

N-67

April 1967

NOTE GLOSTATISTIQUE N° 71Le Krigeage en représentations transitives.

Nous nous proposons, dans cette Note, d'adapter le procédé du krigeage aux représentations transitives. Les équations auxquelles nous parviendrons sont pratiquement équivalentes à celles du krigeage intrinsèque, la fonction intrinsèque $\gamma(h)$ cédant ici la place au covariogramme transitif $g(h)$. Ce résultat n'est pas dépourvu d'une certaine importance méthodologique. On pouvait craindre, en effet, que les méthodes usuelles de krigeage ne soient inapplicables au cas des fonctions aléatoires non stationnaires. En fait, le $\gamma(h)$ que l'on obtient, en appliquant les méthodes de calcul habituelles dans le cas d'une variable régionalisée non-stationnaire, possède, en réalité, la signification d'un covariogramme transitif. Il est donc équivalent d'utiliser les méthodes de krigeage intrinsèque avec un tel $\gamma(h)$, ou celles du krigeage transitif avec le $g(h)$ correspondant. Cela revient à dire que les méthodes de krigeage telles que nous les utilisons en pratique, avec les $\gamma(h)$ construits selon les procédés habituels, conservent toute leur valeur, même si la variable régionalisée n'est pas stationnaire.

Nous examinerons successivement le krigeage discontinu, puis le krigeage continu et ce que l'on pourrait appeler le cokrigeage, c'est à dire l'estimation d'une variable $f_1(x)$ à partir des valeurs connues prises par une autre variable $f_2(x)$ dans un domaine donné. Nous indiquerons, pour terminer une application possible au problème dit de la déconvolution des sondages radiocarottés, c'est à dire de la reconstitution des teneurs en Uranium à partir des radioactivités que l'on observe dans les sondages.

I- Krigeage discontinu.

Soit $f(x)$ une variable régionalisée, et $g = f \star f$ son covariogramme transitif, défini par :

$$g(h) = \int f(x) f(x+h) dx$$

On suppose connues les valeurs $f(y+x_1), \dots, f(y+x_n)$ prises par $f(x)$ en n points de prélèvements $y+x_1, \dots, y+x_n$, et on cherche à estimer la valeur $f(y)$ de f en point y , ou, plus généralement, une moyenne pondérée de la forme :

$$(1) \quad \hat{Q}(y) = \int f(x) \varphi(x-y) dx = \int f(x+y) \varphi(x) dx$$

φ étant une fonction de prélèvement donnée (il n'est pas nécessaire de la supposer de somme unité). On va prendre un estimateur linéaire de la forme :

$$\hat{Q}^*(y) = \sum_{i=1}^n \lambda_i f(y+x_i)$$

et déterminer les coefficients λ_i par la condition de rendre minimale l'intégrale :

$$(2) \quad I = \int [\varphi(y) - \hat{Q}^*(y)]^2 dy$$

Cette intégrale I jouera le rôle d'une variance d'estimation. On voit que ce procédé consiste à rendre minimale l'erreur moyenne, le terme "moyenne" se rapportant ici, non pas à une espérance mathématique, mais à l'ensemble des positions que le système de points $(y, y+x_1, \dots, y+x_n)$ occupa dans l'espace lorsqu'on le déplace dans son ensemble sans le déformer (translation y). Explicitons I :

$$I = \iiint f(x+y) \varphi(x) f(x'+y) \varphi(x') dx dx' dy \\ - 2 \sum_i \lambda_i \iint f(y+x_i) f(y+x) \varphi(x) dx dy \\ + \sum_{i,j} \lambda_i \lambda_j \int f(y+x_i) f(y+x_j) dy$$

Il vient, donc :

$$(3) \quad I = \langle g, \varphi * \check{\varphi} \rangle - 2 \sum_i \lambda_i \int \varphi(x) g(x-x_i) dx + \sum_{i,j} \lambda_i \lambda_j g(x_i-x_j)$$

Cette expression est minimale lorsque les λ_i vérifient le système :

$$(4) \quad \sum_j \lambda_j g(x_i-x_j) = \int \varphi(x) g(x-x_i) dx$$

et la valeur minimale correspondante de I est :

$$(5) \quad I_0 = \langle g, \varphi * \check{\varphi} \rangle - \sum_i \lambda_i \int \varphi(x) g(x-x_i) dx$$

Le système (4) ne diffère des équations habituelles du krigeage intrinsèque que par l'absence de la condition de normalisation et du multiplicateur de Lagrange qu'elle fait apparaître. Du point de vue pratique, cette différence ne joue pas un grand rôle, et le système (4) peut être regardé comme équivalent au système que l'on obtient dans le cadre de la théorie intrinsèque.

Remarque L'identification est encore meilleure si, au lieu d'un schéma intrinsèque, nous prenons en considération une fonction aléatoire $f(x)$ d'espérance nulle :

$$E[f(x)] = 0$$

de variance a priori finie $K(0)$ et de covariance :

$$K(h) = E[f(x) f(x+h)]$$

On doit alors minimiser la variance de $Q(y) - \hat{Q}^*(y)$, qui est :

$$E \left[\left(\varphi(y) - \hat{\varphi}^*(y) \right)^2 \right] = \iint k(x-x') \varphi(x) \varphi(x') dx dx' \\ - 2 \sum_i \lambda_i \int k(x-x_i) \varphi(x) dx + \sum_{i,j} \lambda_i \lambda_j K(x_i-x_j)$$

Ce minimum est atteint avec des coefficients λ_i vérifiant :

$$(4)' \quad \sum_j \lambda_j k(x_i-x_j) = \int k(x-x_i) \varphi(x) dx$$

et la variance minimale correspondante est :

$$(5)' \quad D^2 \left[\varphi(y) - \hat{\varphi}^*(y) \right] = \langle k, \varphi * \check{\varphi} \rangle - \sum_i \lambda_i \int k(x-x_i) \varphi(x) dx$$

Il suffit de remplacer la covariance $K(h)$ par le covariogramme transitif $g(h)$ pour retrouver les équations (4) et (5). On voit que, dans cette analogie, on passe du transitif à l'aleatoire stationnaire d'ordre deux en diluant dans l'espace entier la variable régionalisée $f(x)$ - ce qui implique la condition $E[f(x)] = 0$ - et en remplaçant l'intégration dans l'espace par un passage à l'espérance mathématique.

II-Krigeage continu.

Nous nous proposons toujours d'estimer l'intégrale $Q(y)$ définie en (I), mais cette fois nous connaissons les valeurs prises par $f(x)$ sur un ensemble infini de points, que nous désignerons par V_y (translaté d'un ensemble V donné dans une translation y : l'idée est ici encore d'imprimer une translation y à l'ensemble constitué par le domaine où les valeurs de f sont connues et la zone que l'on cherche à estimer). On forme alors un estimateur du type :

$$\hat{Q}^*(y) = \int_V \lambda(x) f(y+x) dx = \int_{V_y} \lambda(x-y) f(x) dx$$

où $\lambda(x)$ représente une fonction nulle en dehors de V qu'il s'agit

de déterminer. Considerons donc le carré de la différence $Q(y) - Q^*(y)$, et intégrons en y , ce qui donne :

$$\begin{aligned} I &= \int [\varphi(y) - \varphi^*(y)]^2 dy \\ &= \iiint f(x+y) \varphi(x) f(x'+y) \varphi(x') dx dy dx' \\ &\quad - 2 \iiint f(x+y) \varphi(x) f(x'+y) \lambda(x') dx dy dx' \\ &\quad + \iiint f(x+y) \lambda(x) f(x'+y) \lambda(x') dx dx' dy \\ &= \langle g, \varphi * \check{\varphi} \rangle - 2 \iint g(x-x') \varphi(x) \lambda(x') dx dx' \\ &\quad + \iint g(x-x') \lambda(x) \lambda(x') dx dx' \end{aligned}$$

Pour exprimer que cette intégrale est minimale, nous donnons à $\lambda(x)$ une variation $\delta\lambda(x)$ nulle en dehors de V , et nous exprimons que la variation de I est nulle. On obtient ainsi une équation intégrale qui est une simple transposition de (4) :

$$(4)'', \quad \int_V \lambda(x) g(x-x') dx = \int \varphi(x) g(x-x') dx \quad (\forall x' \in V)$$

et qui doit être vérifiée pour tout point x' appartenant à V . La valeur minimale de l'intégrale I , qui représente l'erreur associée à ce krigeage, est alors :

$$(5)'', \quad I_0 = \langle g, \varphi * \check{\varphi} \rangle - \iint g(x-x') \varphi(x) \lambda(x') dx dx'$$

III-Cokrigeage.

Nous nous donnons maintenant une coregionalisation à deux composantes $f_1(x)$ et $f_2(x)$, et son covariogramme transitif matricielle :

$$(6) \quad g_{ij}(h) = \int f_i(x) f_j(x+h) dx \quad (i, j = 1, 2)$$

Nous nous proposons d'estimer une moyenne pondérée $Q_1(y)$ de la forme :

$$Q_1(y) = \int f_1(x) \varphi(x-y) dx = \int f_1(x+y) \varphi(x) dx$$

à partir des valeurs supposées connues prises dans un domaine V_y par la deuxième variable $f_2(y)$. Comme ci-dessus, on va chercher à former un estimateur du type :

$$(7) \quad Q_1^*(y) = \int_V \lambda(x) f_2(x+y) dx = \int_{V_y} \lambda(x-y) f_2(x) dx$$

où $\lambda(x)$ est une fonction identiquement nulle à l'extérieur de V , et qu'il va s'agir de déterminer. Pour cela, on introduit l'intégrale :

$$\begin{aligned} I &= \int [Q_1(y) - Q_1^*(y)]^2 dy \\ &= \langle g_{11}, \varphi * \check{\varphi} \rangle - 2 \iint g_{12}(x'-x) \varphi(x) \lambda(x') dx dx' \\ &\quad + \iint g_{22}(x'-x) \lambda(x) \lambda(x') dx dx' \end{aligned}$$

On exprime ensuite que cette intégrale est minimale, en écrivant que la variation de I est nulle pour toute variation $\delta\lambda(x)$ identiquement nulle à l'extérieur de V . On obtient ainsi le système (toujours le même) :

$$(4)'''' \quad \int_V g_{22}(x'-x) \lambda(x) dx = \int g_{12}(x'-x) \varphi(x) dx \quad (\forall x' \in V)$$

qui doit être vérifié pour tout x' appartenant à V . La valeur minimale correspondante de l'intégrale I est :

$$(5)'''' \quad I_0 = \langle g_{11}, \varphi * \check{\varphi} \rangle - \iint g_{12}(x'-x) \varphi(x) \lambda(x') dx dx'$$

Cas particuliers. I/ On cherche à estimer la valeur prise par f_1 en un point y , soit $Q_1(y) = f_1(y)$. La fonction φ est remplacée par la mesure de Dirac, et on obtient :

$$\left\{ \begin{array}{l} \int_V g_{22}(x'-x) \lambda(x) dx = g_{12}(x') \quad (\forall x' \in V) \\ I_0 = g_{11}(0) - \int g_{12}(x) \lambda(x) dx \end{array} \right.$$

2-Le domaine V est identique à l'espace entier, autrement dit les valeurs de la deuxième variable f_2 sont connues partout. On obtient alors des équations de convolution :

$$\left\{ \begin{array}{l} g_{22} * \lambda = g_{12} * \varphi \\ I_0 = \langle g_{11}, \varphi * \varphi \rangle - \langle g_{12}, \lambda * \varphi \rangle \end{array} \right.$$

IV-Application à la déconvolution des sondages radiocarottes

I) / Methode de Coulomb-Grammakov.

Dans un gisement d'Uranium, désignons par :

$f_0(x)$ la teneur en Uranium au point x

$f_1(x)$ la teneur en radium au point x

$f_2(x)$ la radioactivité que l'on pourrait mesurer au point x

Les deux variables $f_0(x)$ et $f_1(x)$, considérées comme des variables aléatoires, sont supposées lognormales et fortement corrélées; le coefficient de corrélation, cependant, n'est pas égal à l'unité, de sorte qu'il n'y a pas de relation fonctionnelle entre f_0 et f_1 . La courbe de régression donnant la valeur probable de la teneur en U en fonction de la teneur en Ra est une droite en coordonnées logarithmiques, mais en général ce n'est pas une droite en coordonnées arithmétiques.

Au contraire, f_1 et f_2 sont reliées par un opérateur linéaire. L'émission radioactive du point x est, en effet, proportionnelle à la teneur $f_1(x)$ de ce point en radium, et s'amortit, en fonction de la distance r, selon une loi en $\frac{1}{r^2} e^{-\mu r}$. Désignons par $\alpha(x)$ cette

fonction d'amortissement, avec $r = |x|$, soit :

$$(8) \quad \alpha(x) = \frac{1}{r^2} e^{-\mu x}$$

La radioactivite $f_2(y)$ que l'on pourrait observer en un point y est alors une moyenne, ponderee par la fonction d'amortissement $\alpha(x-y)$, des teneurs $f_1(x)$ en radium de tous les points x du gisement, soit :

$$(9) \quad f_2(y) = \int \alpha(x-y) f_1(x) dx$$

(En fait, cette fonction $\alpha(x)$ ne serait vraiment de la forme (8) que s'il etait possible de faire des observations ponctuelles. Du fait que les sondes utilisees possedent leurs caracteristiques geometriques propres, la veritable fonction de ponderation α que l'on doit utiliser dans l'equation (9) est quelque peu plus compliquee que (8). Mais nous n'aurons pas à tenir compte de cette difficulte.)

Si donc l'on connaissait les radioactivites $f_2(y)$ en tous les points du gisement, il suffirait de resoudre l'equation integrale (9) c'est à dire de deconvoluer f_2 pour reconstituer la teneur $f_1(x)$ en radium en tous les points x du gisement. Il suffirait ensuite d'utiliser la regression de f_0 en f_1 pour obtenir les valeurs probables de la teneur en U en chaque point de l'espace.

Mais naturellement, on ne connait pas $f_2(y)$ en tous les points du gisement. On connait seulement les profils des radioactivites mesurees le long d'un nombre fini n de sondages verticaux. Le problème que l'on se pose alors consiste à reconstituer le profil des teneurs en radium à partir du profil des radioactivites.

Ce problème n'est pas soluble, en general (il est indetermine)

à moins que nous n'introduisions des hypothèses supplémentaires : l'hypothèse que l'on fait dans la méthode de Grammakov, et dans la méthode qui en dérive mise au point par Coulomb au C.E.A., consiste à supposer que les teneurs en radium restent constantes dans les plans horizontaux (au moins au voisinage de chaque sondage).

A partir de maintenant, nous utiliserons des notations explicites, en designant par (x, y, z) les coordonnées d'un point de l'espace. L'hypothèse de Grammakov et de Coulomb consiste donc à admettre que la teneur en radium est une fonction $f_1(z)$ de la seule coordonnée z . La radioactivité devient alors une fonction $f_2(z)$ constante également dans les plans horizontaux (au moins au voisinage de chaque sondage). L'équation (9) qui représentait une convolution dans l'espace à 3 dimensions, cède alors la place à l'équation plus simple :

$$(II) \quad f_2(z) = \int_{-\infty}^{+\infty} \omega(\zeta - z) f_1(\zeta) d\zeta$$

qui représente une convolution dans l'espace à une seule dimension. La fonction de pondération à une seule variable $\omega(z)$ se déduit très simplement de la fonction d'amortissement initiale $\alpha(x, y, z)$ à trois variables, selon la formule :

$$(I2) \quad \omega(z) = \iint \alpha(x, y, z) dx dy$$

L'opération (I2) permettant de passer de α à ω est une montée d'ordre deux. Si α est de la forme (8), on trouve, en passant en coordonnées polaires :

$$\omega(z) = 2\pi \int_0^{\infty} \frac{e^{-\mu \sqrt{\rho^2 + z^2}}}{\rho^2 + z^2} \rho d\rho$$

ρ designant le rayon vecteur $\sqrt{x^2 + y^2}$ du plan des (x, y) , soit encore en prenant $r = \sqrt{x^2 + y^2 + z^2}$:

$$(I3) \quad \omega(z) = 2\pi \int_z^{\infty} e^{-\mu r} \frac{dr}{r}$$

que selon la direction verticale. Dans ce cas, le problème n'admet plus de solution rigoureuse, mais seulement une solution optimale au sens des moindres carrés.

4-De toutes manières, on s'intéresse aux teneurs en U et non en Ra. Au lieu de reconstituer d'abord f_1 à partir de f_2 par une méthode \bar{K} rigoureuse, puis f_0 à partir de f_1 par une méthode de régression, n'est-il pas possible de reconstituer directement f_0 à partir de f_2 par une méthode de moindres carrés?

Dans le cas d'une régression linéaire de f_0 en f_1 , cette méthode directe serait parfaitement légitime, et n'entraînerait la perte d'aucune information. Mais la régression lognormale n'est pas linéaire en général. La méthode que nous proposons, qui est un cokrigage de f_0 par f_2 , entraînera donc une certaine détérioration de notre information, puisque nous remplaçons implicitement la régression curviligne, qui est la meilleure possible, par une simple régression linéaire. Cette substitution est cependant nécessaire, si nous voulons rester dans le cadre de l'algèbre linéaire, et l'on peut penser, sous réserve de vérification, qu'elle n'aura pas de conséquences trop graves. (Dans le même ordre d'idée, on sait que D.G. Krige a été conduit, pour le gisement d'or du Rand, à substituer un simple krigage linéaire à son estimateur lognormal, théoriquement le meilleur possible, mais lourd à mettre en œuvre en pratique. Il considère comme très faible la perte de précision qui en résulte.)

2°/ Méthode du cokrigage transitif.

La méthode que nous proposons va donc consister à former directement un estimateur linéaire $f_0^*(z)$ de la teneur $f_0(z)$ au point de cote

z d'un sondage en fonction des radioactivites $f_2(z)$ mesurees le long de ce sondage :

$$(15) \quad f_0^*(z) = \int_{-\infty}^{+\infty} \lambda(z-\beta) f_2(\beta) d\beta$$

Pour determiner la fonction $\lambda(z)$, nous ecrivons que l'expression:

$$(16) \quad I = \sum \int_{-\infty}^{+\infty} [f_0(\beta) - f_0^*(\beta)]^2 d\beta$$

est minimale. Le signe \sum signifie que l'on fait la somme de toutes les integrales correspondant à chacun des sondages etalonnees (pour lesquels on connait à la fois $f_0(z)$ et $f_2(z)$.)

Explicitons l'expression de I, en tenant compte de (15). I, vient :

$$I = \sum \left\{ \int [f_0(\beta)]^2 d\beta - 2 \iint f_0(\beta) \lambda(\beta-\beta') f_2(\beta) d\beta d\beta' + \iiint \lambda(\beta-\beta') \lambda(\beta'-\beta'') f_2(\beta) f_2(\beta'') d\beta d\beta' d\beta'' \right\}$$

Introduisons donc la matrice des covariogrammes, qui se presente sous la forme habituelle, à ceci près que les expressions obtenues doivent etre sommees sur les differents sondages qui servent d'etalon.

$$\left\{ \begin{aligned} g_{00}(h) &= \sum \int_{-\infty}^{+\infty} f_0(\beta) f_0(\beta+h) d\beta \\ g_{02}(h) &= \sum \int_{-\infty}^{+\infty} f_0(\beta) f_2(\beta+h) d\beta \\ g_{22}(h) &= \sum \int_{-\infty}^{+\infty} f_2(\beta) f_2(\beta+h) d\beta \end{aligned} \right.$$

Ces fonctions peuvent etre construites à partir des donnees experimentales. L'integrale I à minimiser s'ecrit :

$$I = g_{00}(0) - 2 \int g_{02}(h) \lambda(h) dh + \iint g_{22}(u-v) \lambda(u) \lambda(v) du dv$$

La fonction λ et la valeur minimale I_0 de I sont alors donnees par le système habituel :

$$(I7) \quad \left\{ \begin{array}{l} \int_{-\sigma}^{+\sigma} g_{22}(u-\sigma) \lambda(u) du = g_{02}(\sigma) \\ I_0 = g_{00}(0) - \int_{-\sigma}^{+\sigma} g_{02}(z) \lambda(z) dz \end{array} \right.$$

Il s'agit d'une equation de convolution que l'on pourra resoudre, numeriquement, à l'aide des procedes habituels de discretisation.

L'erreur commise sur l'estimation d'une teneur individuelle peut et re evaluee à partir de I_0 . Si, en effet $L = \sum L_i$ represente la longueur utile cumulee de tous les sondages etalons, la variance-valeur moyenne de l'erreur quadratique- a pour valeur :

$$(I8) \quad \sigma^2 = \frac{I_0}{L}$$

Cette variance (I8) represente l'erreur commise dans l'estimation d'une teneur individuelle, et on ne peut rien en deduire en ce qui concerne l'erreur globale commise dans l'estimation du gisement dans son ensemble. L'evaluation de cette dernière erreur necessitera la mise au point d'une theorie speciale .

3°/ Cokrigeage reciproque.

Il sera interessant de comparer directement les resultats fournis par l'estimateur (I6) avec ceux que l'on obtient par la methode de Coulomb-Grammakov. Mais cette comparaison peut aussi etre faite en sens inverse.

En effet, si l'on suppose qu'une relation fonctionnelle telle que (II) existe reellement, on doit pouvoir en determiner un reflet en cherchant le meilleur estimateur donnant la radioactivite $f_2(z)$ en fonction des teneurs en Uranium $f_1(z)$ (ou meme, si celles ci sont connues, en fonction des teneurs en radium¹): la comparaison sera alors

beaucoup plus directe.) Cet estimateur sera de la forme :

$$f_2(z) = \int_{-\infty}^{+\infty} \omega'(z-\xi) f_0(\xi) d\xi$$

(on mettra f_1 au lieu de f_0 si l'on introduit les teneurs en radium.)

La fonction ω' et l'expression minimale I_0 sont données par le système habituel :

$$(I8) \quad \int_{-\infty}^{+\infty} g_{00}(u-v) \omega'(u) du = g_{20}(v) = g_{02}(-v)$$

$$I_0 = g_{22}(0) - \int_{-\infty}^{+\infty} g_{20}(v) \lambda(v) dv$$

Il serait intéressant, en particulier, de comparer la fonction de pondération ω' ainsi obtenue empiriquement avec la fonction théorique ω utilisée dans la méthode de Grammakov et de Coulomb, et notamment d'examiner si la décroissance à l'infini est bien celle que suggère la fonction integro-exponentielle (I3).