D-281-III

CENTRE DE PERFECTIONNEMENT TECHNIQUE

C.P.T.

LE KRIGEAGE UNIVERSEL

RECHERCHE D'ESTIMATEURS OPTIMAUX EN PRESENCE D'UNE DERIVE

par

G. MATHERON

Ingénieur en Chef des Mines

Docteur es-Sciences

Directeur du Centre de Morphologie Mathématique

Ecole Nationale Supérieure des Mines de Paris

Communication faite
aux
JOURNEES D'ETUDE
DE
QUELQUES TENDANCES NOUVELLES
DE
LA GEOLOGIE

—— PARIS ——
11-12 MARS 1970

LE KRIGEAGE UNIVERSEL

(Recherche d'estimateurs optimaux en présence d'une dérive)

0 - INTRODUCTION

0.1 - CRITIQUE DES METHODES DE MOINDRES CARRES

Nous nous proposons, dans ce qui suit, de formuler en termes de fonctions aléatoires non stationnaires le problème de l'estimation des dérives* *(ou tendances)*, et de montrer que ce problème admet une solution optimale, bien différente de celles auxquelles conduisent les méthodes dites de "trend surface analysis" : ces dernières, qui consistent en général à ajuster un polynome par une méthode de moindres carrés, donnent un sentiment de malaise; on a l'impression, en effet, de faire violence à la nature, en lui imposant de force une expression polynomiale, qui n'a aucune raison à priori de présenter le moindre rapport avec la structure réelle du phénomène que l'on veut représenter. D'une manière plus précise, nous adresserons trois reproches principaux à ces méthodes de moindres carrés :

Tout d'abord, ces méthodes entraînent souvent une confusion du mode opératoire et du concept. Peu d'auteurs se donnent la peine de définir la signification de ce "trend" qu'ils estiment par une méthode de moindres carrés. On a souvent l'impression que ce fameux trend n'est rien de plus que le résultat numérique auquel conduit un mode opératoire - c'est-à-dire, peut-être, un pur et simple artefact -. A l'analyse, il semble que le terme peu clair de "trend" se réfère tour à tour, et parfois même simultanément dans un même contexte, à trois notions bien distinctes au moins : si $Z(x)$ est la variable régionalisée à laquelle on s'intéresse, et $P(x)$ le polynome ajusté par moindres carrés à partir des valeurs expérimentalement connues aux points x_1, x_2, ..., on peut attribuer à la valeur en x de $P(x)$ l'une ou l'autre des significations incompatibles suivantes :

a - $P(x)$ est *(une estimation de)* l'espérance à priori $E(\mathbf{Z}(x))$: c'est toujours ce sens a/ que nous attribuerons au terme "dérive".

b - $P(x)$ est une estimation de la vraie valeur *(inconnue)* $\mathbf{Z}(x)$ prise par la variable régionalisée au point x : c'est à ce sens b/ que se rattache notre terme "krigeage ponctuel". Plus précisément, le krigeage ponctuel sera le meilleur estimateur linéaire de $\mathbf{Z}(x)$ construit à partir des données disponibles $\mathbf{Z}(x_1)$, $\mathbf{Z}(x_2)$, ...

c - On attribue parfois, enfin, à $P(x)$ le sens d'une "moyenne mobile". $P(x_0)$ serait alors *(une estimation de)* la valeur moyenne de $\mathbf{Z}(x)$ dans une zone plus ou moins grande *(à préciser)* entourant un point donné x_0. Dans notre terminologie, le krigeage désignera le meilleur estimateur linéaire de cette moyenne mobile.

Ces distinctions ont une grande importance théorique et pratique. Il est essentiel de définir avec précision l'objectif que l'on vise *(a/, b/, ou c/)*. En cartographie sous-marine (1), ce sont les vraies valeurs $\mathbf{Z}(x)$ que l'on doit estimer et cartographier *(cas b/)*. En exploitation minière (3), (4), (5), on s'intéresse à la teneur moyenne d'un panneau de taille donnée *(cas c/)*.

---

* Pour éviter tout anthropomorphisme, nous utiliserons toujours le terme "dérive" au lieu de "tendance", ou "trend".

Dans certaines études à caractère plus fondamental, enfin, où l'on cherche à reconstituer les mécanismes généraux qui ont donné naissance au phénomène que l'on étudie, c'est plutôt dans la dérive elle-même *(cas a/)* que l'on peut espérer trouver un reflet de la structure de ces mécanismes.

Nous en arrivons ainsi au deuxième grief que l'on peut formuler à l'encontre des méthodes de moindres carrés. Il n'est pas possible que le même polynome $P(x)$ résolve à la fois les trois problèmes a/, b/ et c/. Nous verrons que $P(x)$ constitue en fait une solution du problème a/. Mais, en général, ce n'est pas la meilleure solution possible : ces méthodes passe-partout qui utilisent toujours les mêmes polynomes, quelles que soient les caractéristiques structurales du phénomène que l'on étudie, n'ont aucune chance en général de conduire à un optimum.

En troisième lieu, enfin, les méthodes de moindre carré ne permettent en aucune façon d'évaluer l'erreur que l'on commet en estimant la dérive à l'aide du polynome $P(x)$. La variance des résidus, contrairement à ce que l'on croit parfois, n'est pas une variance d'estimation : la variance des écarts $Z(x_1) - P(X_1)$ aux points x_1 où les données sont connues est, par construction, systématiquement plus faible *(et même beaucoup plus faible)* que celle de l'écart $Z(x) - P(x)$ en un point x différent des x_1. Ainsi, la variance des résidus n'est pas la variance d'estimation de $Z(x)$ *(problème b/)*. Elle n'est pas davantage la variance d'estimation de la dérive *(problème a/)*, car il n'y a cette fois plus aucun lien conceptuel entre un écart $Z(x_1) - P(x_1)$ et la qualité de $P(x_1)$ considéré comme un estimateur de l'espérance à priori $E[Z(x_1)]$.

Cependant, si les méthodes de moindres carrés se révèlent peu satisfaisantes, l'objectif qu'elles visaient, et qu'elles n'atteignent pas ou atteignent mal, correspond, lui, à un problème très réel et très important. Il existe réellement des phénomènes que l'on ne peut absolument pas assimiler à des *(réalisations de)* fonctions aléatoires stationnaires. Pour reprendre l'exemple de la cartographie des fonds sous-marins, il est certain que la profondeur va en augmentant lorsqu'on s'éloigne des côtes. Nous allons essayer de formuler et de résoudre dans le cadre de la théorie des fonctions aléatoires non stationnaires les problèmes importants que posent ces dérives.

02 - POSITION DU PROBLEME ET HYPOTHESES GENERALES

La variable régionalisée $z(x)$ que l'on étudie sera interprétée comme une réalisation d'une fonction aléatoire $Z(x)$ non stationnaire en général. Dans un premier groupe d'hypothèses, nous supposerons que $Z(x)$ admet des moments d'ordre 1 et 2 :

$$(I)\qquad \begin{cases} E[Z(x)] = m(x) \\ E[Z(x)\,Z(y)] = m(x)\,m(y) + C(x,y) \end{cases}$$

Nous dirons que la fonction $m(x)$ est la dérive de la F.A. $Z(x)$. Il nous semble, en effet, que la seule définition conceptuellement claire de la notion de dérive est bien celle-ci : moment d'ordre 1 d'une F.A. non stationnaire. La fonction $C(x,y)$ - qui dépend séparément des deux points x et y est la covariance *(non stationnaire)* habituelle.

Dans bien des cas *(comme le montre la pratique de la Géostatistique)*, (3), (5), ces hypothèses (I) seront encore trop restrictives, et devront être remplacées par les hypothèses (I') suivantes qui expriment que les accroissements de la fonction aléatoire $Z(x)$ *(et non $Z(x)$ elle-même)* admettent des moments d'ordre 1 et 2 :

$$(I')\qquad \begin{cases} E[Z(x) - Z(y)] = m(x) - m(y) \\ \frac{1}{2} D^2[Z(x) - Z(y)] = \gamma(x,y) \end{cases}$$

La dérive $m(x)$ n'est plus ici déterminée qu'à une constante près, car, en général, l'espérance $E[Z(x)]$ n'existe plus : le point de vue des hypothèses (I') revient donc à étudier la fonction aléatoire $Z(x)$ à une constante additive près. La fonction $\gamma(x,y)$ est le demi-variogramme habituel. *(Dans l'espace à une seule dimension, des processus aussi usuels que le mouvement brownien ou les processus de Poisson donnent des exemples simples de F.A. sans espérance, vérifiant les hypothèses (I') mais non les hypothèses (I)).*

Nous avons à résoudre les deux problèmes *(liés, mais distincts)* suivants : connaissant les valeurs numériques prises sur un ensemble S *(l'ensemble des points où l'on dispose de données expérimentales)* par une réalisation $z(x)$ de la F.A. $Z(x)$, nous devons :

1°) estimer la fonction $m(x)$ *(sur S et à l'extérieur de S)* : c'est le problème a/ ci-dessus,

2°) estimer $z(x)$ en un point $x \notin S$ *(problème b/)*, ou, plus généralement, estimer une "moyenne mobile" $\int \mu(dx)\, z(x)$ où μ est une mesure donnée dont le support est disjoint de S *(problème c/)*.

De plus, nous devons être capables de représenter par des variances d'estimation les erreurs commises dans ces opérations, et nous efforcer de choisir nos estimateurs de manière à minimiser *(si faire se peut)* cette variance d'estimation.

En ce qui concerne ce dernier point, indiquons que nous nous limiterons en fait à rechercher le meilleur estimateur linéaire de $m(x)$ ou de $\int \mu(dx)\, z(x)$ que l'on puisse former à partir des valeurs numériques des $z(y)$, $y \in S$: les estimateurs non linéaires sont beaucoup trop compliqués pour qu'il soit possible de les mettre en oeuvre dans les applications, et d'autre part leurs propriétés ne sont plus liées aux seuls moments d'ordre 1 et 2, qui figurent dans les hypothèses (I) ou (I'), mais font intervenir la totalité de la loi spatiale de la F.A. $Z(x)$ - *(dans le cas où cette loi spatiale est gaussienne, on sait d'ailleurs que le meilleur estimateur possible coïncide avec le meilleur estimateur linéaire)*.

Ainsi formulé en toute généralité *(les fonctions $m(x)$ et $C(y,x)$ étant complètement inconnues)* notre problème est manifestement insoluble, et d'ailleurs probablement dépourvu de sens. Mais, dans les problèmes concrets, la notion de dérive ne peut présenter une signification réelle que si la fonction $m(x)$ correspondante varie d'une manière continue et régulière relativement à l'échelle à laquelle on travaille *(et aux données expérimentalement disponibles)* : si la fonction $m(x)$ était irrégulière et chaotique à cette échelle, on devrait la considérer elle-même comme une réalisation d'une nouvelle fonction aléatoire. Du point de vue de l'interprétation physique *(et non mathématique)* la notion de dérive est ainsi manifestement liée à celle d'échelle. Si l'on étudie des cotes topographiques, à l'échelle de la dizaine de mètres, la notion de montagne se traduit par une dérive; à l'échelle de la dizaine de kilomètres, il ne lui correspond plus qu'une fonction aléatoire, et la dérive, à cette échelle, exprimera plutôt la notion de chaîne de montagnes *(cf. aussi la notion de structure gigogne, (7))*.

Mais cette condition de régularité imposée à priori à la fonction $m(x)$ pour que la notion de dérive ait un contenu physique réel, entraîne aussi qu'une estimation de $m(x)$ doit toujours être plus ou moins possible localement. Cette condition exprime en effet que - sur un certain voisinage V d'un point x_0 donné - la fonction $m(x)$ peut être approchée avec une excellente précision par une fonction de la forme :

$$m(x) \simeq \varphi(x) = \sum_{\ell=0}^{k} a_\ell f^\ell(x) \tag{II}$$

où les $f^\ell(x)$ sont des fonctions connues, choisies une fois pour toutes *(par exemple des polynomes)*, et les a_ℓ des coefficients inconnus : quant au voisinage V de x_0 sur lequel l'approximation (II) est acceptable, sans être très grand, il doit - si le problème a un sens - contenir un nombre suffisant de points expérimentaux pour qu'il soit possible d'estimer les k+1 coefficients inconnus a_0, a_1, ... a_k -

Il reste le problème de la fonction $C(x,y)$ ou $\gamma(x,y)$ sur laquelle on ne sait rien, non plus, à priori. Mais, ici encore, et pour les mêmes raisons, on peut supposer que C ou γ sont localement assimilables à des fonctions de type connu et ne se déforment qu'assez lentement dans l'espace à l'échelle à laquelle on travaille.

Le cas le plus favorable sera celui d'une fonction γ *(ou C)* de la forme $\gamma(x,y) = \varpi \gamma_0(x-y)$ où γ_0 est une fonction connue, et ϖ un facteur lentement variable, que l'on pourra regarder comme constant sur le voisinage V du point x_0 : les équations qui déterminent les estimateurs optimaux étant linéaires et homogènes, ces estimateurs ne dépendront que de γ_0 et non du facteur ϖ, ce dernier ne se répercutera donc que sur les variances d'estimation et non sur les estimateurs eux-mêmes. Dans certains cas, on pourra prendre l'expression très simple :

$$\gamma(x,y) = \varpi r \qquad (r = |x-y|)$$

(variogramme linéaire). Il suffit, pour cela, en effet, que le vrai variogramme ou la vraie covariance ait un comportement linéaire au voisinage de $x = y$ et jusqu'à des distances comparables aux dimensions du voisinage V ci-dessus : cette circonstance se rencontre plus souvent qu'on ne croit. *(cf. (1))*.

En général, pourtant, en plus du facteur ϖ, il conviendra d'introduire un ou plusieurs paramètres supplémentaires. Par exemple, on prendra

$$\gamma(x,y) = \varpi\, r^{\alpha-1} \qquad (0 > \alpha > 2)$$

ou encore

$$C(x,y) = \varpi\, e^{-ar}$$

Dans le premier cas, le paramètre α est en relation avec le degré de continuité de la variable régionalisée. Dans le deuxième *(covariance exponentielle)*, le paramètre a, ou plutôt son inverse, donne la mesure de la portée du phénomène *(distance au-delà de laquelle les corrélations s'éteignent)*.

Naturellement, le contrôle expérimental d'une hypothèse de ce genre et l'estimation des paramètres correspondants *(ϖ et surtout α ou a)* posera des problèmes assez délicats de statistique mathématique : en raison de leur extrême importance pour les applications, nous consacrerons à ce problème la section III de cette étude.

En résumé le problème que nous devons traiter se trouve schématisé comme suit : on a une F.A. $Z(x)$ vérifiant les hypothèses (I) ou (I'). Sur un voisinage V d'un point x_0, la dérive $m(x)$ est de la forme (II) avec des coefficients a_ℓ inconnus. On connaît à priori *(éventuellement à un facteur près)* la covariance $C(x,y)$ ou le variogramme $\gamma(x,y)$. Enfin, on connaît les valeurs numériques de *(la réalisation de)* $Z(x)$ pour les points x appartenant à un ensemble $S \subset V$. On veut former les meilleurs estimateurs linéaires :

1°) des coefficients a_ℓ inconnus de la dérive;

2°) de la valeur numérique de *(la réalisation de)* $Z(x)$ en $x_0 \notin S$ *(mais $x_0 \in V$)* ou de $\int \mu(dx)\, Z(x)$ pour une mesure μ dont le support est disjoint de S;

3°) nous voulons de plus contrôler la validité de l'hypothèse que nous avons faite en choisissant pour la covariance ou le variogramme une expression mathématique particulière, et estimer les paramètres dont cette expression dépend.

Nous nous contentons, dans ce qui suit, de traiter le cas le plus simple *(et le plus utile en pratique)*, celui où l'ensemble des données est fini. On trouvera un exposé complet et rigoureux en (9).

Nous traiterons d'abord l'estimation de la dérive (1-1) et le krigeage (1-2 et 1-3) dans le cas où il existe une covariance *(connue)* $C(x,y)$, c'est-à-dire dans le cadre des hypothèses (I). Dans la sous-section 1-4, nous transposerons ces résultats au cas où il n'existe qu'un variogramme $\gamma(x,y)$.

1.1 - ESTIMATION OPTIMALE DE LA DERIVE *(Covariance connue)*

Soit $Z(x)$ une F.A. vérifiant les hypothèses (I) avec :

$$m(x) = \sum_{\ell=0}^{k} a_\ell\, f^\ell(x)$$

Les $f^\ell(x)$ sont des fonctions connues, mais les coefficients a_ℓ sont inconnus et le problème consiste à les estimer. L'ensemble S où l'on connaît les valeurs prises par la réalisation de $Z(x)$ est constitué de n points x_α, $\alpha = 1,2,\ldots n$.

Soit x_0 un point *(appartenant ou non à S)* et $b^\ell = f^\ell(x_0)$. Pour estimer la dérive en x_0, soit :

$$m(x_0) = \sum_\ell a_\ell b^\ell$$

nous allons former une combinaison linéaire :

(1-1) $$M = \sum_{\alpha=1}^{n} \lambda^\alpha Z_\alpha$$

et choisir les coefficients λ^α de manière à remplir les deux conditions suivantes, quels que soient les coefficients a_ℓ ($\ell = 0, 1, \ldots k$) :

~ L'espérance E(M) doit être égale à $m(x_0) = \sum a_\ell b^\ell$

~ La variance $D^2(M)$ doit être minimale.

Pour abréger les notations, nous poserons :

$$Z(x_\alpha) = Z_\alpha,\ f^\ell(x_\alpha) = f^\ell_\alpha,\ C(x_\alpha, x_\beta) = \sigma_{\alpha\beta}$$

La première condition exprime que M est un estimateur sans biais. Elle s'écrit :

$$\sum_\alpha \sum_\ell \lambda^\alpha a_\ell f^\ell_\alpha = \sum_s a_s b^s$$

Cette condition est vérifiée quels que soient les a_ℓ si et seulement si on a :

(1-2) $$\sum_\alpha \lambda^\alpha f^\ell_\alpha = b^\ell \qquad (\ell = 0, 1, \ldots k)$$

Nous dirons qu'un estimateur (1-1) de $m(x_0)$ est universel si les coefficients λ^α vérifient (1-2), c'est-à-dire si cet estimateur est sans biais quels que soient les a_ℓ. L'estimateur des moindres carrés, par exemple, est un estimateur universel *(Annexe 3)*. Mais ce n'est pas le meilleur possible.

Moyennant la condition (1-2), on a $E(M) = m(x_0)$, et par suite $E\left[(M-m(x_0))^2\right] = D^2(M)$: la variance de M, qui ne dépend que des $\sigma_{\alpha\beta}$ et non des a_ℓ, donne une mesure de l'erreur commise en estimant $m(x_0)$ à l'aide de M. Cette variance a pour expression :

(1-3) $$D^2(M) = \sum_{\alpha,\beta} \lambda^\alpha \lambda^\beta \sigma_{\alpha\beta}$$

L'estimateur universel optimal correspond donc aux coefficients λ^α qui minimisent la forme quadratique (1-3) compte tenu des conditions (1-2). Le formalisme classique de Lagrange conduit au système :

(1-4) $$\begin{aligned} \sum_\beta \lambda^\beta \sigma_{\alpha\beta} &= \sum_\ell \mu_\ell f^\ell_\alpha \qquad (\alpha = 1, 2, \ldots n) \\ \sum_\alpha \lambda^\alpha f^\ell_\alpha &= b^\ell \qquad (\ell = 0, 1, \ldots k) \end{aligned}$$

La matrice des $\sigma_{\alpha\beta} = C(x_\alpha, x_\beta)$ est toujours strictement définie positive. Nous verrons dans la section 2 que le système (1-4) de 1 + k + n équations à 1 + k + n inconnues *(les n coefficients λ^α et les k+1 paramètres de Lagrange)* est régulier si et seulement si les fonctions *(connues)* $f^\ell(x)$ sont linéairement indépendantes sur S, autrement dit si :

$$\sum_\ell c_\ell f^\ell_\alpha = 0 \ (\alpha = 1, 2, \ldots n) \Rightarrow c_\ell = 0 \qquad (\ell = 0, 1, \ldots k)$$

Nous supposerons toujours cette condition remplie. La solution unique de (1-4) donne alors l'estimateur universel optimal cherché. La variance de cet estimateur optimal est donnée par :

(1-5) $$D^2(M) = \sum_\ell \mu_\ell b^\ell$$

On le voit sans peine en multipliant par λ^{α} la première relation (1-4) et en sommant en α. Comme cette variance ne dépend que des paramètres de Lagrange μ_{ℓ}, on l'obtient comme un sous-produit immédiat de la résolution du système (1-4).

D'autre part, la condition (1-2) est linéaire en b^{ℓ}. Par suite, les coefficients λ^{α} et les paramètres de Lagrange sont eux-mêmes linéaires en b^{ℓ}, soit :

$$\begin{cases} \lambda^{\alpha} = \sum_{\ell} \lambda^{\alpha}{}_{\ell} \, b^{\ell} \\ \mu_{\ell} = \sum_{s} \mu_{\ell s} \, b^{s} \end{cases}$$

Les matrices $\lambda^{\alpha}{}_{\ell}$ et $\mu_{\ell s}$ constituent la solution unique du système suivant :

(1-D)
$$\begin{cases} \Sigma_{\beta} \lambda^{\beta}{}_{\ell} \, \sigma_{\alpha\beta} = \Sigma_{s} \mu_{s \ell} \, f^{s}{}_{\alpha} & (\alpha = 1, \dots n, \ \ell = 0, 1, \dots k) \\ \Sigma_{\alpha} \lambda^{\alpha}{}_{\ell} \, f^{s}{}_{\alpha} = \delta^{s}_{\ell} \end{cases}$$

(avec $\delta^{s}_{\ell} = 1$ pour $\ell = s$ et 0 pour $\ell \neq s$). Le vecteur :

(1-6)
$$A_{\ell} = \Sigma_{\alpha} \lambda^{\alpha}{}_{\ell} \, Z_{\alpha}$$

constitue alors l'estimateur universel optimal du vecteur dérive a_{ℓ} :

On a $E(A_{\ell}) = a_{\ell}$ quelles que soient les vraies valeurs inconnues des a_{ℓ} et la variance $D^2 (\sum_{\ell} A_{\ell} b^{\ell})$ est minimale pour tout vecteur b^{ℓ}. De plus, la matrice des paramètres de Lagrange $\mu_{\ell s}$ coïncide avec la matrice des covariances des A_{ℓ} :

(1-7)
$$\mu_{\ell s} = \mathrm{Cov}\,(A_{\ell} A_s) = E(A_{\ell} A_s) - a_{\ell} a_s$$

On l'obtient à titre de sous-produit de la résolution du système fondamental (1-D). Notons enfin que ce système de $n(k+1) + (k+1)^2 = (k+1)(n+k+1)$ équations à $n(k+1) + (k+1)^2$ inconnues *(les $\lambda^{\alpha}{}_{\ell}$ et les paramètres de Lagrange $\mu_{\ell s}$.)* est régulier si, et seulement si, les $f^{\ell}(x)$ sont linéairement indépendantes sur S. On notera aussi, d'après (1-7), que l'on a toujours $\mu_{\ell s} = \mu_{s \ell}$.

Remarque : Dans le cas gaussien, l'estimateur optimal universel de la dérive, solution du système fondamental (1-D) coïncide avec l'estimateur du maximum de vraisemblance (voir Annexe 2).

1-2 - LE KRIGEAGE UNIVERSEL *(à covariance connue)*

Proposons-nous maintenant de procéder à l'estimation d'une "moyenne pondérée" de $Z(x)$, c'est-à-dire d'une expression de la forme :

(1-8)
$$Z = \int Z(x) \, p(dx)$$

où $p(dx)$ est une mesure donnée dont le support est disjoint de S, et où $Z(x)$ est la réalisation de notre fonction aléatoire. Nous allons pour cela former une combinaison linéaire des données disponibles (les $Z_{\alpha} = Z(x_{\alpha})$), soit :

(1-9)
$$Z^{*} = \Sigma_{\alpha} \lambda^{\alpha} Z_{\alpha}$$

et déterminer les coefficients λ^{α} grâce à des conditions convenables. Si l'on considère $Z(x)$ non plus comme une réalisation, mais comme la F.A. elle-même, Z, Z^* et les Z_{α} sont des variables aléatoires. Nous allons imposer aux λ^{α} les deux conditions suivantes :

~ quels que soient les a_{ℓ}, $E(Z^* - Z) = 0$
~ quels que soient les a_{ℓ}, $D^2(Z^* - Z)$ est minimale

Au lieu de $E(Z^* Z^*) = 0$, on aurait pu s'imposer la condition beaucoup plus forte : espérance conditionnelle de Z relativement aux $Z_{\alpha} = Z^*$. Mais cette condition plus forte nous ferait sortir du cadre des estimateurs linéaires et nécessiterait l'intervention de la loi spatiale de Z(x) au lieu de sa seule covariance.

La première condition $(E(Z^* - Z) = 0)$ se traduit par :

$$\sum_{\ell} a_{\ell} \int p(dx)\, f^{\ell}(x) = \sum_{\ell} \sum_{\alpha} a_{\ell}\, \lambda^{\alpha} f^{\ell}_{\alpha}$$

Elle est réalisée quels que soient les a_{ℓ} si et seulement si :

(1-10) $$\sum_{\alpha} \lambda^{\alpha} f^{\ell}_{\alpha} = \int p(dx)\, f^{\ell}(x)$$

Cette <u>condition d'universalité</u> (1-10) étant remplie, la variance d'estimation $D^2(Z^*-Z)$ ne dépend plus des a_{ℓ} et vaut :

(1-11) $$D^2(Z-Z^*) = \sigma^2_Z - 2 \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha,Z} + \sum_{\alpha,\beta} \lambda^{\alpha} \lambda^{\beta} \sigma_{\alpha\beta}$$

avec

$$\begin{cases} \sigma^2_Z = \int C(x,y)\, p(dx)\, p(dy) \\ \sigma_{\alpha,Z} = \int C(x_{\alpha},y)\, p(dy) \\ \sigma_{\alpha\beta} = C(x_{\alpha},x_{\beta}) \end{cases}$$

Elle est minimale, compte tenu des conditions (1-10) pour des λ^{α} vérifiant le système :

(1-K) $$\begin{cases} \sum_{\beta} \lambda^{\beta} \sigma_{\alpha\beta} = \sigma_{\alpha,Z} + \sum_{\ell} \mu_{\ell} f^{\ell}_{\alpha} \\ \sum_{\alpha} \lambda^{\alpha} f^{\ell}_{\alpha} = \int p(dx)\, f^{\ell}(x) \end{cases}$$

Telles sont les équations du <u>krigeage universel</u>. Le système (1-K) comporte n+1+k équations à n+1+k inconnues *(les λ^{α} et les paramètres de Lagrange μ)*. La covariance C(x,y) étant de type strictement défini positif, ce système est régulier pourvu seulement que les f (x) soient linéairement indépendantes sur S. <u>La variance</u> σ^2_U de l'estimateur Z^* du krigeage universel est donnée par :

(1-12) $$\begin{cases} \sigma^2_U = \sigma^2_Z - \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha,Z} + \sum_{\ell} \mu_{\ell} b^{\ell} \\ (b^{\ell} = \int p(dx)\, f^{\ell}(x)) \end{cases}$$

<u>Krigeage Ponctuel</u> - Dans le cas où la variable Z à estimer est la valeur Z(x) en un point $x \notin S$, le système (1-K) se réduit à :

(1-13) $$\begin{cases} \sum_{\beta} \lambda^{\beta} \sigma_{\alpha\beta} = C(x_{\alpha},x) + \sum_{\ell} \mu_{\ell} f^{\ell}_{\alpha} \\ \sum_{\alpha} \lambda^{\alpha} f^{\ell}_{\alpha} = f^{\ell}(x) \end{cases}$$

et l'estimateur $Z^*(x)$ du krigeage universel de Z(x) :

$$Z^*(x) = \sum_{\alpha} \lambda^{\alpha} Z_{\alpha}$$

admet la variance

$$\sigma^2_U = C(x,x) - \sum_{\alpha} \lambda^{\alpha} C(x_{\alpha},x) + \sum_{\ell} \mu_{\ell} f^{\ell}(x)$$

<u>Remarque</u> : En un point x_α où $Z_\alpha = Z(x_\alpha)$ est connu, le krigeage ponctuel donne $Z^*(x_\alpha) = Z_\alpha$, comme on le vérifie sans peine : le <u>krigeage ponctuel</u> constitue donc une <u>méthode d'interpolation</u> (et même la meilleure possible) entre des points connus.

1-3 - LE THEOREME D'ADDITIVITE

Examinons les rapports existants entre les estimateurs optimaux de la dérive - système (1-D) et du krigeage universel - système (1-K) - <u>En l'absence de dérive</u> ($a_\ell = 0$); l'estimateur optimal de

$$Z = \int p(dx)\, Z(x)$$

serait de la forme :

$$(1\text{-}14) \qquad Z_K = \sum_\alpha \lambda_K^\alpha Z_\alpha$$

avec des coefficients λ_K^α minimisant la variance

$$(1\text{-}15) \qquad D^2(Z_K - Z) = \sigma_Z^2 - 2\sum_\alpha \lambda_K^\alpha \sigma_{\alpha,Z} + \sum_{\alpha,\beta} \lambda_K^\alpha \lambda_K^\beta \sigma_{\alpha\beta}$$

c'est-à-dire vérifiant le système :

$$(1\text{-}16) \qquad \sum_\beta \lambda_K^\beta \sigma_{\alpha\beta} = \sigma_{\alpha Z}$$

et la variance correspondante serait :

$$\sigma_K^2 = \sigma_Z^2 - \sum_\alpha \lambda_K^\alpha \sigma_{\alpha,Z}$$

Comparons les systèmes (1-K) et (1-16). Posons :

$$Z^* = Z_K + Z_D$$

Z_D représente la <u>correction de dérive</u>. Z_K, solution du krigeage simple *(en l'absence de dérive)* étant donnée par (1-14) avec les coefficients λ_K^α solution de (1-16), on voit que Z_D est de la forme

$$Z_D = \sum_\alpha \lambda_D^\alpha Z_\alpha$$

avec des coefficients λ_D^α vérifiant :

$$(1\text{-}17) \qquad \begin{cases} \sum_\beta \lambda_D^\beta \alpha\beta = \sum_\ell \mu_\ell f_\alpha^\ell \\[2ex] \sum_\alpha \lambda_D^\alpha f_\alpha^\ell = \int p(dx)\, f^\ell(x) - \sum_\alpha \lambda_K^\alpha f_\alpha^\ell \end{cases}$$

Ce système est identique à (1-4), à condition de prendre

$$b^\ell = \int p(dx)\, f^\ell(x) - \sum_\alpha \lambda_K^\alpha f_\alpha^\ell$$

Si donc nous désignons par A_ℓ :

$$A_\ell = \sum_\alpha \lambda_\ell^\alpha Z_\alpha$$

l'estimateur optimal de la dérive a_ℓ, avec λ^α_ℓ vérifiant (1-D), nous voyons que les coefficients λ^α_D du terme correctif sont donnés par :

$$\lambda^\alpha_D = \sum_\ell \lambda^\alpha_\ell \left[\int p(dx)\, f^\ell(x) - \sum_\beta \lambda^\beta_K f^\ell_\beta\right]$$

L'estimateur $Z^* = Z_K + Z_D$ se met donc sous la forme

$$Z^* = \sum_\alpha \lambda^\alpha_K Z_\alpha + \sum_\alpha \lambda^\alpha_D Z_\alpha = \sum_\alpha \lambda^\alpha_K \left(Z_\alpha - \sum_{\ell,\beta} f^\ell_\alpha \lambda^\beta_\ell Z_\beta\right) + \sum_{\alpha,\ell} \lambda^\alpha_\ell Z_\alpha \int p(dx)\, f^\ell(x)$$

Compte tenu de l'expression (1-6) de l'estimateur A_ℓ de la dérive, ceci s'écrit encore :

(1-18) $$Z^* = \sum_\ell A_\ell \int p(dx)\, f^\ell(x) + \sum_\alpha \lambda^\alpha_K \left[Z_\alpha - \sum_\ell A_\ell f^\ell_\alpha\right]$$

Le sens de cette relation est le suivant : on obtient l'estimateur Z^* du krigeage universel en appliquant le krigeage ordinaire *(coefficient λ^α_K)* aux résidus $Z(x) - \sum_\ell A_\ell f^\ell(x)$, c'est-à-dire aux valeurs de la F.A. $Z(x)$ corrigées de la dérive estimée $\sum_\ell A_\ell f^\ell(x)$. Si les valeurs vraies des a_ℓ étaient connues, l'estimateur optimal serait évidemment :

(1-19) $$Z^* = \sum_\ell a_\ell \int p(dx)\, f^\ell(x) + \sum_\alpha \lambda^\alpha_K (Z_\alpha - \sum_\ell a_\ell f^\ell_\alpha)$$

La relation (1-18) est identique à (1-19), à ceci près que les a_ℓ sont remplacées par leurs estimations *(optimales)* A_ℓ. Ce théorème d'additivité n'est évidemment valable que si les A_ℓ sont effectivement l'estimateur optimal de la dérive, solution de (1-K). Il ne s'étendrait pas à des estimateurs quelconques des a_ℓ *(par exemple aux estimateurs de moindre carré)*.

Dans les applications, si l'objectif est d'estimer une "moyenne mobile" $\int p(dx)Z(x)$, il suffit de résoudre directement le système (1-K)(n+1+k équations) sans qu'il soit nécessaire d'expliciter l'estimation de la dérive *(ce qui nécessiterait la résolution du système (1-D) à (k+1)(n+k+1) équations)*. Mais le théorème d'additivité (1-18) montre que le système plus simple (1-K) tient compte, implicitement, de cette dérive.

Dans le cas particulier du krigeage ponctuel *(estimation de $Z(x)$ en $x \notin S$)*, l'estimateur $Z^*(x)$ obtenu en résolvant (1-15) se met sous la forme :

$$Z^*(x) = \sum_\ell A_\ell f^\ell(x) + \sum_\alpha \lambda^\alpha_K (Z_\alpha - \sum_\ell A_\ell f^\ell_\alpha)$$

Cette relation exprime que l'on a le droit de prolonger à l'extérieur de S l'estimation optimale $\sum_\ell A_\ell f^\ell(x)$ de la dérive, et de kriger les résidus comme s'il n'y avait pas de dérive.

Ce théorème d'additivité s'étend aux variances d'estimation.

En effet, avec $Z^* = Z_K + Z_D$, on trouve :

$$D^2[Z-Z^*] = D^2(Z-Z_K) + D^2(Z_D) - 2\,\mathrm{Cov}\,(Z-Z_K,\ Z_D)$$

Mais les équations (1-16) du krigeage ordinaire expriment que $Z-Z_K$ a une covariance nulle avec chacun des Z_α, donc aussi avec Z_D *(qui est une combinaison linéaire des Z_α)*. Il reste donc :

$$D^2[Z-Z^*] = D^2(Z-Z_K) + D^2(Z_D)$$

et on voit que la variance du krigeage universel Z^* est la somme :

(1-20) $$\sigma^2_U = \sigma^2_K + \sigma^2_D$$

de la variance σ^2_K du krigeage simple :

$$\sigma^2_K = \sigma^2_K - \sum_\alpha \lambda^\alpha_K \sigma_{\alpha Z}$$

et de la variance de la correction Z_D de dérive, soit :

$$\sigma^2_D = \sum_{\ell, s} \mu_{\ell s}\, b^\ell b^s$$

avec $\mu_{\ell s}$ = Cov (A_ℓ, A_s) et $b^\ell = \int p(dx)\, f^\ell(x) - \sum_\alpha \lambda^\alpha_K f^\ell_\alpha$

1-4 - CAS OU IL N'EXISTE QU'UN VARIOGRAMME $\gamma(x,y)$

Plaçons-nous maintenant dans le cadre des hypothèses (I') de l'introduction : il existe un demi-variogramme $\gamma(x,y)$, mais non, en général, de covariance $C(x,y)$, ni même d'espérance $E[Z(x)] = m(x)$. Pour des coefficients λ^α quelconques, la combinaison $\sum_\alpha \lambda^\alpha Z_\alpha$ n'a donc pas, en général, de variance finie, ni même d'espérance. Moyennant la condition :

(1-21) $$\sum_\alpha \lambda^\alpha = 0$$

on voit cependant que $\sum_\alpha \lambda^\alpha Z_\alpha$ est une combinaison linéaire d'accroissements de la F.A. $Z(x)$: dans ce cas, $\sum_\alpha \lambda^\alpha Z_\alpha$ admet une espérance et une variance finies.

La dérive

$$m(x) = \sum_{\ell=0}^{k} a_\ell f^\ell(x)$$

n'est ici déterminée qu'à une constante près. Si l'on prend $f^0(x) = 1$ *(conformément à un usage naturel)*, on voit que le coefficient a_0 est indéterminé, les a_ℓ n'étant réellement définis que pour $\ell = 1,2,\ldots k$.

Estimation de la dérive - Pour obtenir les estimateurs optimaux A_ℓ des coefficients a_ℓ ($\ell = 1,2,\ldots k$) on raisonne comme dans la section 1-1, à cette réserve près que toutes les combinaisons linéaires des Z_α doivent vérifier la condition supplémentaire (1-21). L'indice ℓ et les indices latins ne prenant plus cette fois que les valeurs 1,2,... k, on voit facilement que ces estimateurs sont de la forme

(1-22) $$A_\ell = \sum_\alpha \lambda^\alpha_\ell Z_\alpha$$

avec une matrice λ^α_ℓ ($\alpha = 1,2,\ldots n$; $\ell = 1,2,\ldots k$) vérifiant le système :

(1-D') $$\left\{\begin{array}{l} \sum_\beta \lambda^\beta_\ell \gamma_{\alpha\beta} = -\sum_s \mu_{s\ell} f^s_\alpha - c_\ell \\ \sum_\alpha \lambda^\alpha_\ell f^s_\alpha = \delta^s_\ell \\ \sum_\alpha \lambda^\alpha = 0 \end{array}\right. \qquad \left(\begin{array}{l} \alpha,\beta = 1,2,\ldots n \\ \ell, s = 1,2,\ldots k \\ \gamma_{\alpha\beta} = \gamma(x_\alpha, x_\beta) \end{array}\right)$$

Compte tenu du fait que ℓ ne prend plus la valeur 0, ce système ne diffère pas réellement de (1-D) (avec $\sigma_{\alpha\beta} = \gamma_{\alpha\beta}$ et $C_\ell = \mu_{0\ell}$), à condition d'exclure de celui-ci les équations correspondant à $\ell = 0$.

Les paramètres de Lagrange $\mu_{\ell s}$ coïncident encore avec la matrice des covariances :

$$\mu_{\ell s} = \mathrm{Cov}\,(A_\ell, A_s) = E(A_\ell A_s) - a_\ell a_s$$

et l'on a, bien entendu, $E(A_\ell) = a_\ell$ quels que soient les a_ℓ.

<u>Remarque</u> : *On peut même compléter le système (1-D') en réintroduisant les indices $\ell = 0$. On obtient (1-D), avec $\sigma_{\alpha\beta} = -\gamma_{\alpha\beta}$. Mais l'estimation*

$$A_o = \sum_\alpha \lambda_o^\alpha Z_\alpha$$

de la constante a_o (qui est, en réalité, indéterminée) ne vérifie pas la condition (1-21), puisque l'on a $\sum \lambda_o^\alpha = 1$: elle conduit donc, en général, à une variance infinie, ce qui est bien conforme à l'hypothèse faite (indétermination de a_o, ou inexistence de $E(Z(x))$).

<u>Krigeage Universel</u> - De la même manière, pour estimer une moyenne pondérée $\int p(dx)\, Z(x)$, avec $\int p(dx) = 1$, on formera une expression de la forme

$$Z^* = \sum_\alpha \lambda^\alpha Z_\alpha \tag{1-23}$$

avec cette fois $\sum \lambda^\alpha = 1$, de manière à ce que l'espérance $E(Z-z^*)$ existe toujours. En imposant à cette espérance d'être nulle quels que soient les a_ℓ, et en écrivant que la variance de $Z-Z^*$ est minimale, on obtient le système

$$\text{(1-K')}\quad \begin{cases} \sum_\beta \lambda^\beta \gamma_{\alpha\beta} = \int \gamma(x_\alpha, y)\, p(dy) - \sum_{\ell=1}^{k} \mu_\ell\, f^\ell_\alpha - C \\ \sum_\alpha \lambda^\alpha f^\ell_\alpha = \int p(dx)\, f^\ell(x) \\ \sum_\alpha \lambda^\alpha = 1 \end{cases}$$

La variance de $Z^* - Z$ est alors :

$$\sigma_U^2 = - \iint p(dx)\, p(dy)\, \gamma(x,y) + \sum_\alpha \lambda^\alpha \int \gamma(x_\alpha, y)\, p(dy) + \sum_{\ell=1}^{k} \mu_\ell\, f^\ell_\alpha + C \tag{1-24}$$

Dans le cas du <u>krigeage ponctuel</u>, l'estimateur optimal $Z^*(x)$ de $Z(x)$, $x \notin S$ est de la forme (1-23), avec des λ^α vérifiant :

$$\text{(1-25)}\quad \begin{cases} \sum_\beta \lambda^\beta \gamma_{\alpha\beta} = \gamma(x_\alpha, x) - \sum_{\ell=1}^{k} \mu_\ell f^\ell_\alpha - C \\ \sum_\alpha \lambda^\alpha f^\ell_\alpha = f^\ell(x) \\ \sum_\alpha \lambda^\alpha = 1 \end{cases}$$

et la variance correspondante est *(compte tenu de $\gamma(x,x) = 0$)*

$$\sigma_U^2 = \sum_\alpha \lambda^\alpha \gamma(x_\alpha, x) + \sum_{\ell=1}^{k} \mu_\ell f^\ell{}_\alpha + C \tag{1-26}$$

Le théorème d'additivité subsiste sans changement : si l'on désigne par

$$Z_K = \sum_\alpha \lambda_K^\alpha Z_\alpha$$

la solution du krigeage simple *(en l'absence de dérive)*, les λ_K^α vérifient le système :

$$\sum_\beta \lambda_K^\beta \gamma_{\alpha\beta} = \int \gamma(x_\alpha, x)\, p(dx) - C_K$$

$$\sum_\alpha \lambda_K^\alpha = \int p(dx)$$

On met Z^* sous la forme $Z^* = Z_K + Z_D$, et la correction de dérive Z_D est encore de la forme

$$Z_D = \sum_{\ell=1}^{k} A_\ell \int f^\ell(x)\, p(dx) - \sum_\alpha \lambda_K^\alpha \sum_{\ell=1}^{k} A_\ell f^\ell_\alpha$$

et on a encore :

$$D^2(Z-Z^*) = D^2(Z-Z_K) + D^2(Z_D)$$

On trouvera dans la section 2-10 un exemple de calcul explicite *(dérive et krigeage universel)* dans le cas d'un variogramme linéaire $\gamma(x,y) = \varpi|x-y|$ et de l'espace à une seule dimension.

BIBLIOGRAPHIE

(1) JOURNEL, A. (1969) - Rapport d'Etude sur l'estimation d'une variable régionalisée : application à la cartographie automatique - (Service Hydrographique de la Marine, Paris).

(2) LANDKOFF, N.S. (1966) - Osnovy sovremennoï teroii potentziala, Moscou, Izd. Nauka, 516 pages.

(3) MATHERON, G. (1965) - Les variables régionalisées et leur estimation, Paris, Masson, 306 pages.

(4) MATHERON, G. (1967) - Kriging, or polynomial interpolation procedures ?. Can. Min. and Met. Bull., LXX, 240-244.

(5) MATHERON, G. (1968) - Osnovy prikladnoï Geostatistiki, Moscou, Editions Mir, 408 pages.

(6) PLESNER, A.I. (1965) - Spektralnaïa teoria lineinikh operatorov. Moscou. Izd. Nauka, 624 pages.

(7) SERRA, J. (1968) - Les structures gigogne : morphologie mathématique et interprétation métallogénique - Mineralium Deposita, 3, 135-154.

(8) YOSIDA, K. (1968) - Functional Analysis, Springer, Berlin, 466 pages.

(9) MATHERON, G. (1969) - Le krigeage universel : Les Cahiers du Centre de Morphologie Mathématique - Fontainebleau.