

**Ecole des Mines
de Paris**



L'INFÉRENCE STATISTIQUE AUTOMATIQUE DES FAI-k

Jean-Paul CHILÈS

Rapport technique du CGMM

N-584

Décembre 1978

CENTRE DE GÉOSTATISTIQUE
35 rue Saint-Honoré
77305 FONTAINEBLEAU (France)
Tél. : 33-1 64 69 47 81 Fax : 33-1 64 69 47 05
<http://cg.ensmp.fr>

INTRODUCTION

L'objet de cette note est de présenter les fondements des choix qui ont été faits dans la mise en oeuvre de méthodes automatiques d'inférence statistique dans le programme KRIGEPACK. On verra en même temps quelle précision on peut obtenir. Les grandes lignes des méthodes utilisées dans KRIGEPACK et BLUEPACK ont déjà été présentées^x. On trouvera ici deux études différentes, de caractère plutôt expérimental que théorique :

- la première porte sur la détermination de l'ordre k de la FAI prise comme modèle ; celui-ci est déterminé par comparaison d'erreurs de krigeage pour différentes hypothèses sur k ; on étudie quel est le type de configuration de points qui permet le mieux de faire une inférence correcte de k .

- la seconde est consacrée à l'inférence de la covariance généralisée, dans l'hypothèse où celle-ci est de type polynômial ; la méthode employée est une régression de variances expérimentales à des variances théoriques ; les différentes combinaisons linéaires utilisées ayant des variances d'ordre de grandeur très variées, on doit les pondérer ; on étudie l'influence du choix du pondérateur sur la précision des résultats, ainsi que le type et le nombre des combinaisons linéaires à construire.

On suppose ici que la covariance est isotrope. On trouvera néanmoins en addendum quelques réflexions concernant la possibilité d'utiliser des schémas anisotropes.

Volontairement, nous n'aborderons pas les autres aspects de l'inférence, comme l'exploitation du variogramme brut, l'élaboration de modèles plus complexes dans le cas où on dispose d'un grand nombre de données, le contrôle des ajustements obtenus, etc... Ces points sont en effet déjà présentés dans les articles cités.

^xJ.P. CHILES - Géostatistique des phénomènes non stationnaires dans le plan, thèse, 1977

P. DELFINER - Linear estimation of non-stationary phenomena, NATO-ASI Rome, 1975

I - DETERMINATION DE L'ORDRE D'UNE FAI

1 - POSITION DU PROBLEME

Soit $Z(x)$ une variable régionalisée considérée comme une réalisation de représentation de FAI- k_0 de covariance généralisée $K^0(h)$. $K^0(h)$ est connue ou non ; nous supposerons cependant que $K^0(h)$ est une covariance polynômiale. k_0 est inconnu, et le problème posé est celui de sa détermination.

Nous travaillerons toujours dans un domaine borné D . Or nous savons que toute FAI- K^0 ^{à covariance polynômiale} admet des représentations stationnaires sur D . Nous pourrions donc, sur D , assimiler $Z(x)$ à la somme d'une dérive polynômiale $m(x)$ de degré k_0 et d'une représentation $Y(x)$ stationnaire de moyenne nulle et de covariance $C^0(h)$.

$$Z(x) = m(x) + Y(x) \quad \text{avec } m(x) = a_1 f^1(x)$$

la somme $a_1 f^1(x)$ étant étendue aux monômes de degré inférieur ou égal à k_0 .

En pratique on travaille généralement en voisinage glissant, et on utilise des dérivées d'ordre 0 à 2. En voisinage unique, on peut être conduit à utiliser des dérivées cubiques ou exceptionnellement des dérivées d'ordre supérieur. Mais dans ce cas il s'agit plutôt d'un modèle qu'on se fixe a priori (par exemple, pour estimer une régionale, mais sans trop savoir ce que l'on estimera en fait) que d'une hypothèse à vérifier. Nous étudierons ici essentiellement le choix entre $k_0 = 0, 1$ ou 2 , et nous donnerons des indications pour le cas où on veut également tester $k_0 = 3$.

2 - LA METHODE PROPOSEE

On propose ici une méthode qui consiste à estimer des points connus à partir d'autres données, et ceci pour différentes hypothèses sur le degré k . On cherche alors à déterminer le degré k_0 optimum par comparaison des erreurs quadratiques obtenues.

Pour préciser, considérons un point connu x_0 et d'autres données x_α , $\alpha = 1$ à n . Construisons alors un estimateur de $Z(x_0)$ à partir des $Z(x_\alpha)$ en prenant l'estimateur de krigeage associé à un degré k et à une covariance généralisée $K(h)$; en pratique, nous utiliserons, soit $K(h)$, soit la covariance $C(h)$ stationnaire sur D qui lui est associée. Notons tout de suite que le mieux est de prendre pour $K(h)$ ou $C(h)$ la covariance réelle $K^0(h)$ ou $C^0(h)$.

Mais celle-ci est généralement inconnue. En pratique on prendra alors soit l'effet de pépité, soit une covariance monômiale. Les coefficients λ^α de l'estimateur $Z^X = \lambda^\alpha Z_\alpha$ de Z_0 sont solution du système :

$$\begin{cases} \lambda^\beta C_{\alpha\beta} = \sum_{l=0}^k \mu_l f_\alpha^l + C_{0\alpha} & \alpha = 1, 2, \dots, n \\ \lambda^\alpha f_\alpha^l = f_0^l & l = 0, 1, \dots, k \end{cases}$$

(on utilise ici les notations synthétiques habituelles à une dimension)

L'espérance du carré de l'erreur $Z^X - Z_0$ est alors :

$$\sigma^2(k) = E[\lambda^\alpha Z_\alpha - Z_0]^2 = E[\lambda^\alpha (m_\alpha + Y_\alpha) - (m_0 + Y_0)]^2 = \sigma_1^2(k) + \sigma_2^2(k)$$

avec :

$$\sigma_1^2(k) = \lambda^\alpha \lambda^\beta C_{\alpha\beta}^0 - 2\lambda^\alpha C_{0\alpha}^0 + C_{00}^0$$

$$\sigma_2^2(k) = (\lambda^\alpha m_\alpha - m_0)^2$$

Voyons l'évolution de chacun de ces termes en fonction de k :

- pour $\sigma_1^2(k)$: Si la covariance $C(h)$ est la covariance réelle $C^0(h)$, la variance $\sigma_1^2(k)$ croît quand k augmente. En effet, à degré k fixé, la combinaison linéaire λ est, parmi celles de l'espace des mesures autorisées Λ_k , celle qui minimise la variance $\sigma_1^2(k)$. Comme $\Lambda_{k+1} \subset \Lambda_k$, $\sigma_1^2(k+1) \geq \sigma_1^2(k)$. Si par contre $C(h)$ n'est pas la covariance réelle, on peut avoir le phénomène inverse.

- pour $\sigma_2^2(k)$: deux cas se présentent :

a) $k \leq k_0$: les conditions d'universalité ne filtrent pas les monômes de degré supérieur à k . On en déduit :

$$\sigma_2^2(k) = \left[\sum_{l=k+1}^{k_0} a_l (\lambda^\alpha f_\alpha^l - f_0^l) \right]^2$$

b) $k \geq k_0$: la combinaison linéaire appartient à $\Lambda_k \subset \Lambda_{k_0}$. Donc $\sigma_2^2(k) = 0$

Conséquences

L'idéal serait de pouvoir caractériser k_0 par $\sigma^2(k_0) < \sigma^2(k) \forall k \neq k_0$. Les combinaisons linéaires λ_k utilisées devraient donc satisfaire les deux conditions suivantes :

- λ_k ne filtre pas les dérivées de degré supérieur à k , c'est-à-dire que $\lambda_k \in \Lambda'_k = \Lambda_k \cap \Lambda_{k+1}$, de façon que pour $k < k_0$, on ne puisse avoir $\sigma_2^2(k) = 0$.
- $\sigma_1^2(k)$ croît aussi peu que possible.

Ces conditions sont plus ou moins antagonistes. En effet si à la limite on trouvait des mesures telles que $\sigma_1^2(k+1) = \sigma_1^2(k)$, on aurait alors $\lambda_{k+1} = \lambda_k$; donc λ_k filtrerait les dérivées d'ordre $k+1$, ce qui est contraire à la première condition.

Il s'ensuit qu'on ne pourra jamais déceler une dérive très faible. Le problème consiste à trouver des combinaisons linéaires (c'est-à-dire des ensembles de points x_α et une covariance $C(h)$) telles que le seuil de détection de la dérive soit le plus faible possible.

3 - SCHEMA DE L'ETUDE

On se place d'abord dans le cas où la covariance $K^0(h)$ est connue. Le problème consiste alors à chercher les configurations de points x_0 et x_α qui permettent le mieux de détecter une dérive. Il ne s'agit pas d'une étude théorique, mais d'une étude pratique. On s'est placé en voisinage glissant et on a construit des voisinages de 16 points sélectionnés par octant ; ces points ont ensuite été classés par distance croissante au centre du voisinage, et on a alors étudié diverses configurations de points x_0 et x_α . Pour aboutir à une mise en oeuvre simple, on n'a considéré que des configurations du type : estimation du point n° m (après classement) par les points p à q ($m \in [p, q]$). Les 14 configurations considérées sont essentiellement des types suivants :

- estimation du point central par ses voisins ou par l'auréole externe ;
- estimation d'un point de l'auréole interne par l'auréole externe ;
- estimation d'un point de l'auréole externe par l'auréole interne, ou par tous les points qui sont plus près du centre que lui.

Les résultats sont évidemment relatifs aux données étudiées (du moins à leurs coordonnées x et y). Aussi a-t-on étudié quatre modes

d'implantation des données :

- maille aléatoire pure
- maille aléatoire stratifiée
- maille régulière
- profils discrétisés formant deux réseaux orthogonaux

Les résultats fournis sont des moyennes sur neuf voisinages

Cette première partie, où $K^0(h)$ est supposé connue, permet de réduire le nombre des configurations intéressantes. On étudie ensuite le cas où la covariance $K^0(h)$ n'est pas connue. On analyse l'influence de la covariance $K(h)$ utilisée dans la construction des mesures.

Avant d'exposer les résultats de cette étude, il nous faut expliciter les critères qui ont été utilisés pour caractériser les performances d'une configuration.

4 - CARACTERISATION DES PERFORMANCES D'UNE CONFIGURATION DE POINTS

A - Les variances élémentaires

Soit k_0 le degré réel, inconnu. Pour comparer les résultats obtenus dans les hypothèses k et k_0 , nous utiliserons les variances $\sigma^2(k)$ et $\sigma^2(k_0)$, qui se décomposent chacune en deux termes :

- $\sigma_1^2(k)$ et $\sigma_1^2(k_0)$ permettent d'apprécier la variation de la variance d'estimation, en l'absence de dérive, selon le nombre de conditions d'universalité.

- $\sigma_2^2(k)$ mesure la part de la dérive d'ordre k_0 qui n'est pas filtrée par les combinaisons linéaires d'ordre k . On a bien sûr $\sigma_2^2(k) = 0$ dès que $k \geq k_0$. Dans le cas contraire, $\sigma_2^2(k)$ dépend des coefficients a_1 de la dérive. Nous considérerons ici la variance $\sigma_2^2(k)$ associée au monôme $f^{k_0}(x) = x^{k_0}$.

(N.B. - En fait, à deux dimensions, les termes de degré k_0 sont les k_0+1 monômes $x^m y^n$ tels que $m+n = k_0$; on prend pour $\sigma_2^2(k)$ la plus faible des variances associées à chacun des monômes. Ce choix est motivé par la raison suivante : si on avait :

$$m(x) = \sum_{m+n=k_0} a_m x^m y^n$$

On aurait alors :

$$\sigma_2^2(k) = \left[\sum_{m+n=k_0} a_m (\lambda^\alpha x_\alpha^m y_\alpha^n - x_0^m y_0^n) \right]^2 = \left[\sum_m a_m T^m \right]^2$$

$$\text{avec } T_m = \lambda^\alpha x_\alpha^m y_\alpha^n - x_0^m y_0^n \quad m+n = k_0$$

Si on considère seulement les vecteurs a des coefficients a_m de norme 1 ($\sum a_m^2 = 1$) et tels que $a_m T^m > 0 \quad \forall m$, la dérive la plus filtrée est celle qui correspond à $\sigma_2^2(k)$ minimum, c'est-à-dire à $a_m = \pm \delta_m^{m_0}$, où m_0 est défini par : $|T^{m_0}| > |T^m| \quad \forall m \neq m_0$.

B - Le seuil de détection de la dérive

A partir des variances précédentes, on peut construire d'autres critères. En particulier nous utiliserons fréquemment, pour $k < k_0$, le seuil de détection de la dérive défini par :

$$s(k) = \frac{\sigma_1^2(k_0) - \sigma_1^2(k)}{\sigma_2^2(k)}$$

Plus s est faible, plus une dérive de faible amplitude a des chances d'être détectée. (Si on ne travaille pas avec la vraie covariance $C^0(h)$, on peut avoir s négatif ; nous reviendrons là dessus plus loin).

Comme les variances, s(k) dépend également de k_0 , $C^0(h)$, $C(h)$. Comme l'ordre de grandeur de $\sigma_2^2(k)$ dépend grandement de k et k_0 , on ne peut comparer directement les seuils obtenus pour des couples (k, k_0) différents. Il est donc bon de chercher l'ordre de grandeur des variations de $\sigma_2^2(k)$ en fonction de k et k_0 .

C - Ordre de grandeur de $\sigma_2^2(k)$

$\sigma_2^2(k)$ est l'erreur quadratique commise en approximant la dérive réelle d'ordre k_0 par un polynôme de degré $k < k_0$. Comme ici on a travaillé sur des voisinages couvrant le carré $\left[-\frac{1}{2}, +\frac{1}{2}\right] \times \left[-\frac{1}{2}, +\frac{1}{2}\right]$, on a calculé à une dimension l'erreur quadratique σ_2^2 que l'on commet en approximant sur $\left[-\frac{1}{2}, \frac{1}{2}\right]$ le monôme x^{k_0} par un polynôme de degré $k < k_0$ déterminé par moindres carrés. On a :

$$I(a_0, \dots, a_k) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(x^{k_0} - \sum_{l=0}^k a_l x^l \right)^2 dx$$

Les a_1 sont déterminés par

$$\frac{\delta I}{\delta a_1} = 0 \quad \forall l = 0, 1, \dots, k$$

On en déduit alors le minimum de I , c'est-à-dire précisément la variance σ_2^2 cherchée.

A deux dimensions, on obtiendrait évidemment les mêmes résultats pour les monômes extrêmes x^{k_0} et y^{k_0} . On peut faire les calculs pour les monômes intermédiaires ; on trouve alors des résultats comparables. On trouvera ces divers calculs en annexe. Pour ce qui nous intéresse, les résultats qui suivent nous suffiront :

approximation de x par a	:	variance :	$\frac{1}{12}$
" " de x^2 par $a+bx$:	" :	$\frac{1}{180}$
" " de x^2 par a	:	" :	$\frac{1}{180}$
" " de x^3 par $a+bx+cx^2$:	" :	$\frac{1}{2800}$
" " de x^3 par $a+bx$:	" :	$\frac{1}{2800}$
" " de x^3 par a	:	" :	$\frac{1}{448}$

On en déduit qu'un seuil de l'ordre de 1 pour $k=0$ et $k_0=1$ est de qualité comparable à un seuil de l'ordre de $\frac{2800}{12}$ pour $k=2$ et $k_0=3$. En d'autres termes, si on ne veut pas laisser passer une dérive linéaire de pente α , une sévérité analogue impose de ne pas laisser passer une dérive en x^2 de coefficient $\sqrt{15}\alpha$ ou en x^3 de coefficient $\sqrt{\frac{700}{3}}\alpha$.

5 - DETERMINATION DU DEGRE k_0 QUAND LA COVARIANCE REELLE EST CONNUE

Nous supposons $K^0(h)$, ou $C^0(h)$, connu. De plus nous nous limitons à $k_0 < 2$.

Éliminons tout de suite le cas des données sur profils : dans ce cas il faut sélectionner un nombre plus réduit de données, ou travailler différemment (une étude à une dimension le long des profils est particulièrement simple et utile, surtout pour la connaissance de la covariance à l'origine). En effet, si on prend toutes les données, il est impossible de travailler avec $k=2$: on observe fréquemment des systèmes singuliers, ou des variances 1000 ou 10000 fois plus fortes que pour $k=0$; les seuls cas où la variance augmente peu correspondent à des configurations bien symétriques, qui justement filtrent très bien des dérives de degré supérieur.

Nous limiterons donc le reste de l'étude aux trois autres types d'implantation des données. Sur chacun d'eux, nous avons étudié les quatorze types de configuration de points exposés au § 3. Rappelons seulement que les configurations sont prises dans des voisinages de 16 points, et que ces 16 points sont classés par distance croissante au centre du voisinage.

Les résultats sont assez similaires pour les différentes mailles. On observe tout d'abord que les configurations de krigeage d'un point intérieur par des points extérieurs donnent de mauvais résultats ; on a essayé ici :

- estimation du point n°1 par les points 2 à 8, ou 2 à 12, ou 9 à 15, ou 6 à 16.
- estimation du point n°5 par les points 9 à 15, ou 6 à 16.

Dans tous les cas, la variance de krigeage $\sigma_1^2(k)$ augmente peu avec k car le voisinage entoure bien le point krigé, mais la variance $\sigma_2^2(k)$ liée à une dérive de degré $k_0 > k$ décroît encore plus lentement. Il s'ensuit qu'on ne peut détecter que les dérives linéaires fortes ; quant aux dérives quadratiques, il est encore moins question de les diagnostiquer.

Pour le choix entre $k = 0$ et le degré réel $k_0 = 1$, les autres configurations sont de qualités assez voisines, les meilleures performances étant tout de même fournies par l'estimation d'un point extérieur par un assez grand nombre de points de l'auréole interne (9 au minimum). S'il s'agit par contre de choisir entre $k=0$ ou 1 et $k_0 = 2$, le classement des différentes configurations est le même, mais les écarts s'amplifient. Le tableau 1 donne à titre indicatif les performances des différentes configurations par rapport à la meilleure, dans le cas de la maille aléatoire stratifiée et d'une covariance pépitique ou linéaire. Cette performance est le quotient du seuil de détection de la dérive par le seuil associé à la configuration la meilleure (estimation de point n° 16 par tous les autres). Les bonnes performances sont donc celles qui sont proches de 1.

Une performance égale à α signifie que la configuration ne permet de diagnostiquer que les dérives $\sqrt{\alpha}$ fois plus fortes que les plus petites dérives diagnostiquées par la configuration la meilleure.

configuration	$K(h)=K^0(h)$	δ			$-h$		
	$k-k_0$	0-1	0-2	1-2	0-1	0-2	1-2
8 par 9 à 15		1	49	27	2	9	8
10 par 1 à 7		3	57	89	2	17	22
10 par 1 à 9		2	9	11	2	6	7
13 par 1 à 7		4	50	64	3	12	15
13 par 1 à 11		1	3	4	2	1	2
16 par 1 à 7		3	24	28	2	7	7
16 par 1 à 11		1	3	4	2	2	2
16 par 1 à 15		1	1	1	1	1	1

Tableau 1 : Performances des différentes configurations par rapport à la meilleure (16 par 1 à 15).

Une performance est d'autant meilleure qu'elle est faible.

N.B. : configuration "16 par 1 à 15" signifie : estimation du point n° 16 par les points 1 à 15.

Pour la suite, nous ne retiendrons que les quatre configurations les meilleures. Le tableau 2 donne les seuils de détection de la dérive pour ces quatre configurations et pour les différentes mailles de données. Pour apprécier les valeurs numériques, il faut se souvenir qu'on travaille en première approximation sur un domaine de diamètre 1 : ainsi $K(h)$ varie, au signe près, entre 0 et 1, de même que les monômes $f^1(x)$. Il s'ensuit que si pour $k=0$ et $k_0 = 1$ le seuil est égal à 0.01, on diagnostiquera une dérive $m(x)=0.1x$, ce qui est excellent compte tenu de l'amplitude des fluctuations permises par $K(h)$, même en l'absence de dérive.

configuration $k^0(h)=k(h)$	10 par 1 à 9			13 par 1 à 11			16 par 1 à 11			16 par 1 à 15		
	a	b	c	a	b	c	a	b	c	a	b	c
0-1	2	2	4	1	1	1	1	1	1	1	1	1
0-2	219	103	223	64	30	37	69	38	26	26	12	9
1-2	175	79	225	48	26	22	53	30	33	27	7	10
0-1	5	3	6	5	3	3	3	3	3	3	2	2
0-2	99	57	72	32	14	19	33	18	24	24	10	7
1-2	62	46	61	27	13	15	26	15	25	23	7	6
1-2	28	30	23	27	11	20	33	14	41	27	15	11

Tableau 2 - Seuil de détection de la dérive (multiplié par 1000) ; on utilise ici la covariance réelle $K^0(h)$

Les résultats sont donnés pour les différents types d'implantation des données:

- a : maille aléatoire pure
- b : maille aléatoire stratifiée
- c : maille régulière

6 - CONSEQUENCES DE L'UTILISATION D'UNE COVARIANCE AUTRE QUE LA COVARIANCE REELLE.

Souvent on ne connaît pas encore la covariance réelle $K^0(h)$. De plus, dans certains cas, et notamment lorsqu'on traite simultanément un grand nombre de variables, il est plus simple de construire les mesures à l'aide d'une covariance $K(h)$ qu'on se donne a priori, par exemple l'effet de pépite pur ou une covariance linéaire. Mais alors la détermination du degré k_0 se complique. En effet, si on ne travaille plus avec la covariance réelle, la variance d'estimation $\sigma_1^2(k)$ n'est plus nécessairement croissante. Elle peut par exemple décroître entre $k=0$ et $k=1$ et inciter à faire intervenir une dérive linéaire même s'il n'y en a pas. Si on pense que le degré réel k_0 est caractérisé par $\sigma^2(k_0) \leq \sigma^2(k) \quad \forall k \neq k_0$, on risque donc de commettre plus d'erreurs que précédemment. On peut évidemment mettre en oeuvre des critères plus serrés, surtout dans le cas où on connaît $K^0(h)$, mais cela ne supprimera pas tous les risques d'erreur. Nous allons donc étudier les conséquences des différents modèles $K(h)$ utilisés ici (δ , $-h$, h^3) sur les mêmes modèles de covariance réelle $K^0(h)$, et ceci pour les quatre types de configuration retenus.

A) Covariance $K(h)$ pépitique

Les différentes configurations et les différentes mailles d'échantillonnage donnent des résultats similaires:

- Si la covariance réelle $K^0(h)$ n'a pas d'effet de pépite, mais est continue, voire dérivable, la covariance pépitique $K(h)$ ne pourra rendre compte de la régularité de la structure ; ceci sera compensé par l'introduction d'une dérive : la variance d'estimation $\sigma_1^2(k)$ va commencer par décroître. On risque donc d'être conduit à des degrés de dérive trop élevés si on associe k_0 à la variance $\sigma^2(k)$ minimum. Ainsi pour une covariance linéaire sans dérive, on trouvera $k_0=1$; pour une covariance en h^3 , on trouvera $k_0=2$.

- Si, par contre, il y a effectivement une dérive, et si la covariance réelle n'est pas un effet de pépite pur, le seuil de détection de la dérive sera plus faible : ainsi pour une dérive quadratique et une covariance linéaire, le seuil est environ divisé par 2.

B) Covariance K(h) linéaire

Les résultats sont encore communs aux différents types de maille et de configuration.

- Si la covariance réelle est un effet de pépite pur et s'il y a une dérive, la covariance linéaire va prendre à son compte une partie de la dérive. Il s'ensuit que le seuil de détection de la dérive sera plus élevé : 5 à 10 fois plus pour une dérive linéaire, environ deux fois plus pour une dérive quadratique.

- Si, par contre, la covariance réelle est en h^3 , on est pratiquement assuré, comme précédemment, de trouver $k_0=2$.

C) Covariance K(h) en h^3

On a étudié à titre indicatif son influence pour l'étude du choix entre dérive linéaire et dérive quadratique. Elle a évidemment pour conséquence d'augmenter, en général, le seuil de détection de la dérive : pour l'effet de pépite, il est multiplié par 2 à 50, et pour une covariance linéaire par 2 à 15.

D) Conclusion

- lorsqu'on fait le choix de k_0 selon la méthode évoquée plus haut ($\sigma^2(k_0) \leq \sigma^2(k) \quad \forall k \neq k_0$) :

Si l'on préfère augmenter le degré de la dérive plutôt que l'abaisser, il faut prendre pour $K(h)$ l'effet de pépite pur. Dans le cas contraire, mieux vaut utiliser une covariance linéaire, ou éventuellement en h^3 .

- lorsque l'on connaît la vraie covariance $K^0(h)$ et qu'on veut faire un choix plus fin, on peut calculer les variances d'estimation $\sigma_1^2(k)$. La comparaison de ces variances avec les variances totales $\sigma^2(k)$ permet de déterminer s'il y a une dérive ou non. Le tableau 3 donne l'évolution de la variance d'estimation pour les différents types de mesures en fonction du degré k et des covariances $K(h)$ et $K^0(h)$. Ce ne sont bien sûr que des indications. Il est préférable de calculer ces variances pour les mesures effectivement construites et pour les covariances $K^0(h)$ qui nous intéressent.

configuration		10 par 1 à 9			13 par 1 à 11			16 par 1 à 11			16 par 1 à 15		
K(h)	$K^0(h)$	a	b	c	a	b	c	a	b	c	a	b	c
	0	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
	1	1.67	1.70	1.57	1.63	1.58	1.42	1.81	1.75	1.59	1.38	1.37	1.32
	2	16.77	6.59	5.22	10.59	5.98	3.33	17.91	10.46	5.98	5.39	3.21	2.43
δ	0	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
	1	0.76	0.67	0.66	0.68	0.67	0.61	0.83	0.80	0.72	0.61	0.64	0.58
	2	3.96	1.70	1.42	2.30	1.50	0.95	3.78	2.53	1.59	1.49	0.98	0.70
h^3	1	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
	2	1.12	0.60	0.51	0.65	0.53	0.40	0.79	0.70	0.53	0.54	0.38	0.32
	0	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
δ	1	1.35	1.29	1.22	1.28	1.23	1.15	1.45	1.36	1.27	1.17	1.16	1.12
	2	13.74	4.45	3.67	7.55	4.16	2.37	12.91	7.64	4.08	4.17	2.35	1.79
$-h$	0	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
	1	1.29	1.25	1.18	1.26	1.18	1.12	1.36	1.27	1.22	1.20	1.15	1.10
	2	6.90	3.58	2.79	4.82	3.03	2.02	6.45	4.31	3.04	3.31	2.16	1.63
h^3	1	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
	2	1.25	0.67	0.74	0.95	0.77	0.67	0.93	0.87	0.77	0.74	0.67	0.60
	1	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
δ	2	3.08	1.54	1.50	1.84	1.65	1.26	2.05	1.94	1.71	1.53	1.39	1.18
$-h$	1	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
	2	2.81	1.71	1.53	2.00	1.58	1.29	2.10	1.85	1.51	1.50	1.35	1.18
h^3	1	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
	2	1.81	1.48	1.34	1.55	1.33	1.24	1.62	1.45	1.39	1.38	1.27	1.16

Tableau 3. Evolution de la variance d'estimation $\sigma_1^2(k)$, ramenée à 1 pour $k=0$ (ou $k=1$ pour une covariance en h^3)
 Les différentes mailles d'implantation des données sont : a : maille aléatoire pure
 b : maille aléatoire stratifiée
 c : maille régulière

7 - QUELQUES INDICATIONS POUR DES DERIVES CUBIQUES

On a vu que pour diagnostiquer des dérives quadratiques, les meilleures mesures font intervenir 12 à 16 points, la limite semblant correspondre à l'estimation d'un point extérieur par 10 autres, soit autant qu'il y a de conditions d'universalité pour $k=3$. On peut en déduire que, pour une dérive cubique, la limite serait l'estimation du point n°16 par les 15 autres.

Ceci est confirmé par l'étude des mesures précédentes : Les deux autres configurations autorisées pour $k=3$ donnent un seuil de détection d'une dérive cubique qui, pour une covariance pépitique ou linéaire, est en général 8 à 100 fois plus élevé que la configuration d'estimation du point n°16 par les 15 autres.

A - Cas où les mesures sont construites à l'aide de la covariance réelle $K^0(h)$

Le tableau 4 donne les seuils que l'on obtient avec cette configuration, quand on travaille avec la vraie covariance $K^0(h)$. Compte tenu de ce qui a été vu en 4-C, les résultats obtenus pour $k_0=3$ sont bons.

configuration		16 par 1 à 15		
$K^0(h)=K(h)$	$k-k_0$	Maille aléatoire pure	Maille aléatoire stratifiée	Maille régulière
δ	0-1	1	1	1
	0-2	26	12	9
	1-2	27	7	10
	0-3	294	51	21
	1-3	2522	528	261
	2-3	714	319	534
-h	0-1	3	2	2
	0-2	24	10	7
	1-2	23	7	6
	0-3	271	91	34
	1-3	740	398	104
	2-3	500	127	147
h^3	1-2	27	15	11
	1-3	113	59	33
	2-3	104	35	31
$-h^5$	2-3	190	81	114

Tableau 4 : Seuil de détection de la dérive (multiplié par 1000) ; on utilise ici la covariance réelle $K^0(h)$.

B - Cas où les mesures sont construites à l'aide d'une covariance $K(h)$ fournie a priori.

Si pour construire les mesures, on utilise le système de krigeage associé à un effet de pépite ou à une covariance linéaire, on a les résultats suivants :

- Si la vraie covariance est pépitique ou linéaire, le seuil de détection de la dérive reste assez faible, il est cependant meilleur avec les mesures construites sur l'effet de pépite.

- Si la vraie covariance est en h^3 , le seuil est très faible (de l'ordre de 10), si bien qu'on doit avoir un très bon diagnostic au niveau des dérives cubiques.

- Si la vraie covariance est en h^5 , alors la variance d'estimation $\sigma_1^2(k)$ décroît au moins jusqu'à $k=3$, et si on n'y prend garde, on diagnostique une dérive cubique même si elle n'est que quadratique.

On peut donc reconduire à ce sujet les conclusions du §6. En particulier si on ne veut pas sous-estimer le degré k_0 de la dérive, le mieux est encore d'utiliser les mesures fondées sur l'effet de pépite.

Dans le cas où on veut affiner le choix de k_0 , on peut faire comme il a été proposé au §6. Le tableau 5 donne, à titre indicatif, l'évolution de la variance d'estimation $\sigma_1^2(k)$ jusqu'à $k=3$.

configuration 16 par 1à15.	K(h)	δ			-h			h^3			-h ⁵		
		a	b	c	a	b	c	a	b	c	a	b	c
	Maille												
	0	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
δ	1	1.38	1.37	1.32	1.17	1.16	1.12	1.	1.	1.	1.	1.	1.
	2	5.39	3.21	2.43	4.17	2.35	1.79	1.53	1.39	1.18	1.	1.	1.
	3	112.47	19.68	6.94	75.61	13.62	4.74	8.19	3.17	1.81	1.89	1.55	1.29
	0	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
-h	1	0.61	0.64	0.58	1.20	1.15	1.10	1.	1.	1.	1.	1.	1.
	2	1.49	0.98	0.70	3.31	2.16	1.63	1.50	1.35	1.19	1.	1.	1.
	3	14.64	3.93	1.50	35.08	9.07	3.73	7.65	3.07	1.82	2.53	1.51	1.28
	1	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
h^3	2	0.54	0.38	0.32	0.74	0.67	0.60	1.38	1.27	1.16	1.	1.	1.
	3	1.78	0.46	0.32	2.74	1.26	0.77	4.22	2.32	1.67	2.43	1.49	1.28
	2	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
-h ⁵	3	0.80	0.40	0.27	0.86	0.48	0.36	1.10	0.77	0.68	1.66	1.36	1.23

Tableau 5 : Evolution de la variance d'estimation $\sigma_1^2(k)$, ramenée à 1 pour $k=0$, ou éventuellement 1 ou 2 suivant

la covariance $K^0(h)$. Les différentes mailles d'implantation des données sont :

- a : maille aléatoire pure
- b : maille aléatoire stratifiée
- c : maille régulière

8 - CONSEQUENCES POUR LA MISE AU POINT D'UNE METHODE AUTOMATIQUE DE DETERMINATION DU DEGRE k_0 .

De ce qui précède, on peut dégager une méthode de détermination automatique du degré k_0 de la dérive. Pour cela on construira des voisinages de 11 à 16 points (ou de 16 à 21 points, si on envisage la possibilité d'une dérive cubique). Si on veut déterminer la dérive sur des voisinages plus grands, voire sur l'ensemble du champ si les données sont assez peu nombreuses, il suffit de subdiviser ces grands voisinages en sous voisinages de 16 points environ. Dans chaque voisinage, on classe les points par distance croissante au centre, et on fait l'estimation des points les plus extérieurs par les points intérieurs.

Le mieux semble être d'estimer, pour les différents degrés k , chacun des trois ou quatre derniers points par ceux qui sont avant lui. Mais pour économiser du temps calcul, il est préférable de traiter les points par paquets et par exemple d'estimer les points 14, 15 et 16 par les points 1 à 13. Et si on dispose seulement de peu de voisinages, on peut encore estimer les points 11, 12 et 13 par les points 1 à 10. (Tout ceci pour le cas où on s'arrête à $k=2$; si on pousse jusqu'à $k=3$, on estimerait par paquets les points 16 à 21).

Reste le problème de la covariance à utiliser pour construire les mesures. Si on connaît la vraie covariance, autant l'utiliser. Mais cela est rare. Dans le cas contraire, l'effet de pépète fournit globalement les meilleurs résultats quand on détermine k_0 par la condition $\sigma^2(k) > \sigma^2(k_0) \forall k \neq k_0$. On peut évidemment calculer les variances d'estimation $\sigma_1^2(k)$ associées aux différentes covariances élémentaires et utiliser ces résultats pour choisir k_0 en fonction des covariances $K^P(h)$ possibles selon k_0 . Mais c'est là un raffinement qu'il serait délicat d'automatiser et qui exige beaucoup de mesures et des résultats bien représentatifs de la distribution des données. On sait, en effet, que les variances sont très sensibles aux données extrêmes, et encore plus aux données erronées. Tant qu'on ne considère que les variances $\sigma^2(k)$, on peut remplacer le critère précédent par un critère fondé sur les scores : pour comparer k et k_0 , on comptera le nombre de fois où le degré k_0 aura fourni un meilleur résultat que le degré k , et le nombre de fois où le degré k aura été meilleur que k_0 . Un tel critère a l'avantage d'être peu sensible aux données erronées. Mais il serait difficile à généraliser au cas de la méthode plus fine faisant intervenir les termes $\sigma_1^2(k)$.

II - INFERENCE AUTOMATIQUE D'UNE COVARIANCE GENERALISEE

1 - DELIMITATION DU PROBLEME

Soit $Z(x)$ une représentation de FAI-k de covariance généralisée $K(h)$, dont on connaît une réalisation (que nous noterons également $Z(x)$) en certains points expérimentaux x_α . Nous supposons l'ordre k connu. Par contre $K(h)$ est inconnue. Diverses méthodes (variogramme des résidus, par exemple) permettent d'en faire l'inférence dans certains cas particuliers. Nous ne nous intéresserons pas à elles, mais à des méthodes qui soient applicables quelle que soit la configuration des données. Mais pour pouvoir mettre en oeuvre de telles méthodes sans effectuer de calculs prohibitifs, on doit restreindre la classe des covariances possibles. On sait que les covariances polynômiales, où les paramètres interviennent linéairement, constituent un modèle particulièrement intéressant. Nous supposerons ici que $K(h)$ est de la forme :

$$K(h) = C_0 \delta + \sum_{p=0}^k (-1)^{p+1} b_p h^{2p+1}$$

L'inférence de $K(h)$ se ramène à celle de C_0 (effet de pépite) et des b_p . Nous ne proposons pas ici d'autres méthodes que celles qui ont été mises en oeuvre dans les programmes KRIGEPACK et BLUEPACK (qui ne diffèrent que sur des points secondaires) et qui ont déjà été publiées. Celles-ci reposent sur la comparaison de variances théoriques à des variances expérimentales (de la forme $Z(\lambda)^2$, où λ désigne une combinaison linéaire autorisée). Ici nous donnerons des indications sur le choix et le nombre des combinaisons linéaires à construire, et sur la précision des résultats de l'inférence statistique. Mais, auparavant, voyons comment on détermine les paramètres C_0 et b_p .

2 - INFERENCE DES PARAMETRES DE LA COVARIANCE

La méthode utilisée n'impose pas vraiment que la covariance $K(h)$ soit polynômiale. Il faut simplement qu'elle s'exprime comme combinaison linéaire de modèles élémentaires $K^p(h)$:

$$K(h) = \sum_p b_p K^P(h) \quad b_p \geq 0 \quad \forall_p$$

En pratique évidemment, ces modèles $K^P(h)$ sont habituellement $\delta, -h, h^3 \dots$

Pour déterminer les b_p , on procède de la manière suivante : sur des voisinages analogues à ceux que l'on utilisera pour le krigeage, on construit, à partir des données expérimentales, des mesures appartenant à l'espace Λ_k des combinaisons linéaires autorisées à l'ordre k . Pour chaque mesure λ , attribuant les poids λ^α aux points expérimentaux x_α , on calcule :

- la combinaison linéaire $Z(\lambda) = \sum_\alpha \lambda^\alpha Z(x_\alpha)$

- sa variance expérimentale $V(\lambda) = Z(\lambda)^2$

- les variances théoriques élémentaires $K^P(\lambda) = \sum_\alpha \sum_\beta \lambda^\alpha \lambda^\beta K^P(x_\alpha - x_\beta)$

associées aux différentes covariances élémentaires $K^P(h)$

Ces différentes variances sont liées par la relation :

$$E [V(\lambda)] = K(\lambda) = \sum_p b_p K^P(\lambda)$$

On construit un certain nombre de mesures λ , que nous indiquerons par i . Pour déterminer les b_p , on choisit alors ceux qui minimisent :

$$Q(b) = \sum_i \bar{\omega}_i \left[V(\lambda_i) - \sum_p b_p K^P(\lambda_i) \right]^2$$

où b désigne le vecteur des b_p , et où les $\bar{\omega}_i$ sont des pondérateurs ayant pour but de ne pas attribuer trop de poids aux termes qui ont une forte variance. En pratique, on prend pour $\bar{\omega}_i$ l'inverse de $[K^q(\lambda_i)]^2$, q désignant un modèle élémentaire particulier (nous reviendrons sur son choix).

Cette condition de moindres carrés se résoud simplement. Mais on n'est pas assuré que tous les termes b_p obtenus soient positifs. Si l'un des termes est négatif, on recherche les différentes solutions

comportant un terme nul, ^{et} on retient celle qui minimise $Q(b)$; si aucune n'est admissible, on recherche les solutions comportant deux termes nuls, et ainsi de suite. On est sûr d'aboutir à une solution admissible car les solutions à un seul terme fournissent toujours un coefficient positif.

3 - PRECISION DES RESULTATS OBTENUS

A - Variance d'estimation des paramètres b_p

Moyennant une hypothèse sur les moments d'ordre 4 de la fonction aléatoire, on peut calculer la variance d'estimation des paramètres b_p . En effet, c'est en fait un estimateur b_p^* de b_p que l'on obtient en minimisant la forme quadratique Q . La minimisation de $Q(b)$ conduit au système :

$$\sum_i \bar{\omega}_i \left[V(\lambda_i) - \sum_q b_q^* K^q(\lambda_i) \right] K^p(\lambda_i) = 0 \quad \forall p \in \mathcal{D}$$

où \mathcal{D} désigne l'ensemble des modèles de covariance élémentaires pris en considération.

$$\begin{aligned} \text{Posons : } C^{pq} &= \sum_i \bar{\omega}_i K^p(\lambda_i) K^q(\lambda_i) \\ B^p &= \sum_i \bar{\omega}_i K^p(\lambda_i) V(\lambda_i) \end{aligned}$$

On en déduit :

$$b_p^* = D_{pq} B^q$$

où la matrice $[D_{pq}]$ est l'inverse de la matrice $[C^{pq}]$.

On en déduit :

$$E \begin{bmatrix} b_p^* \end{bmatrix} = D_{pq} \sum_i \bar{\omega}_i K^q(\lambda_i) b_r K^r(\lambda_i) = b_r D_{pq} C_{qr} = b_r \delta_{pr} = b_p$$

$$\text{Var} \begin{bmatrix} b_p^* \end{bmatrix} = D_{pq} D_{pr} \text{cov}(B^q, B^r)$$

$$= D_{pq} D_{pr} \sum_i \sum_j \bar{\omega}_i \bar{\omega}_j K^q(\lambda_i) K^r(\lambda_j) \text{cov}(V(\lambda_i), V(\lambda_j))$$

Pour continuer les calculs, on a donc besoin des moments d'ordre 4 de la FAI-k. Nous ferons l'hypothèse qu'elle est gaussienne. On sait que si X_1, X_2, X_3, X_4 sont des variables aléatoires gaussiennes d'espérances nulles et de covariances σ_{ij} , on a :

$$E \begin{bmatrix} X_1 & X_2 & X_3 & X_4 \end{bmatrix} = \begin{matrix} \sigma_{12} & \sigma_{34} & + & \sigma_{13} & \sigma_{24} & + & \sigma_{14} & \sigma_{23} \end{matrix}$$

On en déduit ici :

$$\begin{aligned} \text{cov} (V(\lambda_i), V(\lambda_j)) &= E [Z(\lambda_i)^2 Z(\lambda_j)^2] - E [Z(\lambda_i)^2] E [Z(\lambda_j)^2] \\ &= K(\lambda_i) K(\lambda_j) + 2K(\lambda_i, \lambda_j)^2 - K(\lambda_i) K(\lambda_j) \\ &= 2K(\lambda_i, \lambda_j)^2 \end{aligned}$$

$$\text{où } K(\lambda_i, \lambda_j) = \text{cov} (Z(\lambda_i), Z(\lambda_j)) = \sum_{\alpha} \sum_{\beta} \lambda_i^{\alpha} \lambda_j^{\beta} K(x_{\alpha} - x_{\beta})$$

Pour pouvoir effectivement calculer la variance de b_p^* , on a donc besoin de connaître $K(h)$, c'est-à-dire les coefficients b_p réels. Il semble qu'on se retrouve au point de départ. En fait, comme on cherche surtout un ordre de grandeur de $\text{Var}(b_p^*)$, on peut remplacer $K(h)$ par son estimation $K^*(h) = b_p^* K_p(h)$. De toute façon, on peut calculer $\text{Var}(b_p^*)$ dans l'hypothèse où $K(h)$ est une covariance donnée (par exemple un des modèles élémentaires).

B - Remarque sur la signification des coefficients b_p

Les coefficients b_p n'ont pas d'existence en soi ; ce sont seulement des paramètres du modèle. Ce qui nous intéresse, ce n'est d'ailleurs pas la connaissance précise de chaque b_p , mais c'est que la covariance $K^*(h) = b_p^* K_p(h)$ fournisse des variances d'estimation correctes pour le type de configuration que nous aurons à utiliser. Si les b_p^* ont une variance trop forte, il est donc possible que globalement on obtienne une covariance $K^*(h)$ acceptable (mais c'est loin d'être certain) ; cela signifiera cependant que le modèle comporte plus de paramètres qu'il n'est possible d'en inférer correctement, et qu'il vaut donc mieux se restreindre à un modèle moins riche.

C - Cas d'une covariance dépendant d'un seul paramètre

Dans le cas particulièrement simple, et favorable, où la covariance dépend d'un seul paramètre multiplicatif, on peut dégager un ordre de grandeur intéressant. Soit donc :

$$K(h) = b K_0(h)$$

On travaille en fait avec les variables réduites :

$$V'(\lambda_i) = \frac{Z(\lambda_i)^2}{K_0(\lambda_i)}$$

On accorde à chacune le même poids, ce qui reviendrait, sur les variables brutes, à prendre pour $\bar{\omega}_i$ l'inverse de $K_0(\lambda_i)^2$. On obtient alors :

$$b^* = \frac{1}{n} \sum_i V'(\lambda_i) = \frac{1}{n} \sum_i \frac{Z(\lambda_i)^2}{K_0(\lambda_i)}$$

où n désigne le nombre de mesures construites

On en déduit :

$$E \left[b^* - b \right]^2 = \frac{1}{n^2} \sum_i \sum_j \text{cov}(V'(\lambda_i), V'(\lambda_j))$$

Dans l'hypothèse gaussienne, on a :

$$\text{cov}(V'(\lambda_i), V'(\lambda_j)) = \frac{2 K(\lambda_i, \lambda_j)^2}{K_0(\lambda_i) K_0(\lambda_j)} = 2b^2 \frac{K_0(\lambda_i, \lambda_j)^2}{K_0(\lambda_i) K_0(\lambda_j)}$$

D'où :

$$E \left[b^* - b \right]^2 = \frac{2b^2}{n} \sum_i \sum_j \frac{K_0(\lambda_i, \lambda_j)^2}{K_0(\lambda_i) K_0(\lambda_j)}$$

Pour obtenir une variance faible, on a intérêt, si on raisonne à n fixé, à construire des combinaisons linéaires le moins corrélées possible. Si on construit n mesures non corrélées, on obtient :

$$\text{Var} \left[b^* \right] = \frac{2b^2}{n}$$

On voit dans ce cas, pourtant très favorable, qu'il faut tout de même construire $n = 200$ mesures sans corrélation pour obtenir un écart-type relatif de 0.1.

D - Cas général

A deux dimensions, et surtout pour une implantation irrégulière des données, le calcul des covariances $K(\lambda_i, \lambda_j)$ entre $Z(\lambda_i)$ et $Z(\lambda_j)$ est extrêmement lourd, car on utilise couramment plusieurs centaines de combinaisons linéaires, voire quelques milliers. Il n'est donc pas possible de calculer la variance des b_p^* lors de chaque étude pratique. Aussi allons-nous présenter une étude pour des données à maille régulière à une dimension. La transposition à une maille irrégulière à deux dimensions n'est pas évidente. On en tirera néanmoins des résultats qualitatifs intéressants concernant les pondérateurs $\bar{\omega}_i$ à utiliser, ainsi que les types et le nombre de combinaisons linéaires à construire.

4 - ETUDE DE LA VARIANCE D'ESTIMATION DES PARAMETRES DE LA COVARIANCE POUR DES DONNEES A MAILLE REGULIERE A UNE DIMENSION - GENERALITES.

A - Les combinaisons linéaires utilisées

Les données sont supposées implantées au pas 1. A l'aide de ces données, on a construit uniquement des accroissements d'ordre k fondés sur des points prélevés à maille 1 (l entier), c'est-à-dire de la forme :

$$Z(\lambda) = Z(x) - C_{k+1}^1 Z(x+1) + \dots + (-1)^P C_{k+1}^P Z(x+p1) + \dots + (-1)^{k+1} Z(x+(k+1)1)$$

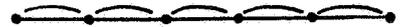
Par raison de symétrie, on prend pour origine de la mesure λ , non pas l'extrémité x mais le milieu $x + \frac{(k+1)1}{2}$. Et pour des raisons de commodité de programmation, on a considéré qu'on disposait, non pas de n points expérimentaux, mais de n origines possibles pour les mesures λ (c'est-à-dire que pour 1 donné, la construction de toutes les mesures possibles fait intervenir $n+(k+1)1$ points). Ceci a conduit également à se limiter à considérer seulement les valeurs impaires de l (sinon, pour k pair, on ferait intervenir deux réseaux de mesures différentes, décalés de $1/2$, selon que l est pair ou impair).

B - Quelques cas simples pour une covariance dépendant d'un seul paramètre

Dans quelques cas simples, on peut conduire les calculs sous forme explicite jusqu'au bout. Supposons donc que $K(h)$ est de la forme :

$$K(h) = b K_0(h)$$

a) Utilisation des mesures associées à $l = 1$



Pour estimer b , on construit toutes les mesures associées à $l = 1$. Comme on n'utilise qu'un type de combinaisons linéaires, le pondérateur \bar{w}_i n'a aucun rôle. Si on désigne par n le nombre d'origines possibles (c'est-à-dire approximativement le nombre de points si n est grand), on obtient :

- pour $k = 0$

$$K_0(h) = \delta \Rightarrow \text{var} [b^*] = \frac{3}{n} - \frac{1}{n^2} \approx \frac{3}{n}$$

$$K_0(h) = -h \Rightarrow \text{var} [b^*] = \frac{2}{n}$$

- pour $k = 1$

$$K_0(h) = \delta \Rightarrow \text{var} [b^*] = \frac{35}{9n} - \frac{2}{n^2} \approx \frac{3.89}{n}$$

$$K_0(h) = -h \Rightarrow \text{var} [b^*] = \frac{3}{n} - \frac{1}{n^2} \approx \frac{3}{n}$$

$$K_0(h) = h^3 \Rightarrow \text{var} [b^*] = \frac{9}{4n} - \frac{1}{4n^2} \approx \frac{2.25}{n}$$

- pour $k = 2$

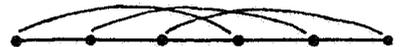
$$K_0(h) = \delta \Rightarrow \text{var} [b^*] = \frac{231}{50n} - \frac{3}{n^2} \approx \frac{4.62}{n}$$

$$K_0(h) = -h \Rightarrow \text{var} [b^*] = \frac{35}{9n} - \frac{2}{n^2} \approx \frac{3.89}{n}$$

$$K_0(h) = h^3 \Rightarrow \text{var} [b^*] = \frac{23}{9n} - \frac{2}{3n^2} \approx \frac{2.56}{n}$$

$$K_0(h) = -h^5 \Rightarrow \text{var} [b^*] = \frac{2855}{1089n} - \frac{226}{363n^2} \approx \frac{2.62}{n}$$

b) Utilisation des mesures associées à $l > 1$



On obtient des résultats plus mauvais (sauf pour l'effet

de pépite). Par exemple pour $k = 0$, on obtient :

$$K_0(h) = \delta \Rightarrow \text{var} [b^*] = \frac{3}{n} - \frac{1}{n^2} \approx \frac{3}{n}$$

$$K_0(h) = -h \Rightarrow \text{var} [b^*] = \frac{2(2l^2+1)}{3ln} - \frac{l^2-1}{3n^2} \approx \frac{4l}{3n} \quad \text{si } l > 1$$

c) Utilisation conjointe des mesures fondées sur la maille 1 et des mesures associées à $l > 1$.



Ici le pondérateur $\bar{\omega}_i$ intervient. On le prend égal à l'inverse de $[K_0(\lambda_i)]^2$. Malgré ce qu'on pourrait croire, la prise en compte de ces deux types de mesures n'améliore pas nécessairement les résultats. Toujours pour $k = 0$, et pour l impair, on obtient :

$$K_0(h) = \delta \Rightarrow \text{var} [b^*] = \frac{5}{2n} - \frac{3l+1}{4n^2} \approx \frac{5}{2n} \quad (\text{légère amélioration})$$

$$K_0(h) = -h \Rightarrow \text{var} [b^*] = \frac{2l^2+9l+1}{6ln} - \frac{(l^2-1)(l+3)}{12ln^2} \quad (\text{détérioration par rapport au § a)}$$

On retrouve le résultat bien connu concernant la précision d'un variogramme expérimental $\gamma^*(h)$: si $\gamma(h) = bh$, la variance de dispersion relative $\frac{E(\gamma^*(h) - \gamma(h))^2}{\gamma(h)^2}$ croît rapidement avec h . Ce qui a pour conséquence que pour ajuster la pente d'un schéma linéaire, on prend surtout en compte les premiers points du variogramme.

C - Evolution de la variance en fonction du nombre de données

Précisons tout de suite qu'on se situe dans l'optique d'un ajustement local de la covariance (disons pour $h < 10$ pour fixer les idées). Et on suppose qu'on dispose de plusieurs dizaines ou de plusieurs centaines de données. Si on se fixe une combinaison de mesures λ (c'est-à-dire qu'on construira par exemple toutes les mesures fondées sur une maille 1 ou 3), la variance d'estimation des coefficients b_p^* est en première approximation inversement proportionnelle au nombre de données. Cela tient à ce qu'on construit un nombre de mesures proportionnel au nombre de données, et à ce que la covariance entre $Z(\lambda_i)$ et $Z(\lambda_j)$ est nulle dès que les mesures

λ_i et λ_j ont des supports éloignés. Ceci est d'ailleurs confirmé par les résultats qui précèdent, et par des résultats plus complets pour $n = 100$ et $n = 400$ (rappelons que n est le nombre d'origines possibles pour la construction des mesures ; pour les cas qui nous intéressent, il diffère peu du nombre de données). Pour la suite de l'étude, on peut donc se fixer n . On a pris $n = 100$.

D - Quelques définitions

Par la suite, on appellera :

- type de mesures : les accroissements d'ordre k fondés sur des points prélevés à une maille l donnée; on parlera de mesures serrées si l est petit, de mesures larges si l est grand.

- combinaison : un ensemble de types de mesures ; par exemple la combinaison 1-3 contient les accroissements associés à $l = 1$ et à $l = 3$.

- modèle : un modèle de covariance généralisée dont on se donne les composantes $K_p(h)$, et dont on cherche les coefficients b_p ; ainsi δ -h désigne un modèle linéaire avec effet de pépité.

- pondérateur : le pondérateur $\bar{\omega}$ déjà défini, ou la covariance généralisée utilisée pour le calculer ; pondérer par $-h$ signifie que le pondérateur $\bar{\omega}$ est l'inverse du carré de la variance associée à un schéma linéaire de pente 1. (cf. §2).

E - Plan de l'étude

L'étude des données régulières à une dimension va être détaillée en trois temps correspondant à trois questions :

- Quelle combinaison de mesures utiliser, et comment choisir le pondérateur $\bar{\omega}$?
- Combien de paramètres peut-on raisonnablement estimer ?
- Combien construire de combinaisons linéaires ?

Ensuite, nous verrons quelles conséquences on peut en tirer pour une méthode d'inférence des coefficients b_p dans le cas des données

à implantation irrégulière dans le plan.

5 - CHOIX DU PONDERATEUR $\bar{\omega}$.

A - Introduction

Tant qu'on construit autant de types de mesures qu'il y a de coefficients à estimer, le résultat est indépendant du pondérateur (du moment que pour des mesures du même type, ce pondérateur est constant). On s'intéresse donc ici au cas où on construit plus de types de mesures qu'il n'y a de paramètres. Le mieux, dans le cadre de la formulation du § 2, est de prendre pour $\bar{\omega}_i$ l'inverse de l'espérance de $\left[V(\lambda_i) - K(\lambda_i) \right]^2$. Dans l'hypothèse gaussienne que nous avons faite pour pouvoir continuer les calculs, on a :

$$E \left[V(\lambda_i) - K(\lambda_i) \right]^2 = 2 K(\lambda_i)^2$$

Comme en pratique on recherche justement les paramètres de la covariance $K(h)$, on prend le pondérateur associé à une covariance qui n'est pas nécessairement la covariance de la FAI-k. Pour éviter de faire des calculs supplémentaires, on prend le pondérateur associé à l'un des modèles élémentaires $K^P(h)$.

Ici, on a considéré plusieurs combinaisons de mesures. Et, pour chacune d'elles, on a calculé les variances des b_p^* selon le pondérateur choisi ($\delta, -h, h^3$). A maille régulière, on pourrait bien sûr se contenter de prendre autant de types de mesures qu'il y a de paramètres à estimer. Mais il faut bien voir qu'à maille irrégulière, on aura en fait autant de types de mesures que de mesures, car on ne construira jamais deux combinaisons linéaires identiques.

Les combinaisons de types étudiées sont les suivantes :

1-3
1-5
1-9
1-3-9
1-3-5-9
1-3-5-7-9-11

B - Covariance dépendant d'un seul paramètre

On a déjà vu que la combinaison 1 - 1 donne de moins bons résultats que la combinaison 1 seule, sauf si la covariance est pépiti- que. Ce résultat est plus général. On aboutit en effet aux faits sui- vants :

- Pour les combinaisons à deux types, la variance varie relativement peu quels que soient la combinaison, le pondérateur et l'ordre k . Par rapport à la variance obtenue avec les mesures du type 1 seul, les cas extrêmes correspondent à une amélioration de 30%, et à une détérioration de 800 %.
- Les combinaisons à plusieurs types donnent des résultats intermédiaires entre ceux des combinaisons 1 - 3 et 1 - 9.
- Chaque covariance $K^P(h)$ est le pondérateur le mieux adapté au modèle de même covariance $K^P(h)$.
- Chaque covariance élémentaire est un mauvais pondérateur pour les modèles plus réguliers ; mais si le modèle est moins régulier que le pondérateur, celui-ci donne, des résultats aussi bons, ou à peine moins bons que le pondérateur associé au modèle.

Conclusion : Prendre le pondérateur associé au modèle recherché, ou à défaut à un schéma plus régulier.

C - Choix de deux types de mesures pour inférer deux paramètres.

Supposons que la covariance est la somme de deux schémas élémentaires. Et proposons-nous d'inférer ses deux coefficients à l'aide de deux types de mesures, de mailles l et $l > 1$ (le choix de la maille l s'explique bien : c'est l'équivalent du premier point du variogramme, qui est le point le mieux connu). Comment choisir l ? A priori, on peut formuler deux raisonnements contradictoires :

- il faut prendre l le plus grand possible, afin de bien différencier le comportement des deux schémas élémentaires.
- il faut prendre l le plus petit possible, afin de ne pas faire intervenir des termes de variance trop grande (rappelons que la variance de $V(\lambda) = Z(\lambda)^2$ est $2 K(\lambda)^2$).

On obtient en fait les résultats suivants :

- Si la covariance réelle est proche du pôle le plus irrégulier, la précision croît quand l augmente.
- Si la covariance réelle est proche du pôle le plus régulier, la précision décroît lorsque l croît.

D - Choix du pondérateur lorsqu'on utilise plusieurs types de mesures.

Supposons qu'il y a toujours deux coefficients à inférer, mais que l'on construit plus de deux types de mesures. Le pondérateur aura une influence sur les résultats. Mais laquelle ?

On peut déjà faire une remarque préliminaire :

- Si on prend un pondérateur plus régulier que la variable elle-même, on va attribuer un poids trop faible aux mesures larges. On perdra donc de l'information, et on fera l'inférence essentiellement à partir des mesures serrées.

- Si par contre, on prend un pondérateur moins régulier que la vraie covariance, l'inférence reposera surtout sur les mesures larges. On est alors dans un cas analogue au précédent, à ceci près que les mesures larges ont une variance plus grande que les mesures serrées.

Il est donc préférable de prendre un pondérateur trop régulier plutôt que l'inverse.

Dans ce qui suit, on envisage les différents modèles de covariance à deux termes, et, pour chacun d'eux, on examine la variance des b_p^* dans le cas où la vraie covariance est l'un des termes extrêmes.

a) modèle $\delta - h$

Pour les combinaisons à deux types, le rapport de la plus mauvaise variance à la meilleure est assez faible: 1 à 3 pour le coefficient du schéma réel (δ ou h), 1 à 5 pour le coefficient du schéma

complémentaire (-h ou δ). Quand on ajoute d'autres types de mesures, les conséquences sont à peu près les mêmes quel que soit l'ordre k, mais varient selon le pondérateur :

- δ apporte une amélioration si $K(h) = \delta$, mais une forte détérioration si $K(h) = -h$.
- -h reste moyen ou apporte une amélioration si $K(h) = -h$, et reste moyen si $K(h) = \delta$.
- h^3 (pour $k > 0$) est tantôt voisin de -h, tantôt un peu moins bon.

Conclusion : Pondérer par -h, ou à la rigueur par h^3 , mais pas par δ .

b) modèle (δ, h^3) ou (-h, h^3)

Pour les combinaisons à deux types, le rapport de la plus mauvaise variance à la meilleure reste faible pour le coefficient du schéma réel : 1 à 3. Par contre, la variance du terme complémentaire peut varier dans des proportions de 1 à 1000 lorsque celui-ci est un terme en h^3 . On s'intéressera donc plus particulièrement aux pondérateurs qui, lorsqu'on mélange plusieurs types de mesures, fournissent des variances du terme complémentaire proches de l'optimum des combinaisons à deux types, voire meilleures. Les résultats dépendent encore essentiellement du pondérateur et peu de k :

- les pondérateurs δ et -h donnent des résultats voisins de l'optimum si $K(h)$ n'a pas de terme en h^3 , mais conduisent à une forte variance du terme complémentaire si $K(h) = h^3$. Cette variance peut être 100 fois plus forte qu'avec une combinaison à deux types (1000 fois si le pondérateur est δ).

- le pondérateur h^3 reste toujours moyen (entre les variances associées aux combinaisons 1-3 et 1-9).

Conclusion : Le pondérateur habituellement le meilleur est sans doute -h, mais il peut conduire à une erreur forte si la vraie covariance est en h^3 . Si donc on n'est pas sûr que la covariance recherchée est loin d'un schéma en h^3 , il est préférable de pondérer par h^3 , qui est plus stable et minimise les risques d'erreur importante. (cf. figure 6).

Pondérateur	Combinaison	Modèle		
		δ	$-h$	h^3
néant	1	.0387	.0299	.0225
	3	.0383	.0449	.0641
	9	.0371	.116	.188
	1-3	.0291	.0385	.0640
	1-9	.0286	.114	.188
δ	1-3-9	.0256	.0864	.188
	1-3-5-7-9-11	.0223	.0967	.212
$-h$	1-3	.0353	.0237	.0632
	1-9	.0382	.0382	.188
	1-3-9	.0349	.0349	.188
	1-3-5-7-9-11	.0332	.0332	.201
	1-3	.0366	.0294	.0358
h^3	1-9	.0387	.0299	.0673
	1-3-9	.0386	.0294	.0619
	1-3-5-7-9-11	.0386	.0293	.0987

Variance $K(h)$ des différents types de mesure.

Variance d'estimation des coefficients d'une covariance à un paramètre lorsque ce coefficient b vaut 1 ; 100 mesures par type.

Variance d'estimation des coefficients d'un modèle de covariance formé de deux schémas lorsque la covariance réelle $K(h)$ est l'un des schémas extrêmes. Les deux chiffres sont les variances de chacun des deux coefficients (d'espérance 1 et 0 ou 0 et 1 selon le cas) ; 100 mesures par type.

Type	$K(h)$	δ	$-h$	h^3
1		6	4	8
3		6	12	216
5		6	20	1000
7		6	28	2744
9		6	36	5832
11		6	44	10648

Pondérateur	Combinaison	$\delta-h$			$\delta+h$			$-h+h^3$					
		δ	$-h$	h^3	δ	h^3	$-h$	h^3	h^3				
néant	1-3	.0670	.0211	.0550	.0937	.0402	.3E-4	.0571	.0669	.0358	.2E-3	.151	.0735
	1-9	.0441	.0013	.0787	.147	.0387	.4E-7	.272	.189	.0306	.5E-5	.625	.192
δ	1-3-9	.0364	.0012	.258	.166	.0295	.3E-7	30.1	.192	.0437	.6E-5	27.8	.213
	1-3-5-7-9-11	.0327	.0006	.591	.184	.0252	.8E-8	1210.	.231	.0790	.3E-5	424.	.287
$-h$	1-3-9	.0480	.0041	.0594	.0902	.0356	.5E-7	2.10	.187	.0252	.5E-5	9.16	.201
	1-3-5-7-9-11	.0444	.0018	.0662	.100	.0345	.2E-7	11.8	.204	.0303	.3E-5	74.1	.237
h^3	1-3-9	.0663	.0203	.0539	.0812	.0394	.8E-5	.135	.109	.0325	.5E-4	.358	.119
	1-3-5-7-9-11	.0618	.0147	.0487	.0817	.0380	.2E-5	.179	.133	.0314	.2E-4	.462	.140

Tableau 6 - Variance des coefficients b_p^* pour $k = 1$

E - Inférence de plus de deux paramètres

Si on recherche un modèle comprenant un effet de pépité, un terme linéaire et un terme en h^3 , le meilleur pondérateur est h^3 , cependant $-h$ est presque aussi bon.

F - Modèles comprenant un terme en $-h^5$

On observe que le pondérateur qui présente le moins de risque de grosse erreur est encore h^3 , et non $-h^5$ (sauf dans le cas où la vraie covariance ne comprend qu'un terme en $-h^5$, mais la différence reste faible) ; si le meilleur pondérateur n'est pas $-h^5$, cela est sans doute dû à la trop grande régularité de ce schéma, qui provoque des variations vraiment énormes de la variance selon la configuration considérée.

G - Conclusion : règle pratique

Lorsqu'on cherche à estimer les coefficients b_p d'une covariance de la forme $b_p K^P(h)$, le pondérateur qui présente le moins de risque de grosse erreur est celui qui est associé au schéma élémentaire $K^P(h)$ le plus régulier. A défaut, on peut également utiliser un pondérateur encore plus régulier (par exemple celui qui est associé à h^3 pour un modèle en $\delta - h$). Toutefois, il ne faut pas prendre un pondérateur dont le comportement à l'origine soit plus régulier que h^3 (sauf si tous les $K^P(h)$ étaient plus réguliers que h^3).

6 - COMBIEN DE PARAMETRES PEUT-ON INFERER ?

On considèrera que l'inférence est possible si son imprécision, caractérisée par l'écart-type des coefficients b_p^* n'a pas trop de répercussions sur les variances des combinaisons linéaires $Z(\lambda)$; on s'attachera surtout aux combinaisons formées de points rapprochés, qui correspondent mieux aux configurations de krigeage usuelles. En se fixant un seuil très lâche, on aboutit aux conclusions suivantes (cf. le tableau 7) :

- si la vraie covariance est à dominante irrégulière (c'est-à-dire essentiellement linéaire avec effet de pépité) et si les

Modèle	Variance des coefficients de :				
	K(h)	δ	-h	h^3	$-h^5$
δ	δ	.458E-01			
-h	-h		.378E-01		
h^3	h^3			.678E-01	
$-h^5$	$-h^5$.241
$\delta-h$	δ	.764E-01	.237E-01		
	-h	.533E-01	.946E-01		
$\delta+h^3$	δ	.464E-01		.106E-04	
	h^3	.438E-01		.979E-01	
$\delta-h^5$	δ	.458E-01			.944E-12
	$-h^5$.159E+02			.241
$-h+h^3$	-h		.408E-01	.101E-03	
	h^3		.139	.105	
$-h-h^5$	-h		.379E-01		.321E-10
	$-h^5$.136E+03		.241
h^3-h^5	h^3			.394E-01	
	$-h^5$.265E+04	
$\delta-h+h^3$	δ	.880E-01	.428E-01	.774E-05	
	-h	.746E-01	.154	.500E-04	
	h^3	.141E+02	.441E+02	.203E+00	
$\delta-h-h^5$	δ	.788E-01	.272E-01		.287E-11
	-h	.573E-01	.105		.235E-10
	$-h^5$.273E+06	.765E+06		.252
$\delta+h^3-h^5$	δ	.471E-01		.686E-04	.626E-10
	h^3	.361E-01		.872E-01	.162E-06
	$-h^5$.222E+04		.674E+04	.311E+00
$-h+h^3-h^5$	-h		.444E-01	.523E-03	.404E-09
	h^3		.116	.985E-01	.165E-06
	$-h^5$.775E+04	.782E+04	.315
$\delta-h+h^3-h^5$	δ	.103	.743E-01	.652E-04	.267E-10
	-h	.108	.256	.382E-03	.174E-09
	h^3	.123E+02	.401E+02	.296	.187E-06
	$-h^5$.114E+07	.365E+07	.369E+05	.381

Variance K(λ) des différents types de mesure (k=2)

Type	K(h)	δ	-h	h^3	$-h^5$
1		20	12	12	132
3		20	36	324	32076
5		20	60	1500	412500
7		20	84	4116	2218524
9		20	108	8748	7794468
11		20	132	15972	21258732

Tableau 7 - Variance des coefficients b^*
pour $k=2$

Variance d'estimation des coefficients des différents modèles de covariance lorsque la covariance réelle est l'un des schémas élémentaires du modèle (les coefficients réels sont donc tous nuls, sauf un qui est égal à 1)

pondérateur associé à h^3
combinaison 1-3-5-7-9-11
100 mesures par type

termes plus réguliers n'ont qu'un rôle accessoire, on peut faire l'inférence de trois ou quatre paramètres ; évidemment on perd en précision, mais dans des limites raisonnables.

- si par contre la vraie covariance est à dominante très régulière (h^3 ou $-h^5$), on ne peut estimer sérieusement que deux paramètres, et encore, si l'on s'agit d'un ajustement en h^3 et $-h^5$, on peut avoir des problèmes.

Conclusion : L'inférence de plus de deux paramètres ne donne des résultats exploitables que si la vraie covariance est à dominante pépétique et/ou linéaire (du point de vue de la contribution des différents termes dans le calcul des variances des mesures $Z(\lambda)$). C'est pourquoi on se limite usuellement à des modèles comprenant au maximum trois schémas élémentaires.

7 - COMBIEN DE MESURES FAUT-IL CONSTRUIRE ?

On a fait des essais en ne prenant pas toutes les combinaisons linéaires, mais en les prélevant à un certain pas. On aboutit à la conclusion qu'on perd peu en précision en ne prenant qu'une combinaison linéaire sur deux; si par contre, on n'en prend qu'une sur trois ou quatre, les résultats sont nettement plus mauvais (variances deux fois plus fortes). A une dimension, il faut donc au moins autant de types de mesures que de coefficients à estimer et il suffit de construire $\frac{n}{2}$ mesures de chaque type, n désignant le nombre de points; comme on estime trois ou quatre paramètres, on construirait environ $2n$ mesures. A deux dimensions, on a peut-être intérêt à en construire plus, mais il est certainement superflu de générer dix fois plus de mesures qu'il n'y a de données.

8 - CONSEQUENCES POUR LA MISE AU POINT D'UNE METHODE AUTOMATIQUE D'INFERENCE DE LA COVARIANCE GENERALISEE.

On se limite bien sûr à un modèle polynômial comprenant des termes en $\delta, -h, h^3$ (si $k \geq 1$), et éventuellement $-h^5$ (si $k \geq 2$).

Les combinaisons linéaires autorisées sont construites sur des points appartenant à des voisinages analogues aux voisinages de krigeage qui seront utilisés ultérieurement ; on obtient ainsi un ajustement local adapté au problème. Au niveau de l'application pratique, différentes méthodes ont été employées :

- dans KRIGEPACK, on construit les combinaisons linéaires autorisées sur des ensembles comprenant autant de points qu'il y a de conditions d'universalité plus un ; les λ^α sont alors définis à une constante multiplicative près, qui disparaît lorsqu'on norme par le pondérateur $\bar{\omega}$; pour obtenir des combinaisons linéaires de types différents, on choisit ces ensembles de points sur différentes auréoles.

- dans BLUEPACK, on utilise plus de points ; pour déterminer la mesure λ , on prend celle qui correspond à l'erreur d'estimation d'une donnée par d'autres points expérimentaux par moindres carrés ; concrètement, chaque voisinage est divisé en deux auréoles, et on construit l'erreur d'estimation de chaque point de l'auréole interne par l'auréole externe, et de chaque point de l'auréole externe par l'auréole interne.

- Notons qu'on pourrait construire des mesures de krigeage analogues à celles qui seront utilisées ultérieurement, mais l'expérience montre que les résultats sont plutôt moins bons; cela tient à ce que les diverses mesures obtenues sont alors trop semblables pour permettre une bonne identification de plusieurs paramètres.

Pour le reste, si l'on veut une méthode assez simple, il suffit d'appliquer celle qui a été décrite, en tenant compte des diverses conclusions qui ont été énoncées et qui doivent rester valables à deux dimensions même pour des données irrégulièrement implantées.

Bien sûr, si on stocke tous les résultats intermédiaires (c'est-à-dire les $V(\lambda_i)$ et les $K^P(\lambda_i)$), on peut faire un travail plus fin : utiliser des pondérateurs différents selon le modèle recherché, éliminer les mesures extrêmes (de variance nettement trop forte ou trop faible), faire une régression robuste plutôt qu'une régression par moindres carrés, etc...

Dans tous les cas, il est indispensable de contrôler la qualité des ajustements obtenus, notamment en étudiant les erreurs de krigeage de points expérimentaux à partir de données voisines. Ceci peut révéler des anisotropies zonales, ou faire ressortir des données suspectes, etc... C'est aussi un moyen de comparer plusieurs ajustements possibles et de déterminer quel est le meilleur.

Pour terminer, précisons qu'une analyse structurale automatique ne saurait dispenser de calculer le variogramme brut, dont on peut toujours tirer des enseignements intéressants.

ADDENDUM : INFERENCE D'UNE COVARIANCE ANISOTROPE

1 - INTRODUCTION

On utilise habituellement comme modèles de covariance généralisée les schémas polynômiaux isotropes du type :

$$K_o(h) = C \delta + \sum_{p=0}^k b_p |h|^{2p+1}$$

Dans la plupart des applications, un modèle isotrope suffit, on travaille, en effet, sur des combinaisons linéaires d'ordre k , et, s'il y a éventuellement une anisotropie, on peut espérer qu'elle est prise en compte par la dérive. Il n'en reste pas moins que dans certains cas, on peut être conduit à rechercher un ajustement à un modèle anisotrope. L'inférence statistique des covariances présentant des anisotropies classiques - comme l'anisotropie géométrique - pose de sérieux problèmes pratiques (on suppose bien sûr que les paramètres de l'anisotropie ne sont pas connus). Aussi a-t-on cherché un type d'anisotropie plus adapté.

2 - LE MODELE ANISOTROPE DES BANDES TOURNANTES

On peut imaginer un modèle de covariance du type :

$$(1) \quad K(h, \theta) = C \delta + \sum_{p=0}^k \int_0^\pi b_p(\phi) |h \cos(\theta - \phi)|^{2p+1} d\phi$$

On reconnaît dans le second terme la covariance d'une fonction aléatoire générée par la méthode des bandes tournantes : il suffit de partir de fonctions aléatoires sur les droites qui aient une covariance généralisée variable selon la direction ϕ de la droite :

$$K_\phi(h) = \sum_{p=0}^k b_p(\phi) |h|^{2p+1}$$

On en déduit une condition suffisante concernant les $b_p(\phi)$ pour que $K(h, \theta)$ soit une covariance généralisée dans \mathbb{R}^2 : il suffit en effet que, pour toute valeur de ϕ , la fonction $K_\phi(h)$ associée soit une covariance généralisée dans \mathbb{R}^1 .

En pratique l'inférence d'une telle covariance est encore

délicate. Il faut déterminer les différentes fonctions $b_p(\phi)$. Même si on discrétise l'intégrale en se limitant à 4 directions, il reste 13 paramètres à ajuster si on s'arrête au degré 5, ou 9 paramètres si on se limite à une covariance de degré 3.

Un modèle moins général, mais plus utilisable est du type :

$$(2) \quad K(h, \theta) = C\delta + \int_0^\pi a(\phi) \sum_{p=0}^k b_p |h \cos(\theta - \phi)|^{2p+1} d\phi$$

C'est un cas particulier du modèle (1) avec $b_p(\phi) = b_p a(\phi)$

Dans le cas isotrope où $a(\phi) \equiv 1$, on a :

$$K(h) = C\delta + b_p I_p |h|^{2p+1} \quad \text{avec} \quad I_p = \int_0^\pi |\cos \phi|^{2p+1} d\phi$$

Nous prendrons donc plutôt le modèle du type :

$$(3) \quad K(h, \theta) = C\delta + \int_0^\pi a(\phi) \sum_{p=0}^k \frac{b_p}{I_p} |h \cos(\theta - \phi)|^{2p+1} d\phi$$

$$\text{avec} \quad I_p = \int_0^\pi |\cos \phi|^{2p+1} d\phi$$

Le modèle isotrope associé à $a(\phi) \equiv 1$ est alors le polynôme de coefficients b_p .

Pour ce qui est de I_p on a :

$$I_p = 2 \int_0^{\frac{\pi}{2}} \cos \phi^{2p+1} d\phi = 2 \int_0^{\frac{\pi}{2}} (1 - \sin^2 \phi)^p d\sin \phi = 2 \int_0^1 (1 - x^2)^p dx$$

Une intégration par parties conduit à la formule de récurrence

$$I_p = 2p(I_{p-1} - I_p)$$

D'où :

$$I_0 = 2 \quad I_p = 2 \frac{2p(2p-2)(2p-4)\dots 2}{(2p+1)(2p-1)(2p-3)\dots 1}$$

3 - DISCRETISATION : LE MODELE ANISOTROPE OCTOGONAL

En pratique, on est amené à discrétiser le modèle précédent. Comme dans le cas de l'analyse structurale des FAI-O où on calcule le demi-variogramme dans quatre directions, nous avons adopté le modèle octogonal issu de la discrétisation du domaine de variation de ϕ en $0, \frac{\pi}{4}, \frac{\pi}{2}$ et $\frac{3\pi}{4}$. Notons x et y les composantes du vecteur h . Le modèle (3) ci-dessus s'écrit alors :

$$K(x,y) = C\delta + a_1 \sum_{p=0}^k \frac{b_p}{I_p} |x|^{2p+1} + a_2 \sum_{p=0}^k \frac{b_p}{I_p} \left| \frac{x+y}{\sqrt{2}} \right|^{2p+1} + a_3 \sum_{p=0}^k \frac{b_p}{I_p} |y|^{2p+1} + a_4 \sum_{p=0}^k \frac{b_p}{I_p} \left| \frac{x-y}{\sqrt{2}} \right|^{2p+1}$$

$$\text{ici } I_p = \sum_{j=0}^3 \left| \cos j \frac{\pi}{4} \right|^{2p+1} = 1 + \frac{2}{(\sqrt{2})^{2p+1}}$$

Pour évaluer le degré d'approximation que représente ce modèle par rapport au modèle continu, cherchons quelle est l'allure de l'extrémité des vecteurs $h = (x,y)$ tels que $K(x,y)$ soit constant, lorsque les coefficients a_i sont tous égaux. Nous nous limiterons aux monômes. On cherche donc l'ensemble des points (x,y) tels que :

$$\frac{1}{I_p} \left[|x|^{2p+1} + \left| \frac{x+y}{\sqrt{2}} \right|^{2p+1} + |y|^{2p+1} + \left| \frac{x-y}{\sqrt{2}} \right|^{2p+1} \right] = 1$$

Par raison de symétrie on se limite à $0 \leq x \leq y$.

- $p=0$: l'équation s'écrit: $\frac{x}{1+\sqrt{2}} + y = 1$

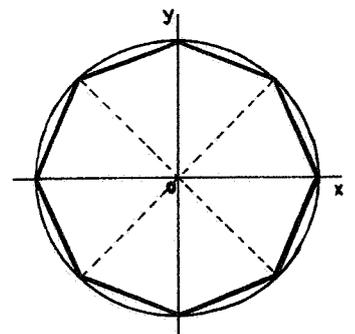
d'où $y = 1 - \frac{x}{1+\sqrt{2}}$

- $p=1$: l'équation se met sous la forme

$$y^3 + Py + Q = 0$$

avec $P = \frac{3x^2}{1+\sqrt{2}}$

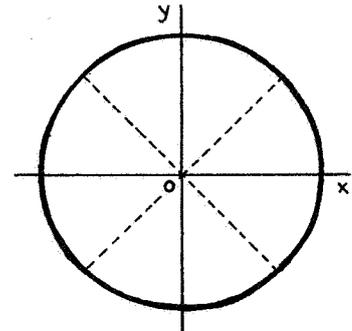
$$Q = \frac{\sqrt{2} x^3}{1+\sqrt{2}} - 1$$



On est dans le cas où $4P^3 + 27Q^2 > 0$ avec $P > 0$ et $Q < 0$. Il y a donc une seule racine réelle $y = \rho \operatorname{sh} \theta$ définie par :

$$\rho = \sqrt{\frac{4P}{3}} \quad \operatorname{sh} 3\theta = -\frac{3\sqrt{3} Q}{2 P \sqrt{P}}$$

D'où le graphe ci-contre qui est presque un cercle parfait (la différence sur le rayon vecteur est inférieure à 1 millième).



On peut donc approcher avec une bonne précision le cas isotrope. Il est à noter que l'anisotropie prise en compte par ce modèle n'est pas une anisotropie géométrique (sauf cas très particulier, $K(x,y)$ ne peut pas se mettre sous la forme $K_0 \left(\frac{h}{p(\theta)} \right)$ où (h,θ) désigne le point (x,y) en coordonnées polaires et où $p(\theta)$ désignerait le rapport d'anisotropie dans la direction θ ; elle se rapproche plutôt d'une anisotropie zonale.

4 - INFERENCE STATISTIQUE

On a vu que dans le cas isotrope l'inférence de la covariance n'est pas toujours évidente. Ici elle est encore plus délicate, et c'est pourquoi nous n'avons pratiquement jamais effectué d'inférence automatique de covariance anisotrope. Nous pouvons néanmoins donner quelques indications :

- Si les coefficients b_p sont connus (à un facteur multiplicatif près), il ne reste que 4 ou 5 paramètres à évaluer (les a_i et éventuellement l'effet de pépité C). On peut alors probablement effectuer l'inférence selon une méthode analogue à celle qui est exposée au chapitre II, mais on est certainement aux limites de l'application de cette méthode. Comme en pratique les b_p sont inconnus, cela signifie qu'on ne peut guère utiliser ce modèle que lorsqu'il y a un seul terme b_p , c'est-à-dire essentiellement pour un modèle comportant un effet de pépité et les termes linéaires x , $|x+y|$, $|y|$, $|x-y|$.

- Pour ajuster les paramètres d'un tel modèle, il faut construire des combinaisons linéaires $Z(\lambda)$ sur des configurations de points présentant un étirement net dans les différentes directions.

- Si la covariance "vraie" présente une anisotropie peu ou moyennement marquée (rapport d'anisotropie inférieur à 2 ou 3), il vaut mieux avoir un bon ajustement à un modèle isotrope plutôt qu'un mauvais ajustement à un schéma anisotrope. Si l'anisotropie est très marquée, alors généralement on s'en doute dès le départ, et le mieux est encore d'évaluer a priori les caractéristiques de cette anisotropie.

ANNEXE : APPROXIMATION D'UN MONÔME PAR UN POLYNÔME DE DEGRÉ INFÉRIEUR
DÉTERMINÉ PAR MOINDRES CARRÉS

A - Cas à 1 dimension

On cherche l'erreur quadratique moyenne que l'on commet en approximant sur $(-1, 1)$ le monôme x^{k_0} par un polynôme de degré $k < k_0$, déterminé par moindres carrés. On a :

$$I(a_0, \dots, a_k) = \frac{1}{2} \int_{-1}^1 (x^{k_0} - \sum_{l=0}^k a_l x^l)^2 dx$$

Les a_l sont solution de :

$$\frac{\delta I}{\delta a_l} = 0 \quad l = 0, 1, \dots, k.$$

On en déduit alors l'erreur quadratique moyenne cherchée.

Notons tout de suite que, après développement de l'intégrant, seuls les termes correspondant à une puissance paire de x apportent une contribution non nulle à l'intégrale. On omettra donc les puissances impaires de x dès le développement. De plus, par raison de symétrie, on se ramène en même temps à une intégrale sur $(0,1)$.

1) approximation de x par une constante a

$$I = \frac{1}{2} \int_{-1}^1 (x-a)^2 dx = \int_0^1 (x^2+a^2) dx = \frac{1}{3} + a^2$$

$$dI = 0 \Rightarrow a = 0$$

$$\text{alors } I = \frac{1}{3}$$

2) approximation de x^2 par $a + bx$

$$\begin{aligned} I &= \frac{1}{2} \int_{-1}^1 (x^2 - a - bx)^2 dx \\ &= \int_0^1 (x^4 + a^2 + b^2 x^2 - 2ax^2) dx \\ &= \frac{1}{5} + \frac{3a^2 + b^2 - 2a}{3} \end{aligned}$$

$$\frac{\delta I}{\delta a} = \frac{6a-2}{5}$$

$$\frac{\delta I}{\delta b} = \frac{2b}{3}$$

$$dI = 0 \Rightarrow a = \frac{1}{3} \quad b = 0$$

$$\text{alors } I = \frac{4}{45}$$

3) approximation de x^2 par une constante a

On a évidemment le même résultat qu'en 2.

4) approximation de x^3 par $a+bx+cx^2$

$$\begin{aligned} I &= \frac{1}{2} \int_{-1}^1 (x^3 - a - bx - cx^2)^2 dx \\ &= \int_0^1 (x^6 + a^2 + b^2x^2 + c^2x^4 - 2bx^4 + 2acx^2) dx \\ &= \frac{1}{7} + \frac{15a^2 + 5b^2 + 3c^2 - 6b + 10ac}{15} \end{aligned}$$

$$\frac{\delta I}{\delta a} = \frac{30a + 10c}{15}$$

$$\frac{\delta I}{\delta b} = \frac{10b - 6}{15}$$

$$\frac{\delta I}{\delta c} = \frac{6c + 10a}{15}$$

$$dI = 0 \Rightarrow a=0 \quad b = \frac{3}{5} \quad c=0$$

$$\text{alors } I = \frac{4}{175}$$

5) approximation de x^3 par $a+bx$

On a évidemment le même résultat qu'en 4.

6) approximation de x^3 par une constante a

$$I = \frac{1}{2} \int_{-1}^1 (x^3 - a)^2 dx = \int_0^1 (x^6 + a^2) dx = \frac{1}{7} + a^2$$

$$dI = 0 \Rightarrow a = 0$$

$$\text{alors } I = \frac{1}{7}$$

B - Cas à 2 dimensions

On pose le problème de la même manière, mais sur $(-1,1) \times (-1,1)$. On obtiendrait évidemment les mêmes résultats qu'en A pour x^k et des résultats analogues pour y^k . Il ne reste donc que le cas des monômes intermédiaires. Comme précédemment, on négligera dans le développement de l'intégrand tous les termes comportant une puissance impaire en x ou en y .

1) approximation de xy par $a+bx+cy$

$$\begin{aligned} I &= \frac{1}{4} \int_{-1}^1 \int_{-1}^1 (xy - a - bx - cy)^2 dx dy \\ &= \int_0^1 \int_0^1 (x^2 y^2 + a^2 + b^2 x^2 + c^2 y^2) dx dy \\ &= \frac{1}{9} + \frac{3a^2 + b^2 + c^2}{3} \end{aligned}$$

$$dI = 0 \Rightarrow a = 0 \quad b = 0 \quad c = 0$$

$$\text{alors } I = \frac{1}{9}$$

2) approximation de xy par une constante a

On a évidemment le même résultat qu'en 1.

3) approximation de $x^2 y$ par $a+bx+cy+dx^2+exy+fy^2$

$$I = \frac{1}{4} \int_{-1}^1 \int_{-1}^1 (x^2 y - a - bx - cy - dx^2 - exy - fy^2) dx dy$$