

**Le cas d'une dérive sur la moyenne : les fonctions aléatoires intrinsèques d'ordre k (FAI-k) et les covariances généralisées.**

Dans le krigeage avec dérive, l'on doit identifier la covariance de  $Z(x)=m(x)+Y(x)$  où  $Y(x)$  est le résidu autocorrélé. On a, si l'on suppose que  $m(x)$  est déterministe, que :  $Cov(Z(x),Z(x+h))=Cov(Y(x),Y(x+h))$ . On peut donc estimer simplement la covariance des résidus.

Différentes approches sont possibles :

- 1- identifier une direction où il n'y a pas de dérive et supposer l'isotropie;
- 2- supposer que la dérive est de faible ampleur et modéliser directement la covariance  $Z(x)$ ;
- 3- estimer simultanément par maximum de vraisemblance  $m(x)$  et  $C(h)$ ;
- 4- recourir aux FAI-k qui sont insensibles au choix particulier de la dérive.

**Principe des FAI-k**

L'idée est essentiellement la même que celle utilisée pour le variogramme.

**FAI-0.** La moyenne inconnue est constante. Pour se débarrasser du problème de devoir estimer cette constante, on travaille sur des incréments d'ordre 0 :  $Z(x)-Z(x+h)$

Ces incréments ont moyenne 0. Ils filtrent donc la moyenne inconnue  $m$ . La variance de ces incréments est par définition  $2\gamma(h)$ . On peut écrire :

$$\text{Var}(Z(x)-Z(x+h)) = [1 \ -1] \begin{bmatrix} 0 & -\gamma(h) \\ -\gamma(h) & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2\gamma(h)$$

On dit que la fonction  $-\gamma(h)$  est la covariance généralisée de l'incrément d'ordre 0  $Z(x)-Z(x+h)$  puisqu'il permet d'en calculer la variance. Ainsi dans le cas stationnaire,  $C(h)$  est aussi covariance généralisée, tout comme  $-\gamma(h)$  d'ailleurs. Dans le cas FAI-0, le variogramme permet de calculer la variance de l'incrément-0 sans connaître la moyenne, ce que ne peut faire la covariance.

Note : en fait on peut calculer la variance de tout incrément d'ordre 0 sachant uniquement le variogramme. Ainsi  $Z(x_1)+Z(x_2)-2Z(x_3)$  est aussi un incrément-0 et sa variance est donnée par :

$$\text{Var}(Z(x_1)+Z(x_2)-2Z(x_3)) = [1 \ 1 \ -2] \begin{bmatrix} 0 & -\gamma(h_{12}) & -\gamma(h_{13}) \\ -\gamma(h_{12}) & 0 & -\gamma(h_{23}) \\ -\gamma(h_{13}) & -\gamma(h_{23}) & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

**FAI-k.** Si la moyenne peut être décrite par un polynôme d'ordre k :  $m(x) = \sum_{l=0}^k a_l f_l(x)$ , alors il est possible de former des incréments d'ordre k permettant de filtrer ce polynôme.

Ex. En 1D pour des données régulièrement espacées à  $x=1,2,3$ ,  $Z(1)-2Z(2)+Z(3)$  filtre une dérive d'ordre 1. Plus généralement,  $Z(x-h)-2Z(x)+Z(x+h)$  filtre une dérive d'ordre 1.

En 2D,  $Z(x-h,y)+Z(x,y-h)+Z(x,y+h)+Z(x+h,y)-4*Z(x,y)$  filtre une dérive d'ordre 1 (un plan).

En fait, toute combinaison linéaire  $\sum_{i=1}^n \alpha_i Z_i$  telle que :  $\sum_{i=1}^n \alpha_i f_l(x_i) = 0 \quad \forall l = 0..k$ , est un

incrément d'ordre k qui filtre le polynôme d'ordre k représentant la moyenne  $m(x)$ . On dit que la combinaison linéaire qui respecte les équations précédentes est « autorisée » puisqu'elle filtre la moyenne.

### Comment obtenir des incréments d'ordre k?

On peut obtenir des incréments d'ordre k par krigeage avec dérive d'ordre k. L'erreur de krigeage  $Z(x) - Z^*(x)$  est par définition une combinaison linéaire autorisée, et ce, peu importe la fonction de covariance (ou de variogramme, ou de covariance généralisée) utilisée. La variance de krigeage est par définition la variance de cette combinaison linéaire autorisée.

On forme des incréments différents en krigeant tour à tour les points où l'on a des données. On peut générer des incréments supplémentaires en modifiant la fonction utilisée dans le krigeage ou en modifiant le voisinage utilisé. En fait, l'on peut former un nombre infini d'incrément généralisés.

### Covariance généralisée

Dans le cas des FAI-0, on a vu apparaître une fonction plus générale que la covariance : le variogramme. Si toute covariance peut être exprimée comme un variogramme, l'inverse n'est pas vrai puisque les variogrammes sans palier ne peuvent être exprimés comme une covariance. Cependant, les variogrammes sans palier peuvent être utilisés pour le krigeage ordinaire (mais pas pour le krigeage simple) puisque la contrainte sur les poids de krigeage filtre la moyenne inconnue (l'estimateur est sans biais). De même, pour une FAI-K, on peut définir de nouvelles fonctions appelées covariances généralisées qui permettent de réaliser le krigeage avec dérive d'ordre k, sans avoir à estimer le polynôme d'ordre k. L'estimateur obtenu est sans biais et l'on peut calculer la variance des incréments (et donc de l'erreur de krigeage) avec les covariances généralisées.

Ainsi, soit un incrément d'ordre k autorisé :  $\sum_{i=1}^n \alpha_i Z_i$  tel que :  $\sum_{i=1}^n \alpha_i f_l(x_i) = 0 \quad \forall l = 0..k$ , la

variance de cet incrément peut s'écrire :

$Var(\sum \alpha_i Z(x_i)) = \sum \sum \alpha_i \alpha_j K(x_i - x_j)$ . On dit alors que  $K(h)$  est une covariance

généralisée.  $K(h)$  est une fonction conditionnellement positive définie (i.e. elle assure que toute variance d'incrément généralisée est positive).

Exemples de covariances généralisées pour le cas  $k=1$  :

- toute covariance stationnaire (e.g. sphérique, exponentiel...) est une covariance généralisée;
- tout variogramme multiplié par -1 (e.g.  $-|h|$ ,  $-|h|^b$  ( $0 < b < 2$ )...) est une covariance généralisée;
- la fonction  $|h|^3$  ;
- la fonction  $|h|^2 \ln(|h|)$  (spline plaque mince).

Plus généralement, pour une dérive d'ordre k :

- $(-1)^{k+1}|h|^b$  avec b dans l'intervalle  $]2k, 2k+2[$  est une covariance généralisée;
- $(-1)^{k+1}|h|^{2k} \ln(|h|)$  est aussi une covariance généralisée (covariance de type spline).

Exemple : dans le cas 1D avec 3 données en  $x=1,2,3$  pour l'incrément  $W=Z(1)-2Z(2)+Z(3)$  avec  $K(h)=|h|^3$

$$\text{Var}(W) = [1 \ -2 \ 1] \begin{bmatrix} 0 & 1 & 8 \\ 1 & 0 & 1 \\ 8 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 8$$

avec  $K(h)=|h|^2 \log(|h|)$

$$\text{Var}(W) = [1 \ -2 \ 1] \begin{bmatrix} 0 & 0 & 2.7726 \\ 0 & 0 & 0 \\ 2.7726 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 5.5452$$

Note : on remarquera que l'on peut ajouter un polynôme pair d'ordre  $2k$  à la covariance généralisée sans que le résultat du calcul de la variance n'en soit affecté

Ainsi, avec  $K(h)=|h|^3+|h|^2$

$$\text{Var}(W) = [1 \ -2 \ 1] \begin{bmatrix} 0 & 2 & 12 \\ 2 & 0 & 2 \\ 12 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 8$$

et avec  $K(h)=|h|^2 \log(|h|)+|h|^2$

$$\text{Var}(W) = [1 \ -2 \ 1] \begin{bmatrix} 0 & 1 & 6.7726 \\ 1 & 0 & 1 \\ 6.7726 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 5.5452$$

### Estimation des paramètres de la covariance généralisée

Les paramètres des covariances généralisées peuvent être estimés en comparant les erreurs au carré d'une validation croisée avec les variances de krigeage (on peut aussi s'intéresser aux erreurs en valeur absolue ou toute autre statistique sur les erreurs). La validation croisée est effectuée en supposant un modèle de covariance et des coefficients inconnus à déterminer. En effectuant la validation croisée, on met à jour les coefficients du modèle. Lorsque la covariance généralisée est polynômiale (e.g.,  $b_1$  effet de pépite +  $b_2$  linéaire +  $b_3$  cubique) les coefficients peuvent être obtenus par une régression linéaire ou la variable à expliquer est les erreurs de krigeage au carré et les variables explicatives sont : l'effet de pépite,  $|h|$  et  $|h|^3$ ). Cependant des contraintes existent sur les coefficients inconnus afin d'assurer l'admissibilité du modèle. Ces contraintes varient selon la dimension du problème (1D, 2D ou 3D). Elles sont données dans Chilès et Delfiner (1999, p. 262).