n$ . Quand le support est très petit par rapport au segment considéré, alors  $n \rightarrow \infty$ , et la part liée à l'effet de pépité devient nulle. Si le support est le segment au complet, alors n est 1 et la variance d'extension est accrue de  $C_0$ .

Dans l'exemple précédent, si le support est constitué de carottes de 1 m, alors

$$\sigma_e^2 \approx 0.14 + 5/10 = .64$$

### **3.5 Combinaison d'erreurs élémentaires pour une estimation globale**

Si l'on a beaucoup d'observations et qu'on désire estimer une moyenne pour un volume important, le calcul de la variance d'estimation utilisant la formule théorique peut devenir laborieux. Il existe toutefois un principe qui permet de simplifier les calculs tout en conservant une bonne précision sur le calcul de la variance d'estimation. Ce principe est celui des extensions élémentaires. Il consiste à décomposer une estimation donnée en une série d'estimations élémentaires approximativement indépendantes et pour lesquelles la variance d'estimation est facilement calculable.

#### **i. ex. estimation de la moyenne d'un champ à partir d'une grille régulière.**

L'estimé est alors simplement:  $\frac{1}{n} \sum_{i=1}^n Z_i$ . C'est comme si on étendait la valeur de chaque point à chaque bloc

correspondant aux dimensions de la grille régulière. Pour chaque extension, on commet une erreur dont la variance est  $\sigma_{ei}^2$ . Cette variance est constante pour chaque bloc. Ces erreurs sont approximativement indépendantes puisque chaque bloc est estimé par son point central et qu'il n'y a donc pas deux estimés utilisant les mêmes données. La variance de l'erreur globale est donc la variance d'une somme de "n" erreurs

élémentaires indépendantes, i.e.  $\frac{\sigma_{ei}^2}{n}$

**ii. ex. estimation de la moyenne d'un champ à partir d'un échantillonnage aléatoire mais à densité uniforme** (échantillonnage aléatoire stratifié).

Dans ce cas, on peut considérer qu'on a implanté aléatoirement un point dans chaque cellule élémentaire. On trouve la taille de cette cellule élémentaire à partir de la taille du champ total et du nombre et de la répartition des points échantillons. La variance de l'erreur pour un bloc « i » est  $\sigma_{e_i}^2 = E[(Z_i - Z_v)^2]$ . Cette variance dépend de l'emplacement précis du point dans le bloc "i". Le point étant placé aléatoirement, on peut calculer la valeur probable de cette variance d'estimation élémentaire par:

$$\frac{1}{v} \int_v E [(Z(x) - Z_v)^2] dx$$

Qui n'est rien d'autre que la variance de dispersion d'un point dans un bloc v, i.e.  $D^2(\cdot|v)$ .

À nouveau, puisque chaque erreur est indépendante, pour l'ensemble du champ, la variance d'estimation sera  $D^2(\cdot|v)/n$ .

**iii. ex. estimation d'une moyenne d'un champ à partir d'un échantillonnage quelconque.**

On peut appliquer la méthode décrite au point ii. sur des sous-domaines où l'hypothèse de répartition aléatoire stratifiée est raisonnable. On calcule alors par la méthode décrite au point ii. les variances d'estimation pour chaque sous-domaine. Pour l'ensemble du champ, l'estimé est obtenu en combinant, selon des poids proportionnels aux volumes des sous-domaines, les différents estimés obtenus. On aura donc:

$$Z_g^* = \frac{1}{V} \sum_{i=1}^p v_i Z_{v_i}^* \quad \text{avec} \quad V = \sum_{i=1}^n v_i$$

et la variance d'estimation sera, puisque chaque erreur est considérée indépendante:

$$\sigma_g^2 = \frac{1}{V^2} \sum_{i=1}^n v_i^2 \sigma_{e_i}^2$$

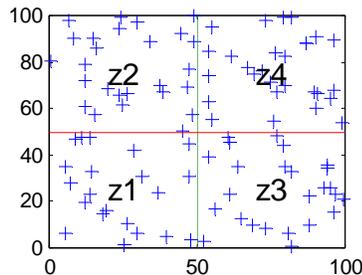
où les variances d'estimation élémentaires sont celles correspondant aux différents sous-domaines que l'on a pu reconnaître.

Note: - Paradoxalement, l'estimateur particulier utilisé pour obtenir l'estimé pour la moyenne du champ n'intervient pas dans le calcul de la variance d'estimation globale par la méthode des erreurs élémentaires. L'estimation aurait pu être obtenue par krigeage, par inverse de la distance, par méthode polygonale, etc.. La raison de cette apparente anomalie est que les estimés globaux obtenus par ces méthodes sont très similaires. Ils consistent plus ou moins en une moyenne, pondérée par la zone d'influence de chaque observation, des valeurs observées. Puisque les estimations sont similaires, il n'est pas surprenant qu'elles aient toutes à peu près la même précision. Ce raisonnement n'est pas vrai si l'on cherche à prédire les variances d'estimation pour une estimation locale, i.e. pour une petite portion du champ. Alors, la technique d'estimation utilisée a une forte influence sur la précision obtenue.

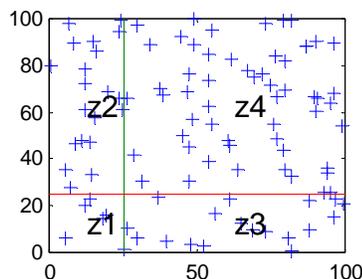
- Si l'on dispose d'un programme permettant d'effectuer le krigeage et de calculer les variances de krigeage (variance d'estimation), alors on peut les combiner suivant le principe précédent, i.e. il suffit de segmenter le domaine en blocs et d'estimer chaque bloc en n'utilisant que les données qui s'y

*trouvent*, puis de combiner les variances de krigeage en fonction de la taille des blocs comme l'indique la formule précédente. Il est important de ne pas avoir des données communes pour l'estimation de deux blocs car alors les erreurs d'estimation ne pourraient plus être considérées comme indépendantes.

Ex. Une zone a été estimée directement par krigeage (variance d'estimation 0.36) puis par combinaison de 4 parcelles (selon 2 scénarios différents). Pour l'estimation de chaque parcelle, on n'utilise que les points s'y retrouvant. On calcule les variances d'estimation pour chaque parcelle et on combine le tout suivant le carré des surfaces de chaque parcelle. Les résultats obtenus par subdivision, dans les deux cas, sont quasi-identiques au résultat direct.



$$\begin{aligned} \sigma^2 \text{ zone globale: } & 0.36 \\ \sigma^2 \text{ pour } z1 : & 1.6 \\ \sigma^2 \text{ pour } z2 : & 1.4 \\ \sigma^2 \text{ pour } z3 : & 1.4 \\ \sigma^2 \text{ pour } z4 : & 1.3 \\ \sigma^2 \text{ combiné : } & 0.35 \end{aligned}$$



$$\begin{aligned} \sigma^2 \text{ pour } z1 : & 7.8 \\ \sigma^2 \text{ pour } z2 : & 1.7 \\ \sigma^2 \text{ pour } z3 : & 1.9 \\ \sigma^2 \text{ pour } z4 : & 0.63 \\ \sigma^2 \text{ combiné : } & 0.36 \end{aligned}$$

#### En résumé :

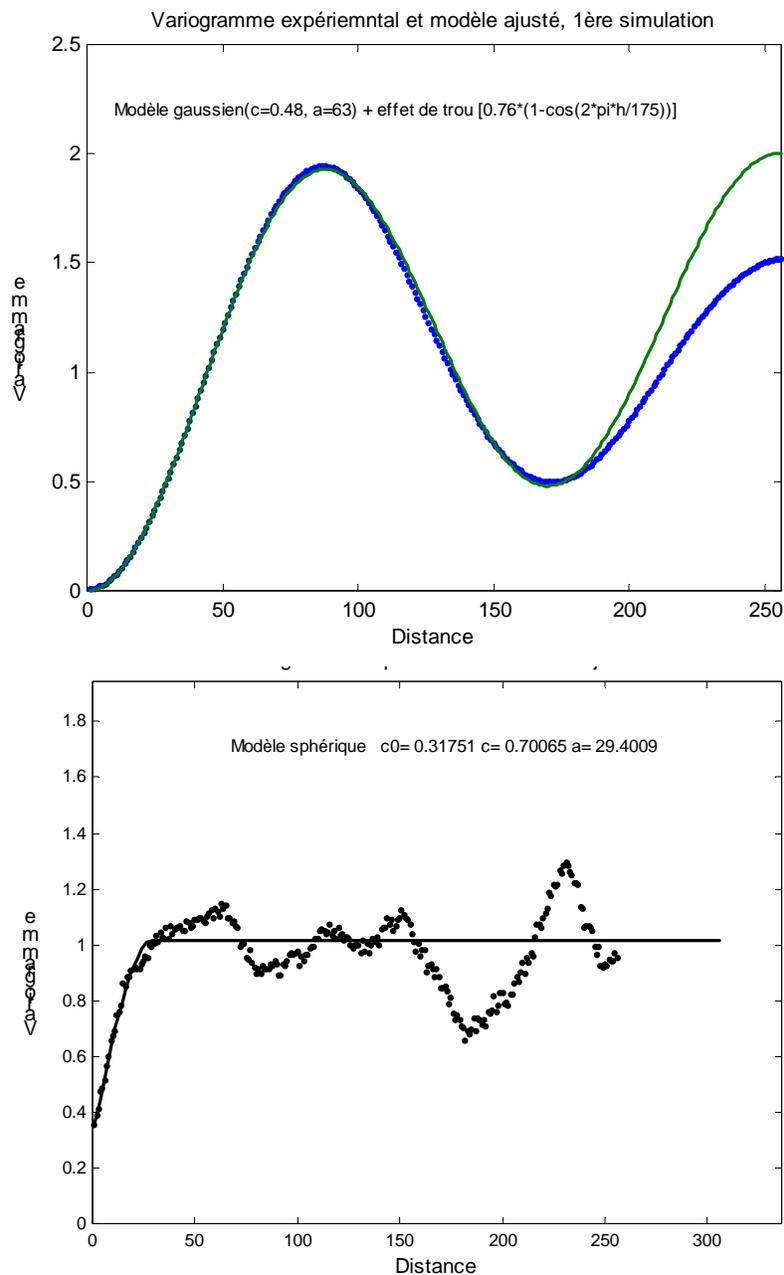
- i. les variances de bloc, de dispersion et d'estimation peuvent toutes être calculées à partir du variogramme ponctuel (ou quasi-ponctuel);
- ii. pour la variance de bloc et la variance de dispersion, si les données sont définies sur un support quasi-ponctuel, alors l'effet de pépité n'intervient pas. Si les données sont définies sur un support non ponctuel, alors l'effet de pépité intervient pour les petits blocs. Pour la variance d'estimation (en particulier d'extension), le traitement de l'effet de pépité dépend du support des données servant à l'estimation relativement au support ayant servi à l'obtention du variogramme;
- iii. pour le calcul de la variance d'estimation d'une quantité moyenne sur un grand champ, on peut recourir à des approximations basées sur des erreurs élémentaires à peu près indépendantes. Ceci n'a pour but que de simplifier les calculs et devrait fournir de bonnes approximations des valeurs que l'on obtiendrait en utilisant la formule générale pour la variance d'estimation.

### 3.6 Un exemple simulé pour la variance de dispersion

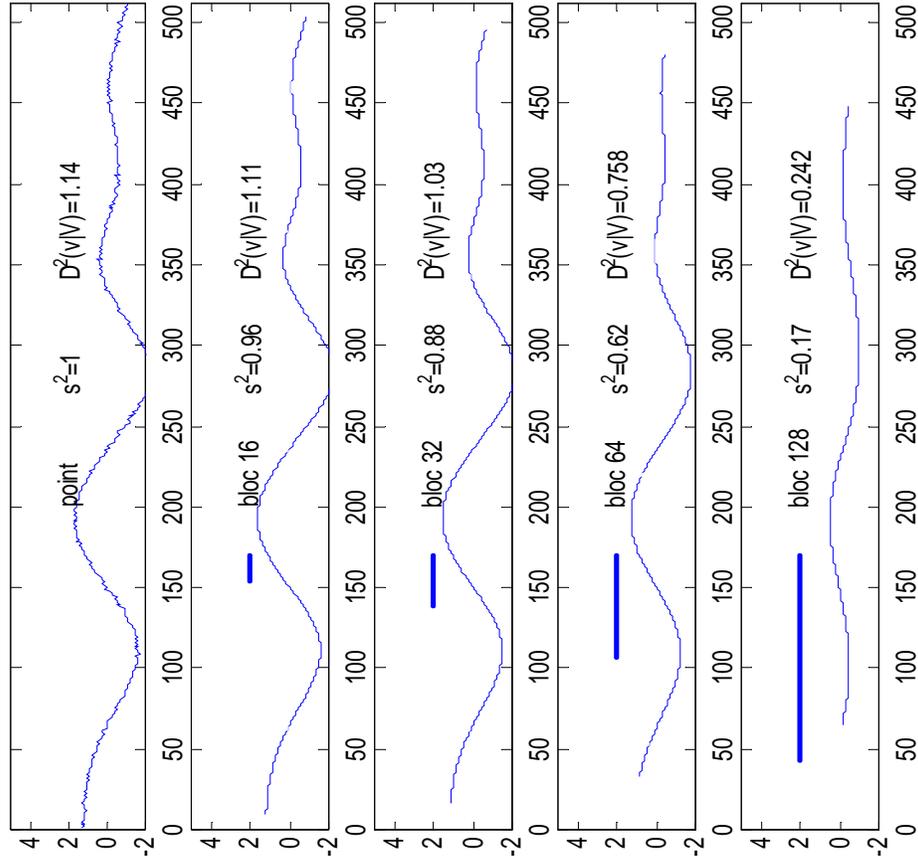
On a simulé 2 séries de 512 données (en 1D) à un intervalle régulier de 1. Les variogrammes expérimentaux et ajustés pour les 2 séries sont illustrés à la Figure 1. On a ensuite regroupé les données

en paquets de 16, 32, 64 et 128 en glissant une fenêtre le long de la série. Dans chaque cas, on a calculé la variance expérimentale de ces valeurs ( $s^2$ ) et la valeur théorique  $D^2(v|V)$  calculée à partir du variogramme (Figure 2).

On notera que pour une variance ponctuelle égale, les blocs formés sur la série ponctuelle montrant la meilleure structure spatiale (série 1) demeurent beaucoup plus variables que les blocs formés sur la série ayant la moins bonne structure spatiale (série 2). On notera également, l'assez bonne concordance entre les valeurs théoriques et les valeurs expérimentales obtenues, surtout pour la 2e série. L'écart entre celles-ci s'explique par le fait que pour la valeur expérimentale, on considère un seul bloc  $V = 512$ , alors que la valeur théorique représente, par définition, une valeur moyenne de cette variance expérimentale prise sur un très grand nombre de blocs  $V$ .



Simulation 1



Simulation 2

