

FONTAINEBLEAU-CG

C-84

**Les Fonctions Aléatoires Intrinsèques**  
**d'ordre  $k$**

---

**G. MATHERON**  
**P. DELFINER**

**Septembre 1980**



# Les Fonctions Aléatoires Intrinsèques d'ordre $k$

par

P. Delfiner et G. Matheron

---

## TABLE

0 - FAI-0 et FAI- $k$ .....	1
1 - Les ambiguïtés de la notion de dérive.....	3
2 - Passage aux FAI- $k$ .....	6
3 - Représentations des FAI- $k$ .....	8
4 - Compléments sur les accroissements d'ordre $k$ .....	13
5 - Les covariances généralisées.....	14
6 - Compléments sur les CG.....	18
7 - Les covariances polynomiales.....	19
8 - Le krigeage en FAI- $k$ .....	23
9 - Le statut de la dérive en FAI- $k$ .....	24
10 - L'estimation des covariances généralisées.....	26
11 - Conclusions.....	32
Bibliographie.....	35



# Les Fonctions Aléatoires Intrinsèques d'ordre $k$

par

P. Delfiner et G. Matheron

---

## 0 - FAI-0 et FAI-k

La notion de fonction aléatoire intrinsèque d'ordre  $k$  (en abrégé dans ce qui suit : FAI- $k$ ) constitue une généralisation naturelle des fonctions aléatoires intrinsèques (c.a.d. à accroissements stationnaires) de la Géostatistique usuelle : ces dernières correspondent au cas particulier  $k = 0$  et seront désormais appelées FAI-0. Relativement aux fonctions aléatoires stationnaires, ou FAST, le passage aux FAI-0 est marqué par les circonstances suivantes :

~ L'outil de travail essentiel du cas stationnaire, la covariance  $C(h)$ , est remplacé par le variogramme  $\gamma(h)$ . On gagne ainsi en généralité, la classe des variogrammes étant d'extension plus vaste que celle des covariances ( $C(h)$  doit être de type positif; pour le variogramme on exige seulement que  $-\gamma$  soit de type positif conditionnel d'ordre 0). Contrairement à la covariance, le variogramme n'est pas nécessairement borné, et permet de décrire des phénomènes présentant une capacité de dispersion potentiellement illimitée (variance théoriquement infinie). Ainsi le mouvement brownien admet un variogramme linéaire, mais pas de covariance stationnaire.

~ Dans le cas stationnaire, il existe une espérance ou moyenne  $m$  autour de laquelle la FAST fluctue. Il s'agit en quelque sorte d'un phénomène autorégulé, qu'une force de rappel contraint à revenir au voisinage de cette valeur  $m$  dès qu'il s'en est quelque peu écarté. Dans le cas d'une FAI-0 à variogramme non borné, il n'**existe** aucune régulation de ce genre. La particule brownienne n'a aucun souvenir de son passé. A chaque instant, elle repart de sa position actuelle comme s'il s'agissait

d'une nouvelle origine, et ne manifeste aucune tendance préférentielle à revenir à son point de départ initial. Il n'existe pas d'espérance ou valeur moyenne  $m$  constante. Un processus de ce genre n'est, en réalité, défini qu'à une constante arbitraire près, et, aussi bien, on ne l'étudie que par l'intermédiaire de ses accroissements.

~ Dans le cas d'une FAST, toute combinaison linéaire

$$Z(\lambda) = \sum_i \lambda_i Z(x_i)$$

admet la variance

$$\text{Var } Z(\lambda) = \sum_{i,j} \lambda_i \lambda_j C(x_i - x_j)$$

où  $C(h)$  est la covariance centrée. Dans le cas d'une FAI-0, seules admettent une variance finie les combinaisons linéaires, dites autorisées ou admissibles, vérifiant la condition

$$(0-1) \quad \sum \lambda_i = 0$$

(poids total égal à 0). Cette variance se calcule alors comme s'il existait une covariance  $C = -\gamma$  :

$$(0-2) \quad \text{Var } Z(\lambda) = - \sum_{i,j} \lambda_i \lambda_j \gamma(x_i - x_j)$$

Il est bien entendu que cette formule n'est valable que pour les combinaisons linéaires autorisées, c.a.d. vérifiant la condition (0-1). Les autres combinaisons n'ont pas, en général, de variance finie. On voit ainsi qu'au prix d'une restriction opératoire relativement minime (s'imposer de ne manipuler que des combinaisons linéaires de poids nul) nous gagnons la possibilité de traiter une large classe de phénomènes qui, comme le mouvement brownien, échappaient aux modèles stationnaires.

Est-il possible d'aller plus loin dans le même sens, autrement dit, pouvons-nous accéder, au prix de restrictions plus sévères que les conditions (0-1) procéder à des classes plus larges encore de phénomènes non-stationnaires ? La théorie des FAI-k montre que cela est possible, mathématiquement parlant. D'autre

part, cette théorie - et c'est là sa justification du point de vue pratique - survient à point nommé pour lever certaines difficultés ou ambiguïtés que l'on rencontre dans la pratique du krigeage universel.

### 1 - Les ambiguïtés de la notion de dérive

En krigeage universel, le modèle de base est la dichotomie :

$$(1-0) \quad Z(x) = Y(x) + m(x)$$

Ici  $Z(x)$  représente la variable étudiée,  $m(x)$  est la dérive, et  $Y(x) = Z(x) - m(x)$  le "résidu" ou "fluctuation" autour de cette dérive. Mathématiquement parlant, on définit la dérive comme l'espérance  $m(x) = E Z(x)$ , mais du point de vue physique, du moins dans le cas d'un phénomène non-répétitif, cette définition risque de rester purement formelle : une espérance, en effet, est une moyenne prise sur un grand nombre de réalisations du même phénomène, alors qu'il n'en existe qu'une seule dans la nature.

Dans certains cas, cependant, la dichotomie (1-0) a un sens assez clair. Les géophysiciens, par exemple, séparent l'anomalie régionale (= dérive) et l'anomalie locale (= résidu), la première représentant, par exemple, l'influence du socle profond, et la seconde l'effet des masses superficielles. Le cas favorable est celui d'un phénomène "autorégulé", que des causes perturbatrices locales n'écartent jamais beaucoup ni très longtemps d'une "tendance" générale représentable par une fonction simple et régulière (Fig. 1). Dans le modèle, le résidu  $Y(x)$  sera dans ce cas

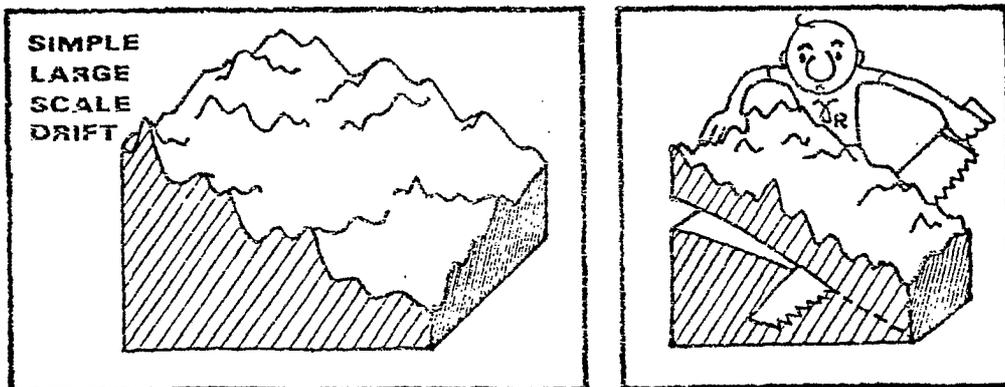


Figure 1 : Une situation favorable pour le Krigeage Universel.

une FAST dont la covariance aura une portée relativement petite (à l'échelle de travail adoptée).

Mais bien souvent la portée de la covariance des fluctuations n'est nullement petite relativement à l'échelle de travail. Et l'on a vu que dans ce cas le variogramme des résidus est entaché de biais inextricables. La dérive elle-même (à supposer qu'on puisse la définir) n'est pas forcément représentable par une expression analytique simple valable sur la totalité du domaine étudié (Fig. 2).

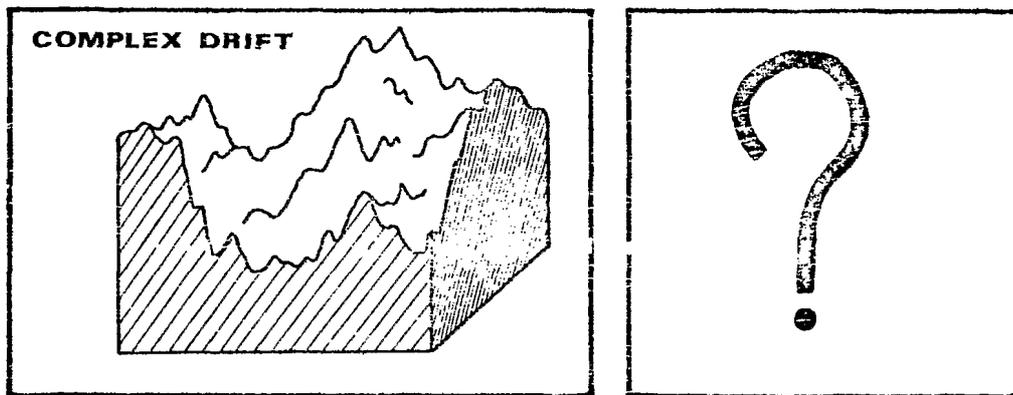


Figure 2 : Un cas embarrassant pour la modélisation de la dérive.

En pareil cas, on travaille le plus souvent par voisinages glissants avec un modèle local de la forme :

$$m(x) = \sum_l a_l(x_0) f_l^1(x)$$

valable seulement sur un voisinage  $V_{x_0}$  de chaque point  $x_0$  : les  $f_l^1$  sont des fonctions données (en général des monomes), mais les coefficients inconnus  $a_l$  varient d'un voisinage à l'autre. Leur estimation devient beaucoup plus difficile (puisque'il y a peu de données dans chaque voisinage glissant) et l'on voit aussi se poser de difficiles problèmes de raccordement. A dire vrai, on ne sait plus exactement ce que l'on est en train de faire sous le nom d' "estimation de la dérive". Il se pourrait bien que, dans un cas comme celui de la figure 2, la dichotomie en dérive et résidu n'ait plus grand sens.

Mais cette dichotomie douteuse (dérive régulière + fluctuation stationnaire) est-elle réellement utile ? En fait, il est possible de s'en affranchir. Considérons, en effet, le système d'équations du krigeage universel (relatif au point  $x_0$ ) :

$$(1-0) \quad \left\{ \begin{array}{l} \lambda^\alpha \sigma_{\alpha\beta} = \sigma_{\beta x} + \mu_1 f_\beta^1 \quad (\text{optimalité}) \\ \lambda^\alpha f_\alpha^1 = f_{x_0}^1 \quad (\text{universalité}) \end{array} \right.$$

Ce qui caractérise vraiment ce système, ce sont les deux circonstances suivantes :

~ on recherche l'estimateur optimal parmi la classe des combinaisons linéaires vérifiant la condition d'universalité

~ la variance d'estimation des estimateurs appartenant à cette classe s'exprime à l'aide d'une fonction  $\sigma(h)$  d'un seul argument vectoriel  $h$ . (En ce qui concerne les autres estimateurs linéaires, nous ne savons pas calculer leur variance d'estimation, à supposer qu'elle existe).

Ces deux circonstances évoquent le passage des FAST aux FAI-0. De fait, elles vont nous conduire aux FAI- $k$ . Changeons les notations pour mieux mettre en évidence l'analogie :

Aux points expérimentaux  $x_\alpha$  ( $\alpha = 1, 2, \dots, N$ ) adjoignons le point à estimer  $x_0$ , et donnons aux indices latins  $i, j, \dots$  les valeurs possibles  $0, 1, 2, \dots, N$ . De même, aux poids  $\lambda^\alpha$  affectés aux points  $x_\alpha$ , joignons le poids  $\lambda^0 = -1$  affecté à  $x_0$ . Alors la condition d'universalité s'écrit :

$$(1-1) \quad \sum_{i=0}^N \lambda^i f_i^1 = 0$$

Nous dirons qu'une combinaison linéaire qui vérifie cette condition est une combinaison autorisée (relativement aux fonctions  $f^1$ ). Comme l'erreur d'estimation s'écrit :

$$Z^*(x) - Z(x_0) = \lambda^i Z_i$$

nous voyons que la condition d'universalité exprime simplement que l'erreur d'estimation doit être une combinaison linéaire autorisée. Comme ces combinaisons autorisées annulent ("filtrent") les fonctions  $f^1$ , nous voyons en particulier que l'erreur serait nulle si  $Z(x)$  était elle-même une combinaison linéaire  $A_1 f^1(x)$  de ces fonctions. Plus généralement, le changement de  $Z(x)$  en  $Z(x) + A_1 f^1(x)$  laisserait cette erreur invariante.

Moyennant cette restriction opératoire (n'utiliser jamais que des combinaisons linéaires  $Z(\lambda) = \sum \lambda^i Z(x_i)$  autorisées), la seconde condition requise pour que le krigeage soit possible est la suivante : il existe une fonction  $K(h)$  (qui ne sera pas nécessairement une vraie covariance) telle que toute combinaison autorisée  $Z(\lambda)$  admette la variance

$$\text{Var } Z(\lambda) = \sum_{i,j} \lambda^i \lambda^j K(x_i - x_j)$$

## 2 - Passage aux FAI-k.

Encore faut-il, évidemment, que l'estimation de cette pseudo-covariance  $K(h)$  soit possible à partir des données. De même que la stationnarité des accroissements d'une FAI-0 permet l'estimation du variogramme, de même ici notre modèle doit comporter une clause de stationnarité portant sur les combinaisons autorisées (que l'on appelle souvent, par analogie, accroissements généralisés). La définition précise est la suivante :

Soit  $\lambda$  une combinaison linéaire attribuant le poids  $\lambda^i$  aux points  $x_i$ . Pour toute fonction  $\varphi$ , nous poserons

$$\varphi(\lambda) = \sum \lambda^i \varphi(x_i)$$

Nous appellerons translatée par un vecteur  $h$  de la combinaison linéaire  $\lambda$ , et nous noterons  $\tau_h \lambda$  la combinaison attribuant les mêmes poids  $\lambda^i$  aux points  $x_i + h$ . Pour toute fonction  $\varphi$ , on aura ainsi :

$$\varphi(\tau_h \lambda) = \sum \lambda^i \varphi(x_i + h)$$

Si  $Z(x)$  est une FA et  $\lambda$  une combinaison linéaire, nous obtenons une nouvelle FA (régularisée de  $Z(x)$  par  $\lambda$ ) en posant :

$$Z_\lambda(x) = Z(\tau_x \lambda) = \sum \lambda^i Z(x + x_i)$$

La clause de stationnarité dont nous avons besoin peut alors s'énoncer ainsi : pour toute combinaison linéaire autorisée  $\lambda$ , la FA  $Z_\lambda(x)$  est stationnaire. On dit souvent, pour abrégé : les accroissements (généralisés) sont stationnaires.

Naturellement, tout ceci n'a de sens que si les translatées des combinaisons autorisées sont encore elles-mêmes des combinaisons autorisées. Autrement dit : l'ensemble des combinaisons autorisées doit être invariant par translation. Si nous nous reportons à la condition (1-1), nous voyons que les fonctions  $f^1$  doivent posséder la propriété suivante :

$$\sum_i \lambda^i f^1(x_i) = 0 \quad \text{pour tout } l \text{ entraîne } \sum_i \lambda^i f^1(x_i + h) = 0$$

pour tout  $h$  (et tout  $l$ ).

Autrement dit, nos combinaisons doivent annuler, outre les fonctions  $f^1(x)$ , toutes leurs translatées  $f^1(x + h)$ . Ces translatées doivent donc être elles-mêmes des combinaisons linéaires des fonctions  $f^1$ , soit

$$f^1(x + h) = B_s^1(h) f^s(x)$$

Si l'on désigne par  $\mathcal{F}$  l'espace vectoriel engendré par les fonctions  $f^1$  (c.a.d.  $f \in \mathcal{F}$  si  $f = \sum C_i f^i$ ), on voit que  $\mathcal{F}$  doit être un espace invariant par translation.

Il en résulte que les fonctions  $f^1$  ne peuvent pas être quelconques. (Nous laissons de côté, évidemment, le cas où les  $f^1$  seraient en nombre infini). Un théorème général affirme en effet que les seuls espaces  $\mathcal{F}$  de dimension finie constitués de fonctions continues (ou même, plus généralement, de fonctions mesurables) qui soient invariants par translation sont constitués

d'exponentielles-polynomes, c.a.d. de fonctions de la forme

$$f(x) = P(x) \exp (\langle a x \rangle)$$

où  $P(x)$  est un polynome, et  $\langle a x \rangle$  est une forme linéaire en  $x$ , à coefficients éventuellement complexes.

On voit que ce n'est pas uniquement pour des raisons de commodité que l'on prend, le plus souvent, pour  $\mathcal{F}$  l'espace des polynomes de degré inférieur ou égal à un nombre  $k$  donné : c'est avant tout pour satisfaire à cette exigence d'invariance par translation.

Nous sommes maintenant à même de donner une première définition (mathématique) des FAI- $k$ , qui généralisera celle des FAI-0.

Nous désignerons par  $\Lambda_k$  l'espace des combinaisons linéaires annulant les polynomes de degré inférieur ou égal à  $k$ , et nous dirons que tout  $\lambda \in \Lambda_k$  est une combinaison linéaire autorisée à l'ordre  $k$  (ou un accroissement généralisé d'ordre  $k$ ). Nous dirons alors qu'une FA  $Z(x)$  est une FAI- $k$  si, pour tout  $\lambda \in \Lambda_k$ , la FA

$$Z_\lambda(x) = Z(\tau_x \lambda) = \sum \lambda^i Z(x_i + x)$$

est une FA stationnaire en  $x \in \mathbb{R}^n$  (autrement dit : si les accroissements généralisés d'ordre  $k$  sont stationnaires) et d'espérance nulle.

(On ajoute la condition d'espérance nulle pour des raisons de commodité, sans que cela nuise à la généralité : si les accroissements d'ordre  $k$  sont stationnaires mais d'espérance non-nulle, les accroissements d'ordre  $k + 1$ , évidemment stationnaires, ont nécessairement une espérance égale à 0 : on a donc affaire à une FAI- $(k+1)$  selon la définition précédente.)

### 3 - Représentation des FAI- $k$ .

Si  $Z(x)$  est une FAI- $k$ , et si les  $A_1$  sont des variables aléatoires (non nécessairement indépendantes de la FA  $Z(x_1)$ ), la FA

$Z_1$  définie en posant

$$Z_1(x) = Z(x) + A_1 f^1(x)$$

est une FAI- $k$  : comme tout  $\lambda \in \Lambda_k$  filtre, par définition, les fonctions  $f^1$ , on a, en effet,  $Z_1(\lambda) = Z(\lambda)$  pour toute combinaison autorisée. Ayant mêmes accroissements d'ordre  $k$  que  $Z$ , la FA  $Z_1$  est donc elle-même intrinsèque à l'ordre  $k$ . Compte tenu de la règle opératoire qui nous impose de ne manipuler que des combinaisons  $\lambda \in \Lambda_k$ , ces deux fonctions aléatoires  $Z$  et  $Z_1$ , en tant que FAI- $k$ , ne sont pas réellement distinctes. Plutôt qu'à une FA particulière  $Z$  ou  $Z_1$ , la notion de FAI- $k$  s'applique à une classe d'équivalence de FA, à savoir la classe de toutes les FA admettant les mêmes accroissements (stationnaires) d'ordre  $k$  que la FA  $Z(x)$ . D'où la seconde (et définitive) définition de la notion de FAI- $k$  :

Une FAI- $k$   $\tilde{Z}$  est une application linéaire de l'espace  $\Lambda_k$  dans un espace de variables aléatoires d'espérance nulle admettant des moments d'ordre 2 finis, telle que, pour tout  $\lambda \in \Lambda_k$ , la FA  $x \rightarrow \tilde{Z}(\tau_x \lambda)$  soit stationnaire en  $x$ .

Cette seconde définition n'est plus générale qu'en apparence. De fait, si une FA (ordinaire)  $Z(x)$  coïncide avec  $\tilde{Z}$  sur  $\Lambda_k$  (c'est-à-dire donne la même valeur à toutes les combinaisons linéaires autorisées), soit

$$Z(\lambda) \equiv \sum \lambda^i Z(x_i) = \tilde{Z}(\lambda) \quad (\lambda \in \Lambda_k)$$

nous dirons que  $Z(x)$  est une représentation de la FAI- $k$   $\tilde{Z}$  (définie au sens de la seconde définition). Il est clair que cette représentation  $Z(x)$ , ayant elle-même ses accroissements stationnaires, vérifie les conditions de la première définition.

Les deux propriétés suivantes montrent qu'une FAI- $k$   $\tilde{Z}$  s'identifie à la classe de ses représentations (de sorte que les deux définitions apparaissent comme équivalentes) :

- a) Toute FAI- $k$   $\tilde{Z}$  admet des représentations

b) Si  $Z(x)$  est une représentation de  $\tilde{Z}$ , toutes les autres représentations sont de la forme  $Z(x) + A_1 f^1(x)$ , où les  $A_1$  sont des variables aléatoires quelconques.

Démonstration : Choisissons (ce qui est toujours possible) des combinaisons linéaires  $\lambda_1$  (c.a.d. des points  $x_\alpha$  et, pour chaque indice 1, des poids  $\lambda_1^\alpha$ ) vérifiant les conditions :

$$(3-1) \quad f^S(\lambda_1) \equiv \sum_{\alpha} \lambda_1^\alpha f_\alpha^S = \delta_1^S$$

(incidemment, on reconnaît ici les conditions d'universalité qui apparaissent lorsqu'on estime les coefficients de la dérive dans le modèle du krigeage universel). Ces  $\lambda_1$  ne sont évidemment pas dans  $\Lambda_k$ . Mais, pour tout point  $x$ , désignons par  $\varepsilon_x$  la combinaison affectant :

~ le poids 1 au point  $x$

~ le poids  $-f^1(x) \lambda_1^\alpha$  à chacun des points  $x_\alpha$

Autrement dit, pour toute fonction  $\varphi$ , on a par définition :

$$(3-2) \quad \varphi(\varepsilon_x) = \varphi(x) - \varphi_\alpha \lambda_1^\alpha f^1(x)$$

Cette combinaison  $\varepsilon_x$  est autorisée à l'ordre  $k$  : de fait, compte tenu de (3-2), on trouve :

$$f^S(\varepsilon_x) = f^S(x) - f_\alpha^S \lambda_1^\alpha f^1(x)$$

Mais, d'après (3-1),  $f_\alpha^S \lambda_1^\alpha = \delta_1^S$  et donc  $f^S(\varepsilon_x) = 0$

Nous pouvons donc définir une FA  $Z(x)$  en posant pour tout  $x$

$$Z(x) = \tilde{Z}(\varepsilon_x)$$

Montrons que  $Z(x)$  est une représentation de  $\tilde{Z}$  : pour tout  $\lambda \in \Lambda_k$  attribuant les poids  $\lambda^i$  aux points  $x_i$  (avec  $\lambda^i f_i^1 = 0$ ), on a

$$Z(\lambda) = \sum_i \lambda^i \tilde{Z}(\varepsilon_{x_i}) = \tilde{Z} \left( \sum_i \lambda^i \varepsilon_{x_i} \right)$$

(puisque  $\tilde{Z}$  est linéaire)

Mais la combinaison  $\sum_i \lambda_i^1 \varepsilon_{x_i}$  est la somme de deux combinaisons linéaires :

- ~ l'une est  $\lambda$  elle-même (poids  $\lambda_i^1$  affectés aux points  $x_i$ )
- ~ l'autre attribue le poids  $-\lambda_i^1 f_i^1 \lambda_1^\alpha$  à chacun des points  $x_\alpha$  : mais, par hypothèse,  $\lambda_i^1 f_i^1 = 0$  puisque  $\lambda \in \Lambda_k$ . On a donc

$$\sum_i \lambda_i^1 \varepsilon_{x_i} = \lambda$$

et par suite  $Z(\lambda) = \tilde{Z}(\lambda)$  :  $Z(x)$  est bien une représentation, ce qui établit le point a).

Passons maintenant au point b). Il est clair que  $Z(x) + A_1 f^1(x)$ , où les  $A_1$  sont des variables aléatoires quelconques, est encore une représentation de  $\tilde{Z}$ . Inversement, soit  $Z_1$  une autre représentation de  $\tilde{Z}$ . On a donc  $Z(\lambda) - Z_1(\lambda) = 0$  pour tout  $\lambda \in \Lambda_k$ . Si nous posons  $Z_0(x) = Z(x) - Z_1(x)$ ,  $Z_0(x)$  est donc une représentation de la FAI- $k$  identiquement nulle ( $Z_0(\lambda) = 0$  pour tout  $\lambda \in \Lambda_k$ ). Considérons la combinaison  $\varepsilon_x \in \Lambda_k$  déjà utilisée ci-dessus. On trouve ainsi :

$$0 = Z_0(\varepsilon_x) = Z_0(x) - f^1(x) \sum_\alpha \lambda_1^\alpha Z_0(x_\alpha)$$

Désignons par  $A_1$  la variable  $\lambda_1^\alpha Z_0(x_\alpha)$  qui peut être absolument quelconque). La relation ci-dessus s'écrit ainsi :

$$Z_0(x) = A_1 f^1(x)$$

ce qui achève la démonstration.

Exemples - Le mouvement brownien  $W(t)$  est une FAI-0 (sur la droite), caractérisée par un variogramme linéaire. Son intégrale

$$W_1(t) = \int_0^t W(\tau) d\tau$$

est une FAI-1. De manière générale, la primitive d'ordre  $k$  défini (à un polynome près de degré  $k$ ) par

$$W_k(t) = \int_0^t \frac{(t - \tau)^{k-1}}{(k-1)!} W(\tau) d\tau$$

est une FAI-k, caractérisée par une covariance généralisée (voir ci-dessous) en  $|h|^{2k+1}$ .

Ces processus existent également dans les espaces à plusieurs dimensions. On peut voir sur la figure 3 des exemples de réalisations de ces FAI-k dans l'espace à 2 dimensions, pour  $k = 0, 1$  et  $2$ . A la structure cellulaire, ou mosaïque de lignes de niveaux de la FAI-0, (variogramme linéaire), se substituent, lorsque  $k$  augmente, des structures d'apparence de plus en plus régulière, où plus rien, semble-t-il, n'évoque la stationnarité classique : pourtant, les accroissements d'ordre  $k$  sont réellement stationnaires, et ces structures relèvent bien du formalisme des FAI-k. On mesure ainsi le degré de généralité du modèle FAI-k, et son aptitude à représenter adéquatement à peu près n'importe quel phénomène naturel.

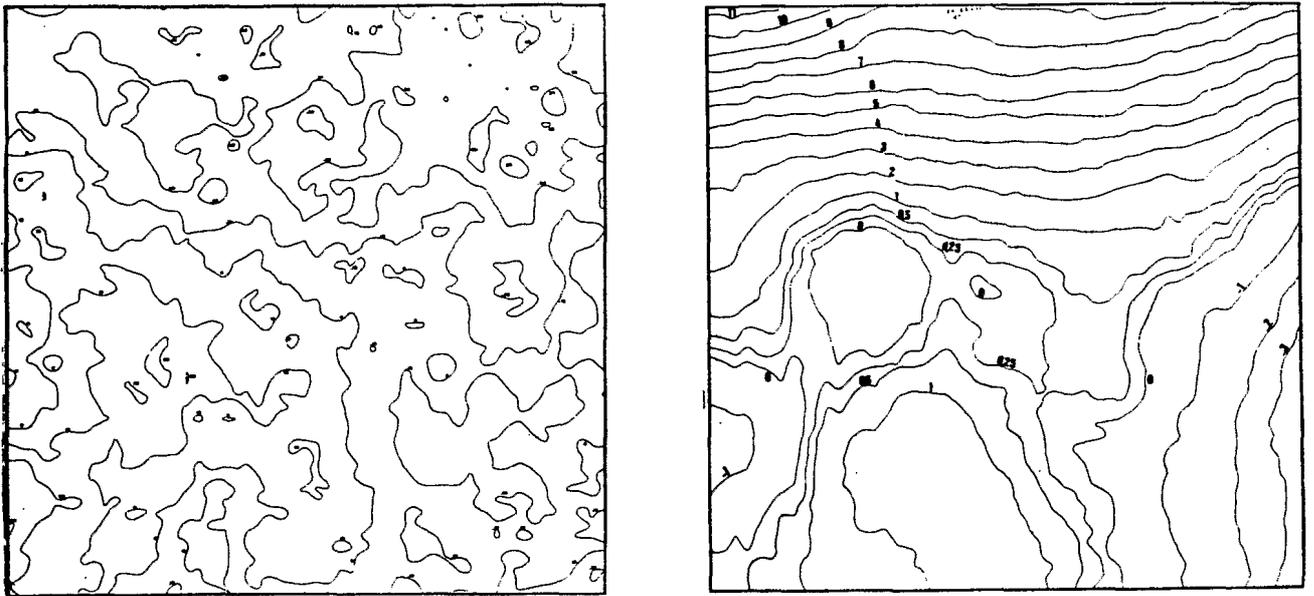


Figure 3 : Réalisations de FAI-k de covariance généralisée  $K(h) = (-1)^{k+1} |h|^{2k+1}$  avec  $k = 0$  (en haut à gauche),  $k = 1$  (ci-dessus),  $k = 2$  (ci-contre) - d'après J.P. Orfeuil.

De même, la primitive d'ordre k+1 d'une FAST quelconque est toujours une FAI-k. Inversement (dans un espace à une ou plusieurs dimensions) si une FA Z(x) est différentiable, et si toutes ses dérivées partielles d'ordre k+1 sont stationnaires, Z(x) est une FAI-k. Cette propriété suffit d'ailleurs à caractériser les FAI-k différentiables. (Il existe évidemment des FAI-k non différentiables. Mais on peut montrer que toute FAI-k continue est somme d'une FAST et d'une FAI-k différentiable).

4 - Compléments sur les accroissements d'ordre k.

Dans l'espace à n dimensions, le nombre des monomes de degré ≤ k est  $C_{n+k}^k$ . Pour former un accroissement autorisé à l'ordre k à partir de N points  $x_i$ , on doit trouver une solution λ non triviale à l'équation

$$\lambda^i f_i^1 = 0$$

Il faut donc que le rang de la matrice des  $f_i^1$  soit plus petit que N. Sauf dans des cas très spéciaux (points alignés par exemple) ce rang est en général  $C_{n+k}^k$ , de sorte qu'il faut prendre  $N \geq C_{n+k}^k + 1$ , c'est à dire au moins

2 points	si k = 0	
n+2 points	si k = 1	
1 + (n+1)(n+2)/2 points	si k = 2	etc...

Exemples :

1) A une dimension, les accroissements successifs d'une fonction f (modulo une constante a donnée) sont définis par :

$$\Delta_1 f(x) = f(x + a) - f(x)$$

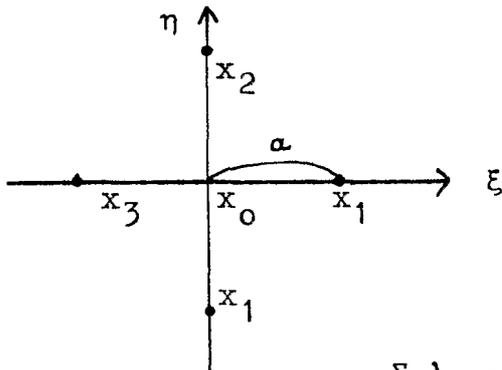
$$\Delta_2 f(x) = \Delta_1 \Delta_1 f(x) = f(x+2a) - 2f(x+a) + f(x)$$

.....

$$\Delta_{k+1} f(x) = \Delta_1 \Delta_k f(x) = \sum_{p=0}^{k+1} (-1)^{k+p} C_{k+1}^p f(x + pa)$$

Comme  $\Delta_1$  diminue d'une unité le degré de tout polynome, les accroissements d'ordre 1 sont autorisés à l'ordre 0 (ils filtrent les constantes) et, de même, les accroissements d'ordre  $k+1$  sont autorisés à l'ordre  $k$  (ils filtrent les polynomes d'ordre  $k$ ).

2) Considérons 5 points disposés en quinconce dans l'espace à 2 dimensions. Prenons  $\lambda_0 = 1$ ,  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$



$= -1/4$ , et désignons par  $\xi_i$  et  $\eta_i$  les deux coordonnées du point  $x_i$ . Cette combinaison linéaire est autorisée à l'ordre 1 : on vérifie sans peine, en effet, les conditions

$$\sum \lambda_i = \sum \lambda_i \xi_i = \sum \lambda_i \eta_i = 0$$

Comme  $\xi_i \eta_i = 0$  pour chaque  $i$ , cette combinaison annule aussi le monôme  $\xi\eta$ . Mais elle n'annule ni  $\xi^2$  ni  $\eta^2$  (de fait, on trouve par exemple  $\sum \lambda_i \xi_i^2 = -a^2/4 \neq 0$ ). Donc, cette combinaison n'est pas un accroissement d'ordre 2.

### 5 - Les covariances généralisées.

De même que toute FAI-0 admet un variogramme  $\gamma(h)$  dont la connaissance permet le calcul de la variance de toute combinaison linéaire autorisée à l'ordre 0, de même, à toute FAI- $k$   $Z(x)$  (continue en moyenne quadratique), on peut associer une fonction  $K(h)$  que nous appellerons covariance généralisée, qui permet le calcul des variances de tous les accroissements d'ordre  $k$ . La définition est la suivante :

Si  $Z$  est une FAI- $k$ , on dit qu'une fonction symétrique  $K(h)$  définie sur  $\mathbb{R}^n$  est une covariance généralisée (CG) de  $Z$  si l'on a

$$E [Z(\lambda) Z(\mu)] = \sum_{i,j} \lambda^i \mu^j K(x_i - y_j)$$

pour tout couple  $\lambda, \mu$  d'accroissements d'ordre  $k$  (attribuant respectivement les poids  $\lambda^i$  aux points  $x_i$  et les poids  $\mu^j$  aux points  $y_j$ ).

En fait, il suffit que la condition ci-dessus soit vérifiée dans le cas  $\lambda = \mu$  (on le voit sans peine en développant  $(Z(\lambda+\mu))^2$ ). Autrement dit, il suffit de vérifier que l'on a

$$(5-1) \quad E [Z(\lambda)^2] = \sum_{i,j} \lambda^i \lambda^j K(x_i - x_j) \quad (\lambda \in \Lambda_k)$$

Ainsi, la donnée d'une CG  $K(h)$ , permettant le calcul des variances de toutes les combinaisons linéaires autorisées, suffira pour mener à bien le programme que nous nous sommes fixés dans le paragraphe 1.

A ce sujet, il existe un théorème d'existence et d'unicité valable pour toute FAI- $k$  continue (c.a.d. admettant au moins une représentation continue en moyenne quadratique).

Théorème d'existence : Toute FAI- $k$  continue admet une CG.

Ce théorème est vrai dans  $\mathbb{R}^n$ , quelle que soit la dimension  $n$ , et pour toutes les FAI- $k$ . Mais la démonstration, dans le cas général, rencontre certaines difficultés techniques, qui obscurcissent l'aspect presque intuitif du résultat. C'est pourquoi nous nous limiterons ici au cas très simple  $n = 1$  et  $Z$  différentiable.

Soit donc, sur la droite réelle, une FAI- $k$  admettant une représentation  $Z(x)$   $k+1$  fois différentiable (au sens m. q.). Comme on l'a vu, la dérivée  $Z^{(k+1)}(x)$  est une FAST, et admet donc donc une covariance  $C(h)$ . Désignons par  $\sigma(x,y)$  la covariance (non stationnaire) de  $Z(x)$ , de sorte que l'on a :

$$(5-2) \quad \frac{\partial^{2(k+1)}}{\partial x^{k+1} \partial y^{k+1}} \sigma(x, y) = C(x - y)$$

Désignons par  $F(h)$  une primitive d'ordre  $k+1$  de la fonction  $C(h)$  et (au signe près) par  $K(h)$  une primitive d'ordre  $2(k+1)$ , par exemple :

$$F(h) = \int_0^h \frac{(h-u)^k}{k!} C(u) du$$

$$K(h) = (-1)^{k+1} \int_0^h \frac{(h-u)^{2k+1}}{(2k+1)!} C(u) du$$

(noter que  $K(h) = K(-h)$ , de sorte que  $K$  est une fonction symétrique comme la covariance  $C(h)$ ).

Intégrons (5-2)  $(k+1)$  fois en  $x$  : on obtient  $F(x-y)$  à un polynôme près de degré  $k$  en  $x$ , dont les coefficients  $b_1$  seront en général des fonctions de  $y$ , soit :

$$\frac{\partial^{(k+1)}}{\partial y^{k+1}} \sigma(x,y) = F(x-y) + b_1(y) f^1(x)$$

Intégrons maintenant en  $y$ , à  $x$  fixé cette fois. En désignant par  $a_1(y)$  des primitives d'ordre  $k+1$  des  $b_1(y)$ , il vient cette fois

$$\sigma(x,y) = K(x-y) + a_1(y) f^1(x) + a_1'(x) f^1(y)$$

Comme  $K(x-y)$  et  $\sigma(x,y)$  sont des fonctions symétriques en  $x, y$ , on a d'ailleurs nécessairement  $a_1(y) = a_1'(y)$ , d'où la forme générale de notre covariance :

$$(5-3) \quad \sigma(x,y) = K(x-y) + a_1(y) f^1(x) + a_1(x) f^1(y)$$

Si maintenant  $\lambda$  et  $\mu$  sont deux combinaisons linéaires autorisées, on a

$$\sum_i \lambda^i f^1(x_i) = \sum_j \mu^j f^1(y_j) = 0$$

et donc, d'après (5-3); il vient

$$\sum_{i,j} \lambda^i \mu^j \sigma(x_i, y_j) = \sum_{i,j} \lambda^i \mu^j K(x_i - y_j)$$

Ainsi,  $K(h)$  est bien une CG.

Théorème d'unicité (à une équivalence près) : Cette covariance  $K(h)$  est unique, à un polynôme pair près de degré au plus égal à  $2k$ .

En effet, supposons que  $Z(x)$  admette deux CG  $K_1(h)$  et  $K_2(h)$  distinctes. Comme la covariance  $\sigma(x,y)$  de  $Z(x)$  admet une représentation de type (5-3) en fonction de chacune des deux CG  $K_1$  et  $K_2$ , on voit que la différence  $K_0 = K_1 - K_2$  est de la forme

$$(5-4) \quad K_0(x - y) = c_1(y) f^1(x) + c_1(x) f^1(y)$$

Comme nous avons supposé  $Z(x)$  différentiable,  $K_0(h)$  admet des dérivées jusqu'à l'ordre  $2k+2$ . Mais on a

$$\frac{\partial^{k+1}}{\partial x^{k+1}} f^1(x) = \frac{\partial^{k+1}}{\partial y^{k+1}} f^1(y) = 0$$

puisque ces fonctions sont des monômes de degré  $\leq k$ . Donc, en dérivant (5-4)  $k+1$  fois en  $x$  et  $k+1$  fois en  $y$  nous trouvons

$$\frac{d^{2(k+1)}}{dh^{2k+2}} K_0(h) = 0$$

$K_0(h)$  est donc un polynôme de degré au plus égal à  $2k+1$ . Comme par ailleurs  $K_0(h)$  est une fonction symétrique, c'est donc un polynôme pair de degré au plus égal à  $2k$ .

Il est facile de voir que pour tout polynôme pair  $P(h)$  de degré  $\leq 2k$ ,  $K(h) + P(h)$  est effectivement encore une CG (pour le voir, il suffit de remarquer que  $P(x - y)$  admet une expression de la forme (5-4) ci-dessus), de sorte que la CG est définie seulement à un polynôme pair près. On devrait donc parler plutôt de la classe d'équivalence des CG de  $Z$ .

Exemple : dans le cas d'une FAI-0, on a

$$K(h) = -\gamma(h) + c \frac{ste}{h}$$

La CG est égale à  $-\gamma$  à une constante arbitraire près.

### 6 - Compléments sur les CG.

Un CG  $K(h)$  ne peut pas être une fonction absolument quelconque. De même qu'une covariance ordinaire doit être de type positif, de même une CG doit être de type positif conditionnel d'ordre  $k$ , c'est à dire vérifier :

$$\sum_{i,j} \lambda^i \lambda^j K(x_i - x_j) \geq 0 \quad (\lambda \in \Lambda_k)$$

Cette condition nécessaire (qui exprime simplement que les variances doivent être positives) est également suffisante, et caractérise ainsi la classe des CG d'ordre  $k$ .

On peut encore caractériser les CG continues par leur représentation spectrale, analogue (quoique plus complexe) à la représentation spectrale des covariances ordinaires (Théorème classique de Bochner). Nous nous limitons ici à donner le résultat relatif au cas  $k = 0$  :

Une fonction  $\gamma(h)$  sur  $\mathbb{R}^n$  est le variogramme d'une FAI-0 continue si et seulement si elle est de la forme :

$$(6-1) \quad \gamma(h) = \int \frac{1 - \cos 2\pi \langle uh \rangle}{4 \pi^2 |u|^2} \chi(du)$$

où  $\chi$  est une mesure positive sur  $\mathbb{R}^n$ , sans atome à l'origine, et vérifiant la condition

$$\int \frac{1}{1 + 4\pi^2 |u|^2} \chi(du) < \infty$$

(NB : avec la définition adoptée pour les FAI-0, les accroissements ont une espérance nulle, et il s'agit ici d'un variogramme centré).

De ce théorème, on peut déduire que  $\gamma(h)$  croît moins vite, à l'infini, que  $|h|^2$ . Plus précisément, on démontre que :

$$\lim_{|h| \rightarrow \infty} \frac{\gamma(h)}{|h|^2} = 0$$

pour tout variogramme  $\gamma$ .

Parmi les variogrammes  $\gamma$ , certains sont en fait de la forme  $\gamma(h) = C(0) - C(h)$  pour une covariance stationnaire  $C(h)$ . La condition nécessaire et suffisante pour qu'il en soit ainsi est que le variogramme soit borné

$$\sup_{h \in \mathbb{R}^n} \gamma(h) < \infty$$

Lorsque cette condition est remplie, on montre que la FAI-0 ne diffère d'une FAST que par une constante (ou, ce qui revient au même, qu'elle admet une représentation stationnaire)

On a des résultats analogues dans le cas général :

Si  $K(h)$  est une CG d'ordre  $k$ , on a

$$\lim_{|h| \rightarrow \infty} \frac{K(h)}{|h|^{2k+2}} = 0$$

Toute FAI-( $k-1$ ) est évidemment une FAI- $k$  (de même que toute FAST est une FAI-0). En sens inverse on obtient le théorème suivant (qui généralise l'énoncé relatif au variogramme borné) :

Pour qu'une FAI- $k$  ( $k > 1$ ) admette une représentation qui soit elle-même une FAI-( $k-1$ ), il faut et il suffit que  $K(h)/|h|^{2k}$  soit borné.

(Noter que  $K(h)$ , définie à un polynôme près de degré  $2k$ , n'est pas nécessairement une CG d'ordre ( $k-1$ ) : car elle peut différer de la CG d'ordre  $k-1$  par des termes de degré  $2k$ . C'est pourquoi  $K(h)/|h|^{2k}$  reste borné sans nécessairement tendre vers 0 à l'infini).

## 7 - Les covariances polynomiales.

Comme,  $k$  augmentant, on impose la clause de stationnarité à des classes de plus en plus restreintes de combinaisons linéaires, on s'attend, corrélativement, à obtenir des classes de plus en plus vastes de FA et de CG (c'était, d'ailleurs, exactement notre objectif initial).

Par exemple, pour  $\alpha > 0$ ,  $|h|^\alpha$  n'est jamais une covariance stationnaire, et n'est un variogramme que si  $\alpha < 2$ . Mais pour

$$0 < \alpha < 2k+2$$

la fonction

$$K(h) = \Gamma\left(-\frac{\alpha}{2}\right) |h|^\alpha$$

est une CG d'ordre  $k$  (on notera que le signe de  $\Gamma(-\frac{\alpha}{2})$  alterne selon la parité de la partie entière de  $\alpha/2$ ). Dans le cas d'un exposant entier impair  $2p+1$ , on trouve que  $(-1)^{p+1} |h|^{2p+1}$  est une CG d'ordre  $k$  dès que  $k$  est supérieur ou égal à  $p$ . Plus généralement, la fonction

$$(7-1) \quad K(h) = \sum_{p=0}^k (-1)^{p+1} b_p |h|^{2p+1}$$

est une CG d'ordre  $k$ , moyennant certaines conditions sur les coefficients  $b_p$  (conditions automatiquement vérifiées dès que les  $b_p$  sont tous  $\geq 0$ ).

Les covariances de cette forme sont appelées (par abus de langage d'ailleurs) CG polynomiales. Ce sont des polynômes relativement au module  $|h|$  du vecteur  $h$ , et non relativement aux coordonnées de ce vecteur. Il n'y a pas de termes pairs dans l'écriture (7-1), mais, comme nous l'avons vu, il est toujours loisible de rajouter des termes en  $|h|^{2p}$  ( $p \leq k$ ), puisque  $K(h)$  n'est défini qu'à un polynôme pair près de degré  $2k$ .

Ce modèle (7-1) est très utile dans les applications. Les différents termes qu'il comporte rendent bien compte des différents aspects de la structure spatiale d'un phénomène. On peut voir sur la figure 3 l'allure des phénomènes associés aux termes purs  $-|h|$ ,  $|h|^3$  et  $-|h|^5$ . La dernière, en particulier, correspond déjà à ce que, dans l'ancienne terminologie, on appellerait sans doute une dérive (composante à variation lente et

suffisamment régulière pour être deux fois différentiable). Il est remarquable que ces différents aspects, difficilement séparables, puissent être pris en charge simultanément par un modèle aussi simple, sans qu'il soit nullement nécessaire de recourir à la douteuse dichotomie dérive + résidu.

En pratique, on adjoint le plus souvent au modèle (7-1) un terme  $C_0 \delta$  représentant un effet de pépité. Le tableau suivant résume ces modèles pour  $k = 0, 1$  et  $2$ .

POLYNOME FILTRÉ (Dérive)	$k$	MODELE DE COVARIANCE GENERALISEE POLYNOMIALE
Constant	0	$K(h) = C_0 \delta - b_0  h $
Linéaire	1	$K(h) = C_0 \delta - b_0  h  + b_1  h ^3$
Quadratique	2	$K(h) = C_0 \delta - b_0  h  + b_1  h ^3 - b_2  h ^5$
Contraintes dans $\mathbb{R}^2$ $C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -(10/3) \sqrt{b_0 b_2}$ dans $\mathbb{R}^3$ $C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -\sqrt{10} \sqrt{b_0 b_2}$		

TABIEAU I : Modèles de CG polynomiale pour  $k \leq 2$

Ces modèles présentent, dans les applications, de très grands avantages, du fait que leurs coefficients interviennent de manière linéaire, ce qui facilite considérablement leur identification.

Du point de vue des calculs numériques, il peut y avoir avantage à remplacer la matrice  $K_{\alpha\beta} = K(x_\alpha - x_\beta)$ , qui est seulement conditionnellement définie positive, par une matrice  $C_{\alpha\beta}$  strictement définie positive. Cela serait possible si la FAI- $k$  admettait une représentation stationnaire : il suffirait alors, en effet, de choisir comme (version de la) CG  $K(h)$  la covariance (stationnaire) de cette représentation. Malheureusement, il résulte des résultats du paragraphe précédent, que les FAI- $k$  à covariance polynomiale ne peuvent pas admettre de représentation stationnaire.

Mais en fait il n'est pas nécessaire d'être aussi exigeant. Il suffit, en effet, que la FAI- $k$   $Z$  admette une représentation

$Y(x)$  qui coïncide localement (i.e. sur le domaine étudié, ou bien sur chaque voisinage glissant) avec une FAST  $Y_1(x)$  (étant entendu que  $Y(x)$  pourra différer de  $Y_1(x)$  en dehors du domaine de travail).

Ceci conduit à la définition suivante : on dit qu'une FAI-k  $Z(x)$  est localement stationnaire sur un domaine borné  $D$  si elle admet une représentation  $Y(x)$  qui coïncide sur  $D$  avec une FAST  $Y_D(x)$ .

Il n'est pas vrai que toute FAI-k soit localement stationnaire (une trop grande régularité est, semble-t-il, incompatible avec cette propriété), mais cette propriété n'est pas exceptionnelle. En particulier, il est possible de démontrer que toute FAI-k à CG polynomiale est localement stationnaire sur tout ouvert borné.

D'après la formule (5-3), la covariance  $C(h)$  de la FAST  $Y_D(x)$  localement équivalente à une représentation de la FAI-k doit être de la forme :

$$C(x - y) = K(x - y) + a_1(y) f^1(x) + a_1(x) f^1(y)$$

On en déduit que  $C(h)$  ne peut différer de  $K(h)$  que par un polynome pair de degré  $\leq 2k$ . Noter que ce polynome dépend nécessairement des dimensions du domaine  $D$ , puisque la FAI-k n'est pas globalement stationnaire. Le tableau suivant donne, en fonction du Diamètre  $D$  de ce domaine (et pour  $k \leq 2$ ) la forme de la covariance stationnaire  $C(h)$  localement équivalente à la CG polynomiale ( $A$  est une constante quelconque, strictement positive).

$\text{in } \mathbb{R}^1 : C_1(h) = b_0 \left( \frac{D}{2} -  h  \right) + b_1 \left( \frac{ h ^3}{6} - \frac{D}{4} h^2 + \frac{D^3}{24} \right) \\ - b_2 \left( \frac{ h ^5}{120} - \frac{D}{48} h^4 + \frac{D^3}{48} h^2 - \frac{D^5}{240} \right)$
$\text{in } \mathbb{R}^2 : C_2(r) = A + \frac{\pi}{2} b_0 \left( \frac{D}{2} - \frac{2}{\pi} r \right) + \frac{9\pi}{2} b_1 \left( \frac{2}{9\pi} r^3 - \frac{D}{8} r^2 + \frac{D^3}{24} \right) \\ + \frac{225\pi}{2} b_2 \left( \frac{D^5}{240} - \frac{D^3}{96} r^2 + \frac{D}{128} r^4 - \frac{2}{225\pi} r^5 \right)$
$\text{in } \mathbb{R}^3 : C_3(r) = A + b_0(D-r) + b_1(r^3 - 2D r^2 + D^3) \\ + b_2(3 D^5 - 5 D^3 r^2 + 3 D r^4 - r^5)$

8 - Le krigeage en FAI-k.

Le modèle FAI-k ayant été justement construit dans ce but, il est clair que les équations du krigeage en FAI-k vont coïncider avec les équations (1-0) du krigeage universel, à ceci près évidemment que la covariance  $\sigma(h)$  doit être remplacée maintenant par la CG  $K(h)$ .

De fait, limitons nous, pour abrégier, au krigeage d'un point  $x_0$ . Nous voulons estimer la valeur numérique  $z(x_0)$  en ce point de la variable régionalisée  $z(x)$ , et cela à l'aide d'une combinaison linéaire  $\lambda^\alpha z_\alpha$  des valeurs  $z_\alpha = z(x_\alpha)$  observées aux points  $x_\alpha$ . Dans notre modèle,  $z(x)$  est interprétée comme une réalisation d'une certaine représentation  $Z(x)$  d'une FAI-k  $\tilde{Z}$ , dont nous avons par ailleurs estimé la CG  $K(h)$  (voir paragraphe 9). Pour qu'il soit possible, dans ce modèle, de minimiser la variance de l'erreur ( $\lambda^\alpha z_\alpha - Z(x_0)$ ), il faut évidemment que cette erreur soit une combinaison linéaire autorisée à l'ordre  $k$ . On doit donc avoir

$$(8-1) \quad \lambda^\alpha f_\alpha^1 = f^1(x_0)$$

c'est à dire les conditions d'universalité du système (1-0). Cette condition étant vérifiée, la variance d'estimation, c.a.d. la variance de la combinaison autorisée  $\lambda^\alpha z_\alpha - Z(x_0)$  est donnée, conformément à la relation (5-1), par la formule :

$$E[\lambda^\alpha z_\alpha - Z(x_0)]^2 = K_0 - 2\lambda^\alpha K_{0\alpha} + \lambda^\alpha \lambda^\beta K_{\alpha\beta}$$

Minimiser cette forme quadratique, sous la condition (8-1), conduit évidemment au système :

$$\begin{cases} \lambda^\alpha K_{\alpha\beta} = K_{\beta 0} + \mu_1 f^1 \beta \\ \lambda^\alpha f_\alpha^1 = f^1(x_0) \end{cases}$$

La boucle est ainsi refermée. Il reste à jeter un coup d'oeil sur les problèmes (d'ailleurs assez délicats) que soulève l'estimation de la CG  $K(h)$ .

### 9 - Le statut de la dérive en FAI-k.

Du fait que la notion de FAI-k couvre une classe de FA (les représentations) définies à un polynome près du type  $A_1 f^1(x)$ , on peut dire qu'elle prend implicitement en charge la notion de dérive (ou, plus exactement, ce qu'il pourrait y avoir de contenu positif dans cette notion ambiguë). Mais la dérive elle-même n'est pas définissable en tant que telle, sauf cas exceptionnel. Pour qu'il soit possible, en effet, de définir univoquement la dérive, il faudrait que la classe des représentations de la FAI-k  $\tilde{Z}$  contienne une FA privilégiée  $Y(x)$  se distinguant de toutes les autres par une propriété remarquable, et pouvant ainsi jouer le rôle de référence absolue: à toute autre représentation  $Z(x)$  serait alors univoquement associée une dérive  $Z(x) - Y(x)$  de la forme  $A_1 f^1(x)$ .

Le seul cas où cette circonstance se présente est celui d'une FAI-k admettant une représentation stationnaire  $Y_{St}(x)$ : toute autre représentation est alors du type :

$$Z(x) = Y_{St}(x) + A_1 f^1(x)$$

Ce n'est pas autre chose que le modèle du krigeage universel (dérive + résidu stationnaire), à ceci près que les coefficients de la dérive sont ici considérés comme des variables aléatoires.

Mais ce cas fait figure d'exception (cf paragraphe 6). En général, une FAI-k n'admet pas de représentation stationnaire, et ne comporte donc pas de dérive définissable univoquement (les modèles les plus utilisés, c.a.d. les FAI-k à covariance généralisée polynomiale, n'admettant jamais de représentation stationnaire, ni donc, non plus, de dérive). Loin de constituer une objection contre notre modèle, cette circonstance joue plutôt en sa faveur: notre but était, en effet, justement de prendre en charge des phénomènes pour lesquels la dichotomie

dérive + résidu est à peu près dépourvue de sens. L'essentiel est que le krigeage (estimation de points ou de valeurs moyennes) reste possible dans ce modèle, et nous avons vu que tel est bien le cas.

NB : Il arrive souvent que, pour une raison ou une autre (et souvent par simple routine), les utilisateurs tiennent, cependant, à estimer une dérive ou "tendance". Du fait que notre information sur le phénomène est, le plus souvent, très fragmentaire, nous devons soigneusement distinguer deux problèmes : celui de la définition (conceptuelle) de la notion de dérive proposée par l'utilisateur, et celui de l'estimation de la dérive ainsi définie.

Pratiquement, cela revient à poser à l'utilisateur la question suivante : à supposer qu'au lieu de connaître la variable en un petit nombre de points  $x_\alpha$  seulement, nous connaissions sa valeur  $z(x)$  en tous les points  $x$  du domaine qui nous intéresse, quel algorithme utiliseriez-vous pour calculer votre dérive  $m(x)$ ? (de fait, si la notion de dérive présente dans l'esprit de l'utilisateur possède un caractère opératoire, cet algorithme doit exister, puisque la dérive doit se trouver univoquement déterminée lorsque l'on connaît tout ce qu'il est possible de connaître sur le phénomène réel. Si cet algorithme n'existe pas, cela signifie que le mot "dérive" ne correspond qu'à une intuition confuse, dépourvue de contenu positif.

Le plus souvent, la réponse de l'utilisateur correspondra à une définition purement conventionnelle de la dérive, et se traduira par un algorithme linéaire (par exemple : ajuster à  $z(x)$  un polynôme de moindre carré, ou filtrer telle bande de basses fréquences). Nous pourrions alors, sur la base des  $z_\alpha$  expérimentaux, estimer cette dérive ainsi définie, en procédant au krigeage de l'algorithme correspondant.

Supposons par exemple que l'utilisateur définisse sa notion de dérive à l'aide d'un polynôme de moindre carré ajusté à  $z(x)$ . L'algorithme correspondant est de la forme

$$m(x) = b_1 f^1(x)$$

avec des coefficients  $b_1$  de la forme

$$b_1 = \int_S z(x) \varphi_1(x) dx$$

où les  $\varphi_1$  sont des polynomes convenablement choisis. Nous pourrions alors estimer par krigeage les coefficients  $b_1$ , qui dépendent linéairement de  $x$ . On obtient ainsi les estimateurs

$$b_1^* = \int_S z^*(x) \varphi_1(x) dx$$

$$m^*(x) = b_1^* f^1(x)$$

où  $z^*(x)$  est le krigeage du point  $x$  : incidemment, on remarque que cette estimation de la dérive s'identifie au polynome de moindre carré - ajusté au krigeage ponctuel  $z^*(x)$  - et diffère par conséquent, de façon notable du polynome que l'on aurait ajusté directement aux seules valeurs expérimentales  $z_\alpha$ .

## 10 - Estimation des covariances généralisées.

### 1 - Principes généraux

Par définition de la covariance généralisée, la variance d'un accroissement généralisé quelconque  $Z(\lambda_m)$  s'écrit :

$$E[Z(\lambda_m)]^2 = \sum_{i,j} \lambda_m^i \lambda_m^j K(x_i - x_j)$$

Pour un modèle polynomial d'ordre  $k$  :

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

D'où :

$$E[Z(\lambda_m)]^2 = C_0 \sum_i (\lambda_m^i)^2 + \sum_{p=0}^k (-1)^{p+1} b_p \sum_{i,j} \lambda_m^i \lambda_m^j |x_i - x_j|^{2p+1}$$

Si l'on pose

$$T_m^0 = \sum_i (\lambda_m^i)^2 \quad \text{et} \quad T_m^{2p+1} = (-1)^{p+1} \sum_{i,j} \lambda_m^i \lambda_m^j |x_i - x_j|^{2p+1}$$

il vient

$$(10-1) \quad E[Z(\lambda_m)^2] = c_0 T_m^0 + \sum_{p=0}^k b_p T_m^{2p+1}$$

C'est l'équation de la régression de  $Z(\lambda_m)^2$  par les  $(k+2)$  termes  $T_m^0; T_m^1, \dots, T_m^{2k+1}$ . Pour déterminer les coefficients de cette régression, on construit un grand nombre d'accroissements tels que  $Z(\lambda_m)$  et l'on minimise :

$$Q(c_0, b_0, \dots, b_k) = \sum_m w_m^2 [Z(\lambda_m)^2 - c_0 T_m^0 - \sum_{p=0}^k b_p T_m^{2p+1}]^2$$

Les poids  $w_m^2$  servent à égaliser les variances de  $Z(\lambda_m)^2$ , et devraient donc, en toute théorie, être égaux aux inverses de  $\text{Var}[Z(\lambda_m)^2]$ . Mais ces variances sont inconnues : elles font intervenir les moments d'ordre 4 de  $Z(\lambda_m)$  ou au moins - dans le cas gaussien - les moments d'ordre 2, qui s'expriment justement en fonction de la covariance à estimer. Leur calcul conduirait à une procédure itérative, mais il est inutile d'aller trop avant dans cette direction puisqu'une autre hypothèse de base des moindres carrés - la non-corrélation des  $Z(\lambda_m)^2$  - n'est de toute façon pas satisfaite. C'est pourquoi on préfère choisir les poids  $w_m$  en se basant sur l'intuition et sur l'expérience acquise. On peut par exemple prendre les poids associés à une covariance généralisée en  $-|h|$  ou en  $|h|^3$  dans le cas gaussien, où :

$$\text{Var}[Z(\lambda_m)^2] = 2[E Z(\lambda_m)^2]^2$$

il vient alors

$$w_m^2 = 1/(T_m^3)^2 \quad \text{si } k > 0 \quad \text{ou} \quad w_m^2 = 1/(T_m^1)^2 \quad \text{si } k = 0$$

De même qu'un variogramme est mieux connu aux courtes distances qu'aux grandes, on a ici intérêt à donner des poids plus forts aux accroissements qui sont "concentrés". C'est pourquoi on pondère par le modèle en  $|h|^3$  chaque fois que cela est possible.

Arrivés à ce stade, plusieurs questions se posent :

- comment choisir l'ordre  $k$  ?
- quelle est la meilleure régression pour (10-1) ?
- comment s'assurer que les coefficients de la régression satisfont aux contraintes sur la covariance généralisée ?

Une réponse à la dernière question consisterait à introduire les contraintes dans l'algorithme de minimisation en utilisant les techniques de programmation quadratique. Mais cette procédure serait très lourde, et l'on préfère essayer plusieurs équations de régression pour retenir celles dont les coefficients vérifient les contraintes : on est ainsi certain qu'au moins les covariances élémentaires, (c.a.d.  $K(h) = C_0 \delta, -|h|, |h|^3, -|h|^5, \text{etc...}$ ) seront admissibles. En pratique, la recherche se limite aux modèles en  $k \leq 2$ , ce qui fait 15 équations de régression à examiner.

Reste à choisir la meilleure des régressions admissibles. Quel doit être le critère de bon ajustement ? Les moindres carrés suggèrent de considérer la somme des erreurs quadratiques  $Q(C_0, \dots, b_k)$  ou sa version normalisée :  $Q(C_0, \dots, b_k)/Q(0, \dots, 0)$ . Sous l'hypothèse gaussienne, le rapport  $E[Q(C_0, \dots, b_k)]/E[Q(0, \dots, 0)]$  est égal à  $2/3$ . Le problème vient de ce que l'on ne dispose pas d'un test significatif, comme le test de Fischer-Snédecor dans le cas des moindres carrés classiques, et que l'on doit donc baser notre choix sur les simples fluctuations expérimentales. En fait, dans la pratique, plutôt que de se limiter au seul critère sur la forme  $Q$ , on considère simultanément plusieurs critères pour contrôler l'ajustement. Il en est de même pour le choix de l'ordre  $k$ .

Pour illustrer ce qui précède, nous donnons ci-après une brève description de la procédure implantée dans le programme BLUEPACK. Les méthodes utilisées dans d'autres programmes, tel que KRIGEPACK, ne sont pas fondamentalement différentes.

## 2 - Description de la procédure utilisée dans BLUEPACK

Elle comprend trois étapes :

- 1) trouver l'ordre  $k$  (0, 1 ou 2)
- 2) calculer les 15 régressions, éliminer celles qui ne conviennent pas et effectuer un premier choix.
- 3) comparer les trois meilleures régressions de la seconde étape pour parvenir au choix final.

Première étape : Pour déterminer  $k$ , l'idée est de supprimer à tour de rôle certaines valeurs et les estimer à l'aide des points expérimentaux du voisinage, en faisant varier uniquement le nombre de conditions de non-biais (1, 3 ou 6), c'est-à-dire l'ordre  $k$  de la FAI.

Pour chaque point expérimental  $Z_\alpha$  que l'on estime, les erreurs  $\hat{Z}_\alpha - Z_\alpha$  sont des accroissements généralisés d'ordre  $k = 0, 1$  ou  $2$ . Ces erreurs sont rangées par ordre de valeur absolue croissante et l'on calcule la moyenne des rangs sur tous les  $Z_\alpha$  estimés. On retient alors le degré  $k$  correspondant au rang moyen le plus faible. L'avantage de cette méthode réside dans sa robustesse vis-à-vis des valeurs expérimentales "aberrantes".

Le critère : moyenne des carrés des erreurs, est également donné par le programme, mais il sert uniquement comme contrôle de second ordre. Dans les cas ambigus, il est conseillé de retenir le degré  $k$  le plus bas possible.

Il convient de noter qu'à ce stade d'identification de la structure, la covariance n'est pas connue et que les estimateurs précédents sont calculés par moindres carrés. Cela ne pose pas de problèmes, pourvu que la configuration des points de données autour du point estimé ne soit pas trop symétrique. En effet un voisinage trop symétrique a tendance à filtrer de lui-même les polynômes, ce qui conduit à sous-estimer  $k$ . Pour éviter cela, les points expérimentaux sont divisés en 2 auréoles concentriques : les données de l'auréole interne servent à estimer les points de l'auréole extérieure, et vice versa.

Etape n° 2 : Une fois que  $k$  est déterminé, on doit effectuer une sélection parmi les régressions acceptables (par exemple pour  $k = 1$ , les équations avec un terme en  $|h|^5$  doivent être éliminées). Le critère utilisé porte sur la somme des erreurs quadratiques :

Soit  $\sigma_m^2 = E[Z(\lambda_m)^2]$  et  $\hat{\sigma}_m^2$  son estimateur

$$\hat{\sigma}_m^2 = \hat{C}_0 T_m^0 + \sum_{p=0}^k \hat{b}_p T_m^{2p+1}$$

Comme  $E[\sum_m Z(\lambda_m)^2] = \sum_m \sigma_m^2$

le rapport  $\rho = E[\sum_m Z(\lambda_m)^2] / E[\sum_m \hat{\sigma}_m^2]$  doit être voisin de 1.

La quantité  $r = \sum_m Z(\lambda_m)^2 / \sum_m \hat{\sigma}_m^2$  est un estimateur biaisé de  $\rho$ .

Pour réduire ce biais, nous utilisons un critère (le "jack-knife") où interviennent les rapports  $r_1$  et  $r_2$  calculés séparément sur chacune des deux auréoles :

$$\hat{\rho} = 2r - \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2} \quad (n_1, n_2 : \text{nombre de points de chaque auréole})$$

On retient la régression qui conduit au  $\hat{\rho}$  le plus voisin de 1.

Etape n° 3 : Cette étape est optionnelle. Elle consiste à kriger réellement quelques points expérimentaux (30 à 50) comme s'ils étaient inconnus, en utilisant l'ordre  $k$  déterminé dans l'étape n° 1 et les trois meilleurs modèles de covariance issus de l'étape n° 2 (classés par distance croissante du critère  $\hat{\rho}$  à la valeur 1). La moyenne expérimentale des carrés des erreurs ( $Z_\alpha^* - Z_\alpha$ ) et celle des carrés des erreurs normées ( $Z_\alpha^* - Z_\alpha$ )/ $\hat{\sigma}_\alpha$  sont calculées et fournissent un critère de décision pour le choix final.

### 3 - Contrôle de la validité du modèle par estimation des points expérimentaux

Il s'agit de l'application systématique de la procédure décrite ci-dessus, mais en utilisant uniquement un modèle spécifié de covariance, et pas nécessairement une covariance polynomiale. En fait, cette méthode est très utile pour tester les modèles de variogrammes ajustés avec les moyens classiques. Si l'ajustement est correct, la moyenne des carrés des erreurs de krigeage  $Z_\alpha^* - Z_\alpha$  est égale à la variance de krigeage théorique  $\sigma_\alpha^2$ . Par conséquent, le rapport

$$\frac{1}{N} \sum_{\alpha} (Z_\alpha^* - Z_\alpha)^2 / \hat{\sigma}_\alpha^2$$

doit être voisin de 1.

A titre indicatif, signalons que la variance de ce rapport est égale à  $2/N$  si les variables  $Z_\alpha^* - Z_\alpha$  sont indépendantes et gaussiennes.

L'étude des erreurs de krigeage  $Z_\alpha^* - Z_\alpha$  ou de leur valeur normée  $(Z_\alpha^* - Z_\alpha) / \hat{\sigma}_\alpha$  peut être menée à l'aide d'outils statistiques simples tels que histogrammes ou nuages de corrélation.

### 4 - Commentaires

Les procédures utilisées jusqu'alors dans les programmes sont empiriques et peuvent sans nul doute être améliorées. Mais il n'est pas évident que l'on aboutisse à des améliorations vraiment notables pour justifier d'un tel travail.

On peut penser que l'une des faiblesses de la méthode provient de ce que les covariances généralisées utilisées sont isotropes. L'inférence de modèles de covariances anisotropes, qui avait été essayée un moment, a dû être abandonnée car elle requiert un trop grand nombre de paramètres. Cette limitation à des modèles isotropes ne serait pas possible dans le cas stationnaire : en revanche dans le cas non-stationnaire, on peut supposer que les anisotropies sont prises en compte par les polynômes filtrés (la "dérive"). Pour  $k = 0$ , les seuls termes filtrés sont les constantes, ce qui est une bonne raison,

entre autres, pour revenir à une analyse structurale habituelle. Une autre raison est qu'un modèle avec portée peut être mieux approprié qu'un variogramme linéaire.

La véritable faiblesse de la méthode vient de son manque de robustesse vis à vis des variables qui ne vérifient pas bien les hypothèses intrinsèques : présence de quelques pics isolés, mélange de zones plates et de zones chahutées, données erronées ou "aberrantes". Ces hétérogénéités ont une grande influence sur le critère à minimiser  $Q(C_0, \dots, b_k)$ . Comme nous l'avons vu ci-dessus, nous disposons heureusement de moyens de contrôle, mais seulement après coup. Pour les futurs développements, il serait peut-être fructueux de s'orienter vers l'utilisation des algorithmes de "régression robuste" proposés par P. Huber. Mais nous devons nous attendre à une augmentation des temps de calculs et de l'espace mémoire nécessaires.

## 11 - Conclusions

Après ce parcours à travers quelques unes des complexités de la théorie des FAI-k (et beaucoup d'autres ont été ignorées), une question vient à l'esprit du praticien : cet effort se justifie-t-il vraiment ? Quels sont les résultats qui n'auraient pu être obtenus avec le simple modèle du krigeage universel ? En fait, deux résultats :

(i) La résolution du problème d'estimation linéaire s'est affranchie du concept plutôt confus de "dérive". Fondée uniquement sur la stationnarité des accroissements généralisés, l'approche en FAI-k ne nécessite plus la dichotomie entre dérive et résidus. Pour atteindre les mêmes buts que dans le modèle de krigeage universel, on fait l'économie de nombreuses hypothèses et de nombreux paramètres.

(ii) Une nouvelle famille de covariances est apparue : les covariances polynomiales, qui dépendent linéairement d'un petit nombre de paramètres, ce qui permet de les identifier au moyen de procédures automatiques.

On envisageait initialement une procédure entièrement automatisée. De fait pour que le krigeage devienne une méthode de routine pour l'interpolation de données non-stationnaires, on ne pouvait le faire dépendre plus longtemps des analyses particulièrement laborieuses des variogrammes des résidus. Les covariances polynomiales sont apparues comme l'outil particulièrement adapté à une reconnaissance automatique, au point d'envisager qu'il suffirait de laisser tourner le programme, depuis l'inférence des paramètres jusqu'aux équations de krigeage, sans avoir à intervenir entre les deux phases. Mais l'expérience a très tôt révélé les risques d'une telle utilisation en "aveugle" et le point de vue qui prévaut maintenant est bien plus modéré.

Plusieurs critères ont été proposés pour contrôler la qualité d'un ajustement, depuis les plus simples jusqu'à la méthode plus sophistiquée de validation croisée par estimation des points expérimentaux. Le temps à consacrer à l'ajustement d'un modèle structural dépend de la confiance que l'on met dans ses propres données et des objectifs poursuivis. Si l'objectif est seulement d'obtenir une carte sommaire, un contrôle de l'ajustement avec les critères les plus simples sera suffisant. Si le problème est plus important, on peut choisir plusieurs modèles et les comparer avec les modèles polynomiaux. En général, une étude assez fine comprendra un calcul de variogrammes selon les méthodes classiques, des ajustements et un test des ajustements par krigeage des points expérimentaux. De plus, toute information supplémentaire sur la physique du phénomène doit être prise en compte, par exemple pour suggérer des transformations sur les variables, des modèles de dérive, des comportements à petite échelle...

Nous ne prétendons pas que les covariances polynomiales généralisées constituent un outil universel. En particulier dans le cas "stationnaire" ( $k = 0$ ), la classe des modèles : variogramme linéaire plus effet de pépite n'est pas assez riche et les covariances bornées à portée finie peuvent être mieux appropriées. Mais l'approche en FAI- $k$  servira à indiquer que  $k = 0$  et à orienter l'utilisateur vers les méthodes classiques. D'autre part, dans le cas où l'on a clairement intérêt à prendre une dérive localement linéaire ou localement quadratique, les covariances polynomiales.

autorisent des modèles à trois ou quatre paramètres, ce qui est habituellement suffisant pour un bon ajustement.

En conclusion, le modèle des FAI-k à covariances polynomi-ales n'est pas forcément toujours le meilleur modèle possible mais il marche généralement bien et il reste jusqu'à ce jour sans concurrent pour sa rapidité de réponse.

## BIBLIOGRAPHIE

REFERENCES SUR LES FAI- $k$ A - SUR LA THEORIE

- CHILES, J.P. (1977) : Géostatistique des phénomènes non stationnaires (dans le plan). Thèse de Docteur-Ingénieur, Université de Nancy.
- DELFINER, P. (1975) : Linear estimation of non-stationary spatial phenomena. Proceedings of NATO ASI, Advanced Geostatistics in the Mining Industry, Rome 1975 - D. Reidel Publishing Co., Dordrecht, pp. 49-68.
- (1977) : Shift Invariance under Linear Models. Ph.D. Thesis, Princeton University.
- GUELFAND, M. and VILENKIN, N.Y. (1967) : Les Distributions, Vol. 4 Dunod, Paris (cf. Chapter 2, Section 4 and Chapter 3, section 3 on generalized processes with stationary  $n$ th order differences).
- MATHERON, G. (1971) : La théorie des fonctions aléatoires intrinsèques généralisées.- Centre de Géostatistique - Tech. Report No. 117.
- (1972a) : Leçons sur les fonctions aléatoires d'ordre 2.- Cours E.N.S.M.P.
- (1972b) : Les covariances généralisées polynomiales.- Centre de Géostatistique - Tech. Report No. 123.
- (1973) : The intrinsic random functions and their applications - Adv. Appl. Prob. 5, pp. 439-468.
- (1974) : Représentations stationnaires et représentations minimales pour les FAI- $k$ .- Centre de Géostatistique - Tech. Report No. 125.
- YAGIOM, A.M. (1962) : Stationary Random Functions. Dover Publications, New York.

B - SUR LES APPLICATIONS

- CHILES, J.P. (1977a) : Géostatistique des phénomènes non stationnaires (cf. above).
- (1977b) : SIMPACK - Notice d'utilisation.- Centre de Géostatistique.
- CHILES, J.P. and DELFINER, P. (1974) : Reconstitution par krigeage de la surface topographique à partir de divers schémas d'échantillonnage photogrammétrique.- Société Française de Photogrammétrie, Bull. n° 57, Janvier 75, pp. 42-50

- DELFINER, P. and CHILES, J.P. (1977) : Conditional simulation : a new Monte-Carlo approach to probabilistic evaluation of hydrocarbon in place. (Submitted to the Society of Petroleum Engineers Journal).
- DELFINER, P. and DELHOMME, J.P. (1973) : Optimum Interpolation by Kriging. Proceedings of NATO ASI, Display and Analysis of Spatial Data, Nottingham 1973 , Wiley and Sons, pp. 96-114.
- DELFINER, P., DELHOMME, J.P. and CHILES, J.P. (1976) : BLUEPACK Manual - Centre de Géostatistique.
- DELFINER, P. and GILBERT, R.O. (1978) : Combining two types of survey data for estimating geographical distribution of plutonium (submitted to the Journal of the American Statistical Association).
- TOTAL - Compagnie Française des Pétroles et Société Nationale ELF-Aquitaine (1973) : KRIGEPACK Manual.

REFERENCES SUR LES REGRESSIONS ROBUSTES

- HOGG, R.V. (1974) : Adaptive Robust Procedures : A Partial Review and some Suggestions for Future Applications and Theory. Journal of the American Statistical Association, Vol. 69, pp. 909-927.
- HUBER, P.J. (1973) : Robust Regression : Asymptotics, Conjectures and Monte-Carlo.- Annals of Statistics, Vol. 1, pp. 799-821.
- MOSTSELLER, F. and TUKEY, J.W. (1977) : Data Analysis and Regression Addison-Wesley Publishing Company.