Traitement des données à support spatial : la géostatistique et ses usages

par

Pierre CHAUVET

N–28/92/G

Octobre 1992

CENTRE DE GÉOSTATISTIQUE

35, RUE SAINT–HONORÉ, 77305 FONTAINEBLEAU (France)

ECOLE DES MINES
DE PARIS

CENTRE DE GÉOSTATISTIQUE,

Ecole Nationale Supérieure des Mines de Paris,

35 rue Saint–Honoré, 77305 FONTAINEBLEAU, France

Pierre CHAUVET. Traitement des données à support spatial : la géostatistique et ses usages. Octobre 1992. 43 pages + planches noir et blanc.

N–28/92/G

Document élaboré à l'initiative et avec le soutien d'EDF.

© ENSMP 1992

Traitement des données à support spatial : la géostatistique et ses usages

par

Pierre CHAUVET

N–28/92/G

Octobre 1992

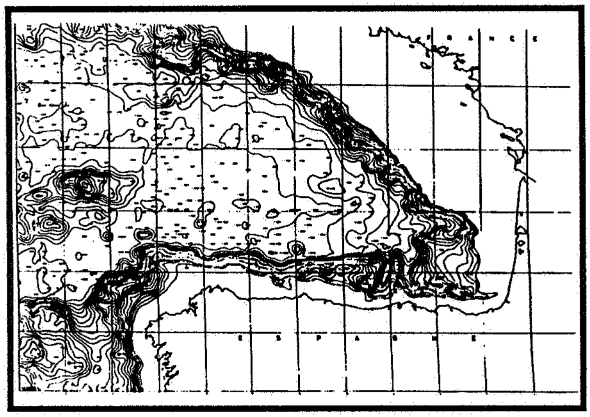
CENTRE DE GÉOSTATISTIQUE
35, RUE SAINT–HONORÉ, 77305 FONTAINEBLEAU (France)

ECOLE DES MINES
DE PARIS

Traitement des données à support spatial : la Géostatistique et ses usages.

Traitement des données à support spatial : la Géostatistique et ses usages

Introduction

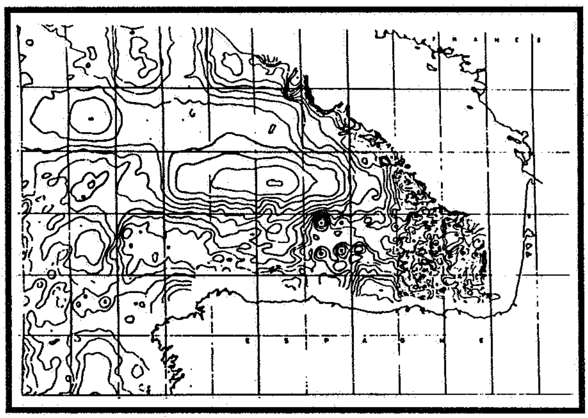


Bathymétrie du Golfe de Gascogne

La carte ci-contre répond à un problème simple : à partir d'un jeu de données de profondeurs du fond marin dans le Golfe de Gascogne, on a voulu en représenter les courbes de niveau. Notons que les données étaient en l'occurrence disposées de façon irrégulière. Pour cela, on est passé très classiquement par l'intermédiaire d'une **estimation** aux nœuds d'une grille très fine, et le tracé des courbes a ensuite été obtenu par un lissage de type spline.

Cet exemple, bien que simpliste, contient en germe bien des interrogations.

Par le truchement de la carte, on souhaite d'abord rendre visible un jeu de **propriétés structurales** du phénomène étudié. La variable "profondeur", en effet, possède des caractéristiques dont une bonne carte doit rendre compte : anisotropie (comportements différents selon les directions : la profondeur ne varie pas de la même façon selon que l'on est parallèle ou perpendiculaire à la côte), hétérogénéités (la structure du fond n'est pas la même au large des côtes espagnole et landaise), tendances (pentes), etc... Il convient donc dans un premier temps de mettre en évidence ces caractéristiques, puis de les décrire mathématiquement afin de pouvoir mettre en œuvre un estimateur qui les prenne en compte. Ce double travail d'**interrogation** des données, puis de **modélisation** de leurs propriétés structurales, constitue sous le nom d'**Analyse Variographique** (*) la phase initiale incontournable de toute étude concrète de Géostatistique.



Ecarts-types de la carte précédente

Mais on peut ne pas vouloir s'en arrêter là. On peut par exemple demander une évaluation de la "**qualité**" de cette première carte. Il faut d'ailleurs commencer par définir cette notion. Car si une profondeur est une grandeur physique clairement définie qui s'impose à chacun, la qualité d'un estimateur est une notion abstraite qui procède d'un **choix** arbitraire — ce dernier mot n'ayant notons-le aucune connotation péjorative.

Il faut donc **convenir** a *priori* **d'un critère** de qualité. Un critère possible est par exemple l'écart-type d'estimation, qui est représenté sur cette seconde carte. Schématiquement, une valeur estimée est d'autant plus fiable que son écart-type d'estimation est faible.

(*) Nous préférons cette expression à "Analyse Structurale", plus souvent employée, mais qui a un autre sens en Géologie.

On peut montrer que cet écart-type dépend de la structure propre à la variable étudiée, mais aussi de la configuration de l'**échantillonnage** — ce qui est intuitivement satisfaisant. Ce critère est particulièrement simple d'utilisation, et constitue d'ailleurs la pierre angulaire de la Géostatistique. On peut aussi l'utiliser pour aller plus loin, et s'interroger par exemple sur la stratégie des campagnes de reconnaissance, sur l'implantation optimale d'un complément d'information, ou encore sur l'impact d'une incertitude de localisation des sondes.

Ces questions demeurent encore élémentaires. Mais on peut se poser des problèmes notablement plus complexes : définir par exemple des probabilités de hauts-fonds ou des **intervalles de confiance** sur la carte des bathymétries, ou bien encore s'interroger sur la longueur d'un câble sous-marin qui reposerait sur le fond entre deux points désignés... Dans un autre ordre d'idées, on peut vouloir relier les paramètres issus de l'Analyse Variographique à des phénomènes géologiques, et donc envisager une **analyse interprétative** des données. On peut aussi s'aider de l'Analyse Variographique pour comparer des méthodes d'échantillonnage : dans l'exemple proposé, les données provenaient de deux campagnes à la mer, effectuées avec deux navires et deux appareillages différents dont il fallait déterminer le mieux adapté.

Toutes ces questions, et bien d'autres encore que l'on peut se poser sur tout phénomène présentant une structure dans l'espace (ou le temps), relèvent fondamentalement de la Géostatistique. La contribution de la Géostatistique au traitement des données à support spatial est toujours double :

- D'une part, elle fournit des **outils d'interrogation** des données, orientés vers la mise en évidence de la structure intrinsèque des variables étudiées et de leurs liens avec leur domaine de définition ;
- D'autre part, elle définit un **cadre théorique** permettant de développer des algorithmes (estimations, simulations numériques, optimisations) répondant aux problèmes que l'on peut rencontrer dans l'étude des processus spatiaux.

Par ailleurs, l'exemple introductif — simpliste — que nous venons de proposer aurait pu être emprunté à bien d'autres domaines : c'est l'une des richesses de la Géostatistique que de ne pas être tributaire d'un champ d'application particulier. C'est pourquoi il ne faut pas, dans la littérature géostatistique et dans la présentation qui va suivre, attacher une importance excessive à des choix de vocabulaire qui semblent privilégier certaines applications. La Géostatistique est, essentiellement, un ensemble de méthodes, et le nom qu'on leur donne est de moindre importance : faut-il dire "estimation" comme dans l'industrie minière ? ou "prédiction" comme pour les processus temporels ? ou "interpolation" comme en cartographie ? ou "assimilation" comme en météorologie ?... Faut-il dire "bruit blanc" plutôt que "effet de pépite" ?... Voilà un exemple de question qui ne sera pas examiné ici...

Historique

Les débuts de la Géostatistique sont d'inspiration exclusivement minière. C'est pour pallier les insuffisances des Statistiques "classiques" constatées dans l'étude des gisements très disséminés, que sont élaborées au début des années 50 des méthodes d'estimation nouvelles. Le néologisme "krigeage", en hommage à D.G. Krige et ses travaux sur les minerais d'or d'Afrique du Sud, reste pour rappeler cette rencontre entre une technique mathématique de régression et les problèmes très concrets d'exploitation des mines d'or. Mais déjà, les applications s'étendent à d'autres minéraux : fer, cuivre, nickel, uranium...

Deux traits caractérisent ce **premier âge de la Géostatistique**. Au niveau pratique d'abord, les moyens de calcul demeurent fort rudimentaires. Aussi les publications abondent-elles en formules d'approximation, courbes, abaques, qui finissent par constituer un véritable capital. Au niveau théorique, on remarque que les formalismes qui s'élaborent se placent souvent dans le cadre d'une loi de distribution donnée. Il s'agit non pas tant du modèle gaussien — inadapté aux variables disséminées — que du modèle log-normal, pour lequel se manifeste un engouement extraordinaire dans les années 1950. Ainsi, pour résumer, le néologisme "Géostatistique" convient parfaitement pour décrire cette première époque.

Mais, avec le **second âge de la Géostatistique**, la référence à des modèles statistiques est abandonnée. Ou bien on élabore des modèles qui ne font pas intervenir les lois de distribution (Géostatistique Linéaire), ou bien on se ramène à des modèles de référence par le biais des anamorphoses. Parallèlement, on cherche à élargir les hypothèses de travail : c'est le développement d'une Géostatistique Non Stationnaire pour traiter des phénomènes présentant une tendance, puis d'une Géostatistique Non Linéaire pour résoudre des problèmes de dépassements de seuil ou de changement de support — la Géostatistique Non Stationnaire–Non Linéaire restant encore actuellement à faire... Des formalismes nouveaux apparaissent, qui dépassent de beaucoup les problèmes classiques d'estimation : Simulations (conditionnelles ou non), Ensembles Aléatoires. Ce foisonnement méthodologique peut être immédiatement mis en valeur grâce à la remarquable amélioration des moyens de calcul.

Il n'est pas facile de parler d'une "**Géostatistique de troisième génération**", qui est actuellement en pleine expansion. Dans un contexte informatique de plus en plus confortable, la Géostatistique se développe dans les directions les plus variées. Cela se manifeste bien sûr au niveau des champs d'application, qui ne se limitent plus désormais aux ressources naturelles (mines ou pétrole). Mais surtout, plus fondamentalement, les recherches s'orientent vers des directions théoriques extrêmement diverses. Il est en particulier intéressant de noter que l'on se remet à prendre en compte les lois de distribution. Cela ne signifie pas un quelconque retour en arrière, mais bien au contraire que l'on dispose — ou que l'on éprouve le besoin — d'outils théoriques nouveaux, qui dépassent de beaucoup les possibilités des fonctions structurales qui jusqu'ici constituaient la pierre angulaire de toute étude géostatistique.

Présentation de la Géostatistique

Variable Régionalisée et Fonction Aléatoire

Le point de départ de toute étude géostatistique est un **jeu de données réparties dans l'espace** (et/ou le temps). Il s'agit de données **numériques** : nous voulons dire par là que la Géostatistique opère sur des quantités. Une information qualitative ne peut être utilisée en tant que telle que s'il est possible de lui appliquer un codage numérique ; sinon, elle sert essentiellement à délimiter l'extension du domaine de validité du modèle géostatistique mis en œuvre.

Mathématiquement donc, on dispose initialement d'une certaine **fonction**, définie sur un espace métrique et prenant ses valeurs dans $\mathbb{R}$ (éventuellement $\mathbb{R}^n$ ou $\mathbb{C}$). Cette fonction, traditionnellement notée $z(x)$, est appelée **Variable Régionalisée**. Des mesures de teneurs minières, d'altitude ou de bathymétrie, de pression atmosphérique, de teneur en polluant... sont autant d'exemples de Variables Régionalisées scalaires ; la mesure du vent est un exemple de Variable Régionalisée vectorielle à 2 dimensions, ou si l'on préfère de Variable Régionalisée complexe...

Il est possible de s'en tenir à ce niveau d'abstraction, et d'entreprendre un travail théorique direct sur la Variable Régionalisée : c'est l'objet de la **Géostatistique Transitive**, qui ne requiert que des outils mathématiques très classiques, et qui surtout n'impose aucune hypothèse restrictive à la variable étudiée. L'inconvénient majeur de cette approche est qu'elle est étroitement tributaire de la notion de **champ**, domaine **borné** en dehors duquel la Variable Régionalisée est supposée nulle : il est la plupart du temps à peu près impossible de faire dans un résultat la part de ce qui est dû à la structure intrinsèque de la Variable Régionalisée et de ce qui est dû à la géométrie de son champ. C'est pourquoi il est souvent indispensable de prendre le risque de franchir un degré d'abstraction.

La démarche consiste à considérer la fonction Variable Régionalisée $z(x)$ comme **réalisation** d'une certaine **Fonction Aléatoire** $Z(x)$. Il s'agit là d'un **choix méthodologique**, librement consenti, non d'une tentative de serrer la réalité de plus près : par ce choix, on ne prétend pas décider de la "nature" déterministe ou aléatoire du phénomène étudié, mais on sélectionne simplement une panoplie d'outils dont on attend une raisonnable efficacité. Et de fait, on dispose alors de tout l'arsenal issu des théories des Probabilités et des Processus Stochastiques.

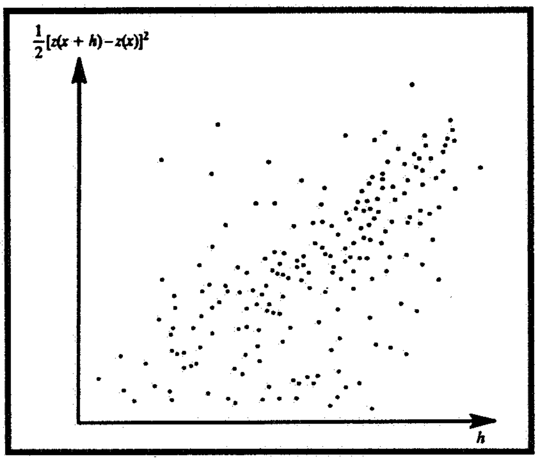
Insistons sur le caractère pragmatique de ce choix. Le modèle probabiliste n'est pas une fin en soi, mais un **outil** que l'on se forge pour répondre à un problème (estimation, simulation...) dont on n'est en général pas maître. Par ailleurs, le fait d'utiliser des probabilités ne préjuge en rien de la nature profonde du phénomène étudié — déterministe ou aléatoire — mais constitue un **choix** stratégique dont l'opportunité ne pourra être jugée que par l'expérience.

Variographie

Dans les modèles probabilistes que nous sommes appelés à mettre en œuvre, l'outil le plus simple pour mesurer une qualité d'estimateur est la **variance**, c'est-à-dire la distance quadratique dans l'espace probabilisé. Ainsi, la mesure quadratique de la distance entre deux valeurs (aléatoires) $Z(x)$ et $Z(y)$ de la Fonction Aléatoire Z sera donnée par la fonction :

$$\gamma(x,y) \quad = \quad \frac{1}{2}\,\mathbf{E}\left[(Z(x)-Z(y))^2\right]$$

(où le symbole $\mathbf{E}$ désigne l'Espérance Mathématique). La théorie permet d'établir que cette fonction γ, appelée le **Variogramme Théorique**, est le seul outil requis pour résoudre les problèmes "linéaires" (estimation sans changement d'échelle, optimisation de réseau d'échantillonnage, qualité d'un estimateur).

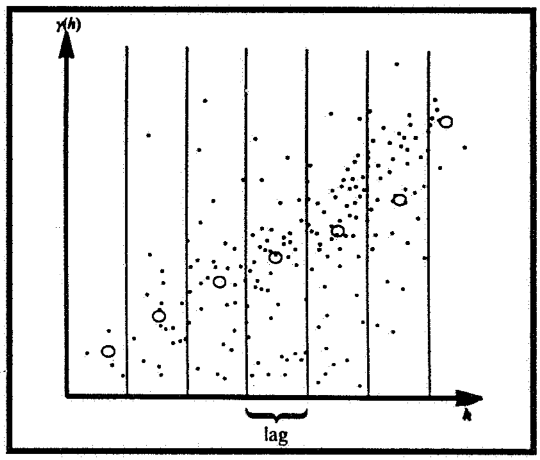


La Nuée Variographique

Au niveau de la Variable Régionalisée, il est donc naturel d'interroger l'écart quadratique entre l'ensemble des données $z(x), z(y)\dots$ La figure ci-contre représente ainsi, pour toute paire de points de données (x, y), le nuage de corrélation entre l'écart quadratique $(z(x) - z(y))^2$ et la distance $||x - y||$.

Ce nuage, appelé **Nuée Variographique**, est un premier outil d'interrogation des données, exhaustif, qui permet de mettre en évidences les valeurs "anormales", les hétérogénéités d'échantillonnage, éventuellement les tendances. Il apporte une information brute, qui n'a fait l'objet d'aucune intervention de l'utilisateur : c'est une image entièrement objective de l'information disponible.

La Nuée Variographique est donc un outil d'analyse extrêmement précieux, à la base d'une "Géostatistique exploratoire" en pleine expansion. Mais elle apporte peu d'information structurale synthétique, et reste difficile à modéliser sans hypothèses fortes. On souhaite donc disposer d'une fonction structurale plus globale, qui puisse exprimer l'**évolution** de l'écart-quadratique entre deux échantillons **en fonction de la distance** entre ces deux échantillons. Plus précisément, on souhaiterait construire une fonction qui, accessible à partir des échantillons, constitue en quelque sorte une **version expérimentale** du Variogramme Théorique requis ultérieurement par les algorithmes géostatistiques. Naturellement, cette fonction est présente dans la Nuée Variographique, mais sous une forme **implicite** qui la rend inutilisable.

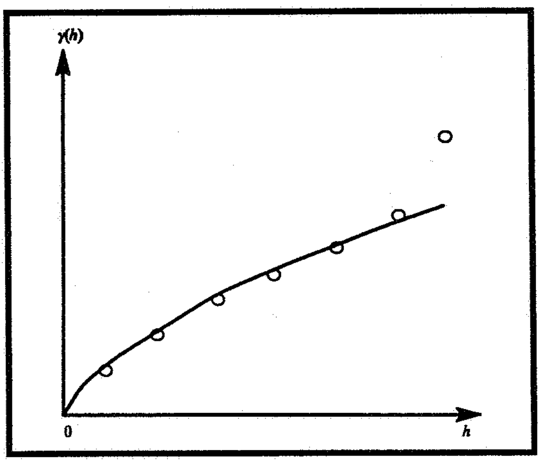


Le Variogramme Expérimental

Aussi va-t-on ranger les couples de données dans un réseau de classes de distances. Par rapport à la Nuée Variographique, on introduit donc une part d'arbitraire puisque, sauf dans le cas particulier où les données sont fournies sur un réseau régulier, le choix des classes de distances est en partie laissé à l'appréciation de l'utilisateur.

L'étape suivante consiste à faire, dans chaque classe de distances, la moyenne des écarts quadratiques correspondants. Pour chaque classe, on obtient donc une valeur unique, l'écart quadratique moyen, et le nuage de corrélation initial est maintenant résumé par une fonction définie pour un petit nombre de valeurs de distances.

Cet **ensemble de valeurs numériques** est appelé **Variogramme Expérimental**.



Le Variogramme Modélisé

Bien que considérablement plus synthétique que la Nuée Variographique, et bien que permettant une bonne interpétation structurale des données, le Variogramme Expérimental n'est pas utilisable en tant que tel dans les formalismes théoriques : il doit être exprimé par une équation, qui permette en particulier de lui attribuer une valeur pour toute valeur de la variable "distance".

L'ultime phase de l'Analyse Variographique consiste donc à ajuster "au mieux" sur le Variogramme Expérimental une courbe d'expression théorique connue.

Cette **fonction** est appelée **Variogramme Modélisé**.

L'usage est de se limiter à un jeu assez restreint de modèles de base pour exprimer le Variogramme Modélisé : on peut renvoyer à la bibliographie pour trouver les formules des variogrammes **exponentiel**, **sphérique**, **monômial**, sans oublier l'**effet de pépite** qui correspond à des données sans structure spatiale et dont l'étude géostatistique revient à des statistiques classiques.

Désignons par $\{x_i\}$ l'ensemble des points de données, par h une distance quelconque, et par $N(h)$ le nombre de couples d'échantillons distants de h — éventuellement à une certaine tolérance près. Alors, en résumé, la Variographie consiste à construire le Variogramme Expérimental

$$\gamma^*(h) \quad = \quad \frac{1}{2\,N(h)} \sum_{i=1}^{N(h)} \left[\left(z(x_i + h) - z(x_i) \right)^2 \right]$$

à l'ajuster "au mieux" par une fonction mathématique, et à assimiler le Variogramme Théorique (requis par la théorie) à ce Variogramme Modélisé.

Naturellement, cette démarche exige des précautions. De nature mathématique d'abord : car toute fonction ne peut représenter un variogramme, et les modèles proposés doivent obéir à un certain nombre de contraintes (valeurs positives ou nulles, parité, caractères de régularité, comportement à l'infini ...). Cependant, la gamme des modèles proposés, et le formalisme développé dans la littérature géostatistique, mettent en général l'utilisateur à l'abri de grosses bévues mathématiques.

En revanche, les précautions à prendre quant à la **signification physique** des opérations sont absolument essentielles. Ainsi par exemple, et ceci quel que soit le jeu de données, il est toujours possible de calculer une fonction expérimentale du type $\gamma^*(h)$ ci-dessus. Par construction, le résultat ne dépendra que du facteur de distance h. Mais cela ne prouve nullement que l'on puisse établir un modèle probabiliste satisfaisant dans lequel le Variogramme Théorique $\gamma(x, y)$ ne dépendrait que de la distance $\|x - y\|$: il peut par exemple parfaitement arriver que le Variogramme Expérimental ne puisse être "raisonnablement" représenté par **aucun** modèle autorisé. Dans un pareil cas, il est évidemment parfaitement vain de vouloir à toute force contraindre la réalité à se plier à nos modèles.

L'essentiel de la démarche variographique est d'assurer le parallèle entre les manipulations numériques réalisées au niveau de la Variable Régionalisée, et les développements théoriques au niveau de la Fonction Aléatoire. On peut imaginer les opérations les plus extravagantes sur un jeu de données : mais ces opérations n'auront de sens que dans le cadre d'un modèle cohérent. A l'inverse, on peut imaginer des développements de pure théorie au niveau de la Fonction Aléatoire : mais ces développements demeureront de la mathématique abstraite si l'on ne sait pas leur associer une interprétation physique. Or ce parallèle, indispensable au **réalisme** d'une étude pratique, ne va pas de soi.

L'élaboration du variogramme est une excellente illustration de ce problème. Pour que la moyenne spatiale qui conduit à la construction du Variogramme Expérimental $\gamma^*(h)$ puisse être interprétée comme une estimation de l'espérance mathématique qui figure dans la définition du Variogramme Théorique, il faut que le modèle probabiliste que l'on cale sur les données ait de bonnes propriétés de **stationnarité** et d'**ergodicité**. Nous reviendrons ultérieurement sur ces notions ; mais ce qui est important de noter ici, c'est qu'il s'agit cette fois de propriétés qui sont **réfutables** : il n'est pas rare que les manipulations numériques effectuées sur les données fassent apparaître des résultats incompatibles avec de tels "bons" modèles. Dès lors, il serait totalement vain de poursuivre des développements mathématiques, même théoriquement corrects, qui n'auraient plus d'interprétation physique raisonnable.

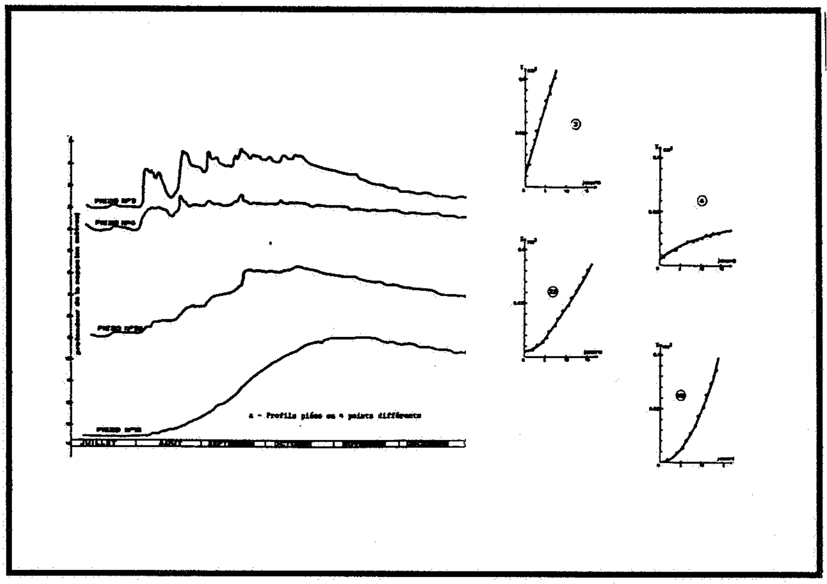
Ces remarques ont surtout pour but de souligner combien une étude géostatistique est tributaire de la qualité de l'étape de Variographie. Le modèle étant non une fin en soi, mais un outil au service des données, il est absolument indispensable de contrôler en permanence l'adéquation du modèle à la réalité. Il est donc parfaitement normal, au cours d'une étude, d'avoir à revenir en arrière parce que des informations nouvelles mettent en question la validité du modèle. En cas de conflit entre données et modèles, le dernier mot doit toujours revenir aux données.

Une dernière remarque. Même adapté au données, le modèle présente des risques. Car le domaine de validité d'une expression mathématique (l'équation d'un Variogramme Théorique par exemple) est illimité. Or le calage du modèle a toujours lieu sur un domaine borné, de sorte que l'on ne peut rien dire de la validité du modèle ni des résultats auxquels il conduit dès que l'on dépasse le cadre dans lequel il a été obtenu. Il n'existe là, par nature, aucun garde-fou. Il est donc de la seule responsabilité de l'utilisateur de ne pas dépasser les limites du raisonnable, et de se garder de faire dire au modèle plus qu'il ne peut. Il ne suffit pas de disposer d'une mathématique correcte, il faut aussi faire preuve de réalisme. Ainsi, à tout travail géostatistique est associée la très importante notion d'**échelle**, qui représente le cadre de validité des opérations envisagées. Tout résultat d'une étude géostatistique doit être référé à son domaine de validité et à son échelle de travail.

Interprétation structurale

Avant d'entrer dans des équations sous sa forme modélisée, le Variogramme Expérimental permet de visualiser et d'interpréter certaines caractéristiques des données. L'exemple proposé ci-dessous (*), réel, en est une illustration.

Les données étudiées sont des cotes piézométriques, relevées toutes les heures pendant la saison des pluies (Juillet–Décembre) à Korhogo (Côte d'Ivoire). Quatre piézomètres, notés respectivement n^0 3, 4 33 et 18, sont examinés ici. Implantés sur le flanc d'une colline, ils correspondent dans l'ordre où ils sont fournis à une profondeur croissante de la nappe d'eau.



Piézométrie à Korhogo

Le profil des niveaux piézométriques s'interprète aisément. Au piézomètre n^0 3, pour lequel la nappe est très proche de la surface, le niveau réagit à chaque pluie. Ainsi le profil piézométrique est-il très erratique, répondant à chaque passée pluvieuse. Au piézomètre n^0 4, l'épaisseur des terrains provoque un effet d'amortissement, et le profil est déjà moins irrégulier. Avec le piézomètre n^0 33, cet effet s'accentue, et les effets de recharge brutale de la nappe phréatique s'effacent au profit d'un comportement de plus grande ampleur. Pour le piézomètre n^0 18, où la nappe est très profonde, les épisodes pluvieux brutaux (à l'échelle de la journée) sont indiscernables ; mais on note par contre une très importante composante saisonnière, qui s'échelonne sur la totalité de la période examinée, et qui correspond à l'accumulation d'eau sur la totalité de la saison pluvieuse.

Ces effets sont bien visibles sur les chroniques, et faciles à interpréter. Ce qui est intéressant du point de vue géostatistique, c'est qu'ils se manifestent immédiatement au niveau des Variogrammes Expérimentaux temporels. Ainsi, le variogramme du piézomètre n^0 3 fait apparaître une croissance importante, de type linéaire, à l'échelle de la quinzaine de jours. De plus, il présente à l'origine une discontinuité, un **effet de pépite**, qui traduit qu'il peut exister de forts contrastes entre deux mesures immédiatement voisines. Cet effet subsiste pour le piézomètre n^0 4, mais la croissance du Variogramme Expérimental est cette fois plus lente. Par comparaison au piézomètre n^0 3, cela signifie que l'écart (quadratique moyen) entre deux mesures séparées du même laps de temps, est plus faible au piézomètre n^0 4 qu'au piézomètre n^0 3 — ou si l'on préfère, que la corrélation entre les deux mesures y est plus forte. On peut également traduire cela en disant que les données du piézomètre n^0 4 présentent une structuration plus marquée que celles du piézomètre n^0 3.

(*) Exemple dû à J.P. Delhomme, cité dans *Journel [1977]*

Cet effet se poursuit avec le piézomètre n^0 33. S'il subsiste un effet de pépite (discontinuité à l'origine), on observe par ailleurs que le Variogramme Expérimental a cette fois une allure parabolique. La conséquence en est que, jusqu'à un intervalle de 6 ou 7 jours, l'écart quadratique moyen est plus faible au piézomètre n^0 33 qu'au piézomètre n^0 4. Mais cet ordre s'inverse pour des intervalles plus importants. Cela s'explique ainsi : au piézomètre n^0 33, le contraste entre deux mesures est essentiellement dû à la composante saisonnière, qui est devenue prédominante ; si deux mesures sont très espacées dans le temps, l'une sera au début de la saison humide et l'autre à la fin, et elles seront de ce fait très contrastées, quels qu'aient pu être les épisodes pluvieux particuliers au cours de la saison. Cette situation, typiquement, correspond en Géostatistique au cas où un modèle stationnaire ne peut être adéquat. Naturellement, cet effet est encore amplifié en ce qui concerne le piézomètre n^0 18, où l'effet de pépite a totalement disparu : le profil piézométrique, comme le variogramme associé, ne représentent plus qu'une ample tendance saisonnière.

Cet exemple est élémentaire : les Variogrammes Expérimentaux ont des formes simples, et l'essentiel des différences entre les quatre situations tient à leurs **comportements à l'origine**. Ce comportement est révélateur du degré de continuité spatiale du phénomène, c'est-à-dire de son comportement aux très petites échelles. Mais d'autres traits du Variogramme Expérimental peuvent également apparaître, qui sont très importants pour la compréhension de la physique des données. C'est le cas en particulier de l'**anisotropie**, qui peut se manifester lorsqu'on travaille dans un espace à plus d'une dimension. Il peut alors arriver que le Variogramme Expérimental n'ait pas les mêmes comportements dans toutes les directions, certaines orientations ayant une importance privilégiée pour des raisons physiques (par exemple dans les processus de sédimentation, ou encore, en climatologie, dans les phénomènes liés au relief, etc...). Il serait alors désastreux de faire un calcul "fourre-tout" du Variogramme Expérimental et de faire une moyenne des calculs dans toutes les directions. Naturellement, si l'on tient compte de l'anisotropie au cours de la modélisation, on augmente le nombre de paramètres du modèle, et donc la liberté de l'utilisateur. Mais dans le même temps, on fait surgir de nouvelles contraintes de cohérence, et donc la phase d'ajustement du Variogramme Théorique peut rapidement devenir très délicate.

Un autre effet important qui peut apparaître sur des données complexes est le phénomène de **structures gigognes** : le Variogramme Expérimental manifeste alors une superposition de composantes d'échelles différentes, que l'on peut essayer de particulariser au cours de la modélisation. Cette fois encore, on peut augmenter considérablement la combinatoire des paramètres. Cette approche est intéressante en ceci qu'elle peut permettre une interprétation de la genèse des données étudiées. L'analyse des structures gigognes, surtout dans ses développements multivariables, s'apparente à l'Analyse des Données et, dans la mesure où l'on sépare des **échelles** différentes, peut être comparée à une Analyse de Fourier : l'intérêt de la méthode géostatistique réside ici en ce que la structure spatiale est fondamentalement prise en compte, et que l'on n'est contraint ni par la dimension de l'espace de travail, ni par une condition de régularité de la maille des données. Notons par ailleurs que si l'un des phénomènes gigognes isolés peut être assimilé à un bruit, il est possible de proposer une méthode géostatistique de **filtrage**, et à cet égard la Géostatistique peut être comparée au Traitement du Signal.

Il faut enfin rappeler qu'il est toujours possible que l'Analyse Variographique conduise à une impasse, par exemple que les résultats expérimentaux obtenus soient incompatibles avec un modèle $\gamma(x, y)$ qui ne dépendrait que de $||x - y||$: il s'agit là typiquement d'une atteinte à l'hypothèse de stationnarité, qu'un utilisateur exercé sait d'ailleurs reconnaître très rapidement. Cette situation est suffisamment importante pour qu'une attention particulière ait depuis longtemps été portée à la Géostatistique Non Stationnaire ; ainsi, la notion par ailleurs intuitive de **tendance** peut être clairement formalisée, et incluse dans les calculs d'estimation ou de simulation. Cependant, pour ne pas alourdir les exemples, nous nous limiterons ici au cas de l'existence d'un variogramme stationnaire.

Modélisation

Revenons une fois encore sur la phase de modélisation. C'est une étape indispensable puisque le formalisme théorique de la Géostatistique "se nourrit" d'une expression mathématique, non d'un ensemble de valeurs expérimentales. C'est aussi une étape cruciale puisque l'on quitte le domaine des certitudes (les données) pour celui de la spéculation. Cette phase présente donc bien sûr des risques, mais elle donne aussi au praticien l'occasion d'inclure dans son travail des informations non numériques, des facteurs subjectifs, son expérience.

Très schématiquement, la modélisation consiste à caler une expression mathématique qui ajuste "au mieux" les quelques points du Variogramme Expérimental. Il s'agit donc du problème somme toute classique de construire une fonction d'interpolation. Nous avons noté cependant qu'il existe un certain nombre de contraintes théoriques sur l'expression finale, et qu'il est par ailleurs souhaitable de faire preuve d'économie dans le choix des paramètres qui caractériseront le modèle.

En pratique, la phase de modélisation est essentiellement interactive. L'utilisation des moyens de visualisation les plus performants est ici particulièrement fructueuse pour représenter et interroger l'ensemble des outils (géo)statistiques — histogrammes, nuages de corrélation, nuées variographiques et corrélations différées, variogrammes, etc... — qui traduisent la structure de la variable étudiée. C'est un véritable **dialogue** qu'il s'agit d'établir avec les données. Et, en Géostatistique de routine, l'instrument privilégié de ce dialogue reste essentiellement le Variogramme Expérimental, qui constitue le principal révélateur de cette structure que l'on veut traduire mathématiquement.

Influence du modèle

Le respect de la structure constitue, par exemple lors de l'ajustement du Variogramme Modélisé, ce que l'on peut appeler les "**contraintes amont**" de la modélisation. Mais le choix d'un modèle a bien sûr des conséquences sur la mise en œuvre des algorithmes : l'ajustement proposé n'est pas indifférent pour la suite de l'étude. Il existe donc également en Analyse Variographique des "**contraintes aval**".

Ce point est important. En général, le Variogramme Expérimental laisse une certaine latitude dans le calage des paramètres du modèle et, sans contrainte supplémentaire, il n'y a pas de critère irréfutable pour trancher entre plusieurs modèles plausibles. On pourrait, bien sûr, envisager des test statistiques. Cependant, un tel choix implique en général une spécification accrue du modèle, que souvent en Géostatistique la quantité d'information disponible ne permet pas. De surcroît, il est **légitime** d'orienter les choix méthodologiques en fonction du problème posé. Si on revient à l'exemple introductif de la bathymétrie du Golfe de Gascogne, il est clair que l'on ne souhaitera pas les mêmes caractéristiques à la carte finale, selon que celle-ci sera destinée à un manuel de Géographie (carte des traits généraux de la région), à un commandant de navire (carte de sécurité, attirant l'attention sur les hauts-fonds), à un commandant de sous-marin (diagnostic de pics : quelle hauteur, quelle implantation), à des pêcheurs, etc...

En résumé : derrière les données qui, elles, sont incontournables, il n'y a pas un "vrai modèle" à trouver, mais seulement un choix de paramètres à ajuster, respectant bien sûr les données, mais également adapté au problème posé. L'essentiel, lors du résultat, est de se bien remémorer les conditions dans lesquelles le travail a été effectué, et donc de toujours associer à ce résultat un **domaine de validité**.

En exemple de l'influence "aval" du choix du modèle, nous présentons ci-dessous les pondérateurs optimaux diagnostiqués pour l'estimation linéaire au centre d'un cercle (*) : ce sont les poids à attribuer à chacune des données pour calculer la valeur estimée au centre, le poids total étant égal à 1. En vis-à-vis de chacun des 12 points de données — ici irrégulièrement espacés — figurent ces pondérateurs.

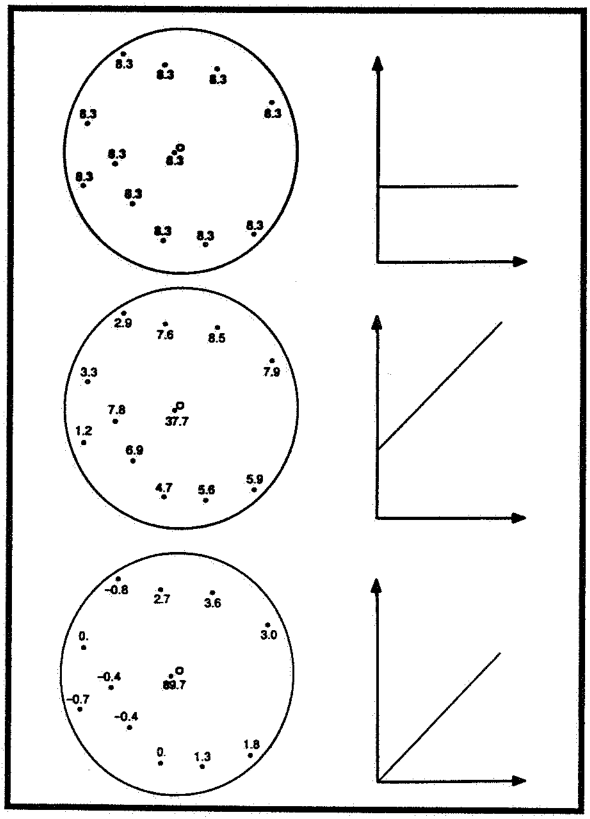
Dans le cas d'un modèle **pépitique pur**, tous les pondérateurs sont égaux, et valent donc $\frac{1}{12}$. Cela vaut en particulier pour la donnée la plus proche du centre à estimer, qui n'a donc pas plus d'importance que les 11 autres. Ce phénomène s'explique aisément : l'effet de pépite pur exprime que l'écart quadratique moyen entre deux données est indépendant de leur distance, autrement dit que la variable étudiée ne présente **aucune structure spatiale**. Il est naturel que cette propriété apparaisse au niveau de l'estimateur. En l'occurrence, on se trouve ici dans le cas des **Statistiques** classiques, non spatiales.

Dans le second cas, il subsiste dans le modèle de variogramme une composante pépitique (pouvant par exemple correspondre à une erreur de mesure) mais la variable manifeste de plus une **structure**, ici de type linéaire. Cela traduit, lorsqu'augmente la distance d'un point de mesure au point à estimer, une **dégradation de l'information** apportée par ce point de mesure. Effectivement, on constate que la donnée proche du centre à estimer joue cette fois un rôle important (près de 40% du poids total de l'information). Comme cela était prévisible, les autres poids ont tendance à décroître quand on s'éloigne du point à estimer.

(*) Les notions d'estimation linéaire et d'optimalité seront précisées aux paragraphes suivants.

Le troisième cas correspond à un modèle linéaire. La variable est cette fois fortement structurée. L'absence d'effet de pépite traduit la continuité de la variable. Comme il n'y a plus de risque d'avoir des écarts de valeurs importants en des points rapprochés, il est clair qu'une information très voisine du point à estimer aura un rôle décisif. Et de fait, le point proche du centre a cette fois un poids de près de 90%. C'est dire que les autres informations sont de moindre importance.

Il est à noter que l'estimateur linéaire optimal peut faire apparaître des **pondérateurs négatifs** : l'expérience montre que c'est une circonstance fréquente pour les variables fortement structurées, et rien dans la théorie n'interdit ce phénomène. Rien n'interdit non plus qu'une donnée ait un poids supérieur à 100%. C'est un point qu'il convient de garder à l'esprit, et qui selon les études peut se révéler tantôt favorable et tantôt néfaste. Ainsi, en topographie ou bathymétrie par exemple, cela signifie qu'il est parfois possible d'obtenir un estimateur qui sort de l'enveloppe des valeurs des données — ce qui n'est pas une mauvaise chose ; mais cela signifie aussi que l'on court le risque d'obtenir une teneur négative en estimation minière...



Influence de l'effet de pépite

Critère de qualité

Dans l'exemple précédent, nous avons parlé d'estimateur (linéaire) **optimal**, ce qui revient à dire que nous disposons d'un critère de qualité pour ce type d'estimateurs. Dans l'ensemble des méthodes géostatistiques, c'est la **variance** qui est adoptée comme critère.

Ce mot de "variance" risque d'être ambigu, et mérite d'être précisé. Afin de définir théoriquement le critère de qualité, nous parlons pour le moment de la variance **au niveau du modèle probabiliste**, c'est-à-dire au niveau de la Fonction Aléatoire. Il s'agit donc d'une **notion mathématique** parfaitement définie. Cela étant, notons qu'en adoptant la variance comme critère de qualité, nous faisons un **choix entièrement libre**. Les avantages en sont immédiats : le calcul théorique de la variance de toute expression linéaire de la Fonction Aléatoire ne requiert que la connaissance du variogramme. C'est dire que ce critère n'est pas trop exigeant en matière de spécification du modèle probabiliste. Mais bien sûr, corrélativement, ce choix présente des limites. Car une variance est un outil assez fruste, qui dans une loi de distribution ne "voit" pas des traits aussi importants que les dissymétries, les multimodalités, etc... Une fois de plus, l'essentiel est d'être conscient de ces limitations au moment du choix méthodologique. Il convient aussi de mentionner que le recours à des outils plus fins exige des modèles considérablement plus spécifiés, et sans doute notablement moins réalistes, que ceux qui suffisent à la Géostatistique Stationnaire.

Le calcul mathématique proprement dit est sans difficulté. Soient v et w deux domaines quelconques de l'espace de travail, et soient $\mid v \mid$ et $\mid w \mid$ leurs mesures (aires dans $\mathbb{R}^2$, volumes dans $\mathbb{R}^3$, etc...). Désignons par $\bar{Z}(v)$ la valeur moyenne sur v de la **Fonction Aléatoire** Z :

$$\bar{Z}(v) \quad = \quad \frac{1}{|v|}\int_v Z(x)\,dx$$

et de même par $\bar{Z}(w)$ sa valeur moyenne sur w. Ces moyennes spatiales sont des **Variables Aléatoires**, ainsi que leur différence $\bar{Z}(v)-\bar{Z}(w)$; cette différence, appelée **Erreur d'Estimation**, représente l'erreur commise (dans le modèle probabiliste) en estimant $\bar{Z}(v)$ par $\bar{Z}(w)$

Nous supposerons toujours que l'estimation de $\bar{Z}(v)$ par $\bar{Z}(w)$ est **sans biais**, c'est-à-dire que

$$\mathbf{E}\left[\bar{Z}(v)-\bar{Z}(w)\right] \quad = \quad 0$$

La **Variance d'Extension $\sigma_E^2(v,w)$ du domaine v au domaine w** est par définition la variance de l'Erreur d'Estimation et, compte-tenu de l'hypothèse de non biais, s'écrit donc

$$\sigma_E^2(v,w) = \mathbf{E}\left[\left(\bar{Z}(v) - \bar{Z}(w)\right)^2\right]$$

Remarquons que cette formule est écrite en toute généralité, sans attribuer de signification particulière aux domaines v et w. En pratique, v sera effectivement un domaine — éventuellement réduit à un seul point — sur lequel on souhaite estimer la valeur moyenne de Z, tandis que w représentera l'ensemble fini $\{x_\alpha\}$ $(\alpha = 1, \ldots, N)$ des N données que l'on souhaite utiliser pour réaliser cette estimation. On aura ainsi :

$$\bar{Z}(w) = \frac{1}{N}\sum_{\alpha=1}^{N} Z(x_\alpha)$$

Quoi qu'il en soit, si l'on pose

$$\bar{\gamma}(v,w) = \frac{1}{|v||w|}\int_v\int_w \gamma(x-y)\,dxdy$$

cette Variance d'Extension s'exprime à l'aide du seul variogramme, par la formule :

$$\sigma_E^2(v,w) = 2\bar{\gamma}(v,w) - \bar{\gamma}(v,v) - \bar{\gamma}(w,w)$$

Cette formule, fondamentale, permet quelques observations :

- La Variance d'Extension $\sigma_E^2(v,w)$ est symétrique en v et w : selon ce critère, l'estimation de $\bar{Z}(w)$ par $\bar{Z}(v)$ est systématiquement de même qualité que celle de $\bar{Z}(v)$ par $\bar{Z}(w)$.
- $\sigma_E^2(v,w)$ **n'est pas une variance conditionnelle**, en ce sens qu'elle ne dépend pas de la valeur particulière des données.
- Par contre, $\sigma_E^2(v,w)$ dépend étroitement de la fonction structurale γ, et en particulier de son comportement à l'origine — c'est-à-dire du degré de continuité de Z.
- La Variance d'Extension dépend à la fois de la géométrie du domaine à estimer — $\bar{\gamma}(v,v)$ —, de la géométrie de l'échantillonnage — $\bar{\gamma}(w,w)$ —, et de leur disposition relative — $\bar{\gamma}(v,w)$ —. Pour la plupart des modèles de variogrammes, $\sigma_E^2(v,w)$ augmente lorsque la distance entre v et w augmente ; par contre, pour une variable totalement non structurée (effet de pépite pur), $\sigma_E^2(v,w)$ ne dépend plus de cette distance.

La formule de la Variance d'Extension est capitale, parce qu'elle permet de calculer la variance d'estimation de toute expression linéaire des données : moyennes, mais aussi intégrales, convoluées et, moyennant quelques précautions de passages à la limite, dérivations. Les applications en sont diverses :

- Pour une configuration d'estimation imposée *a priori*, calculer sa variance. Ce calcul est possible, puisque $\sigma_E^2(v,w)$ ne dépend pas des valeurs particulières prises par les données. La Géostatistique Linéaire permet donc de compléter des estimations linéaires classiques par une évaluation de leur qualité. Ce type de problème est connu sous le nom d'**Estimation Globale**.
- Dans le même ordre d'idées, le critère de la Variance d'Extension permet de définir une stratégie de reconnaissance d'un domaine, d'optimiser un réseau d'échantillonnage ou de rechercher une implantation optimale de données supplémentaires, tout cela en tenant compte de la structure spatiale de la variable étudiée.
- De même, on peut utiliser la Variance d'Extension pour le compactage des données : comment retirer des échantillons en assurant une perte minimale d'information.
- Dans une autre optique, on peut au contraire rechercher, dans une famille d'estimateurs linéaires, celui qui conduira à la Variance d'Extension minimale. Cette technique de calcul d'estimateur linéaire optimal, appelée **Estimation Locale** ou plus usuellement **Krigeage**, constitue une des utilisations les plus fréquentes de la Géostatistique de routine. Elle va être détaillée ci-dessous à titre d'illustration.

Auparavant, au-delà de la formulation mathématique, se pose la question de l'interprétation physique de la Variance d'Extension, c'est-à-dire sa transcription en termes de Variable Régionalisée. En fait, cette transcription ne va pas de soi, et ne peut être établie que moyennant de "bonnes" conditions de stationnarité et d'ergodicité (*) — qu'ici nous supposerons satisfaites. **A ces conditions**, la propriété de non-biais et l'expression de la Variance d'Extension s'interprètent comme suit :

Soit un domaine v (éventuellement réduit à un point) sur lequel on veut calculer la valeur moyenne $\bar{z}(v)$ de la Variable Régionalisée, et soit par ailleurs un domaine w, en général constitué d'un nombre fini de points, sur lequel la Variable Régionalisée est connue et dont la moyenne $\bar{z}(w)$ sera prise comme estimateur de $\bar{z}(v)$. Nous supposons fixées les géométries respectives et la position relative de v et w, et nous appellerons **Configuration d'Estimation** l'ensemble constitué par ces deux domaines v et w.

Alors, compte-tenu des hypothèses, si l'on translatait la Configuration d'Estimation dans tout l'espace et si l'on calculait en chacune de ses implantations l'erreur d'estimation expérimentale $\bar{z}(v)-\bar{z}(w)$, la propriété de non-biais exprime que cette erreur serait statistiquement nulle, et sa valeur quadratique moyenne serait par ailleurs égale à la Variance d'Extension $\sigma_E^2(v,w)$.

En pratique, bien sûr, la situation est un peu plus compliquée, d'abord parce que le domaine de travail est toujours borné et qu'il n'est donc pas question d'effectuer des translations dans la totalité de l'espace, ensuite parce qu'il est assez rare, sauf dans le cas de données régulièrement réparties, de pouvoir exactement translater la Configuration d'Estimation. L'interprétation du non-biais est donc à la fois plus intuitive et moins précise, et peut s'exprimer ainsi : si l'on réalise une estimation avec "à peu près" la même Configuration d'Estimation dans toutes les régions du domaine de travail, alors les erreurs commises auront tendance à s'équilibrer. **Il n'y aura pas d'erreur systématique au niveau du résultat global.** Et la variance expérimentale des erreurs commises sera de l'ordre de la Variance d'Extension.

Le Krigeage

Soit le problème le plus simple possible posé en estimation : interpoler en un point x où elle est inconnue la fonction z dont on connaît les valeurs en N points $z(x_1), \ldots, z(x_N)$. Nous nous limitons ici à un **estimateur linéaire**, c'est-à-dire que nous cherchons à construire un estimateur $z^*(x)$ de la forme

$$z^*(x) \quad = \quad \sum_{\alpha=1}^{N} \lambda^\alpha \, z(x_\alpha)$$

les inconnues du problème étant les pondérateurs λ^α.

Le but est que cette quantité $z^*(x)$ "approche au mieux" la valeur inconnue $z(x)$. Transcrit au niveau du modèle probabiliste, et compte-tenu du choix que nous avons fait d'un critère de qualité, cela signifie que l'on cherche à minimiser la variance de l'Erreur d'Estimation $Z^*(x) - Z(x)$. En supposant pour le moment satisfaites les contraintes de stationnarité, cette variance **n'est autre qu'une Variance d'Extension particulière**, qui tous calculs faits s'écrit :

$$\mathbf{E}\left[(Z^*(x) - Z(x))^2\right] \quad = \quad 2\sum_{\alpha=1}^{N} \lambda^\alpha \, \gamma(x, x_\alpha) \; - \; \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N} \lambda^\alpha \lambda^\beta \gamma(x_\alpha, x_\beta)$$

par simple application de la formule générale. Il s'agit d'une forme quadratique selon les coefficients inconnus λ^α, qu'il convient de minimiser.

On a affaire à une minimisation sans contrainte dans l'hypothèse particulière où la Fonction Aléatoire Z est supposée d'espérance mathématique nulle. Mais il s'agit là d'une circonstance exceptionnelle et la plupart du temps, cette espérance — supposée constante dans le modèle en vertu de l'hypothèse de stationnarité — est inconnue. Aussi, pour garantir que l'Erreur d'Estimation $Z^*(x) - Z(x)$ vérifiera bien

$$\mathbf{E}\,[Z^*(x) - Z(x)] \quad = \quad 0$$

on est amené à imposer la contrainte

$$\sum_{\alpha=1}^{N} \lambda^\alpha \quad = \quad 1$$

(*) Ces deux notions capitales ne seront que brièvement évoquées plus loin, pour éviter dans cette présentation sommaire les réelles subtilités du passage du modèle (Fonction Aléatoire) à la réalité (Variable Régionalisée).

Cette contrainte garantit le **non biais** de l'estimateur de Krigeage. On se placera par la suite dans le cas où cette contrainte doit être explicitée, c'est-à-dire avec des pondérateurs de poids total égal à 1.

Le problème du Krigeage se ramène donc à la minimisation d'une forme quadratique sous contrainte linéaire. Finalement, on trouve un système d'équations linéaires :

$$\begin{cases} -\sum_{\beta=1}^{N} \lambda^\beta \gamma(x_\alpha, x_\beta) + \mu = -\gamma(x, x_\alpha) & \forall\alpha \\ \sum_{\beta=1}^{N} \lambda^\beta = 1 & \end{cases}$$

Les inconnues du système sont les pondérateurs λ^α et le **paramètre de Lagrange** μ. Il suffit après résolution du système de reporter les pondérateurs dans l'expression de l'estimateur $z^*(x)$. Quant à la **Variance de Krigeage**, c'est-à-dire la valeur minimum σ_K^2 prise par la Variance d'Extension $\mathbf{E}\left[(Z^*(x) - Z(x))^2\right]$, elle se simplifie en

$$\sigma_K^2 = \sum_{\alpha=1}^{N} \lambda^\alpha \gamma(x, x_\alpha) - \mu$$

où les λ^α et μ sont les solutions du système de Krigeage.

Nous supposerons bien sûr que le système de Krigeage admet une solution unique, situation à laquelle on cherche naturellement toujours à se ramener. Cela signifie en particulier que le modèle de variogramme n'est pas trop pathologique (les variogrammes périodiques peuvent poser des problèmes...), et surtout que la configuration de Krigeage n'est pas dégénérée. Ce dernier point signifie intuitivement que l'information utilisée doit satisfaire à des contraintes de quantité et d'implantation. Cela étant, le système possède un certain nombre de propriétés qu'il faut garder à l'esprit :

- Le Krigeage est une régression linéaire multiple à résidus corrélés.
- Le Krigeage est — par construction — un **estimateur sans biais**, c'est-à-dire que l'Erreur d'Estimation est d'espérance nulle dans le modèle probabiliste. Au niveau de la Variable Régionalisée, cela signifie que l'erreur moyenne sur les estimations ponctuelles est nulle sur un grand domaine.
- Le Krigeage est un **interpolateur exact**, ce qui signifie que lorsqu'on krige en un point de donnée, on retrouve comme estimateur la valeur de la donnée (avec bien sûr une variance de Krigeage nulle)
- Enfin, par linéarité, et **en supposant fixé le jeu de données utilisées**, le Krigeage d'une combinaison linéaire de valeurs ponctuelles est identique à la même combinaison linéaire appliquée aux Krigeages de ces valeurs ponctuelles. Symboliquement, si * représente l'opérateur "estimation par Krigeage" et $\mathcal{L}$ représente un certain opérateur linéaire,

$$\{\mathcal{L}(Z)\}^* = \mathcal{L}(Z^*)$$

C'est pourquoi il suffit en théorie de présenter les équations du Krigeage ponctuel, tous les autres systèmes (valeurs moyennes, convolutions, dérivations ...) s'en déduisant.

Adéquation du modèle à la réalité

Les formalismes présentés jusqu'ici constituent l'ossature de toute étude de Géostatistique Stationnaire. Les phases d'Analyse Variographique, puis d'estimation (globale et surtout locale) font surgir des difficultés pratiques (calage du modèle, choix d'une échelle de travail, sélection des données en vue du Krigeage, interprétation des résultats) qui font essentiellement appel à l'expérience de l'utilisateur, mais ne posent en général pas de problèmes de nature théorique. La méthodologie de ce que l'on peut appeler la "Géostatistique classique" est bien rodée.

Pourtant, même rigoureuse théoriquement, cette méthodologie ne vaut que s'il y a adéquation entre le modèle mathématique et la réalité étudiée. L'un ne s'identifiant pas à l'autre, il y a en permanence risque de divergence — c'est-à-dire risque que les manipulations mathématiques ou informatiques n'aient plus de signification physique. Ce danger doit rester présent à l'esprit au long de toute étude géostatistique, et ceci d'autant plus que les modèles utilisés seront plus sophistiqués.

Ainsi, dès le début, le choix-même de méthodes probabilistes peut poser problème. Comment, à partir d'un phénomène en général unique, proposer un modèle de Fonction Aléatoire avec toute sa richesse ?

Dans certains cas (séries chronologiques), il arrive que l'on dispose de plusieurs réalisations du processus, et une inférence classique est alors envisageable ; mais, en particulier en ce qui concerne les Sciences de la Terre, le phénomène est unique. Quel sens donner à une probabilité sur un tel phénomène ? Cette question n'est pas de pure forme car, à l'issue des développements de calculs, il faudra bien donner un sens physique aux résultats obtenus. C'est là où les deux hypothèses de **stationnarité** et d'**ergodicité** interviennent.

Stationnarité

Par cette première hypothèse, on fait la supposition que le phénomène étudié présente une certaine **permanence structurale** dans son domaine et que par suite, des observations dans différentes portions de l'espace peuvent être considérées comme autant de **réalisations différentes d'un même processus**, ce processus que l'on cherche à modéliser. Par cette hypothèse, on a levé l'obstacle de la réalisation unique. Mais notons d'une part que cette hypothèse est, essentiellement, **réfutable**, c'est-à-dire qu'il est parfaitement possible qu'aucun modèle stationnaire ne soit compatible avec les données : il faut alors s'orienter vers la Géostatistique Non Stationnaire. Et remarquons par ailleurs que du point de vue des Statistiques, nous ne sommes pas tirés d'affaire puisque, si nous disposons dorénavant de plusieurs réalisations du processus à modéliser, il est clair que **ces réalisations ne sont pas indépendantes**. Il convient de noter que dans la pratique géostatistique, l'usage est de passer outre cette difficulté, dans la mesure où elle n'introduit pas de biais sur l'estimation du comportement à l'origine du variogramme.

Ergodicité

Un modèle est dit **ergodique** si l'inférence de ses paramètres peut être réalisée à partir de l'une quelconque de ses réalisations. Cette seconde contrainte vient peut-être moins clairement à l'esprit dans la pratique. Et pourtant, **nous sommes obligés de supposer nos modèles ergodiques**, faute de quoi il ne serait plus légitime de faire le parallèle entre les moyennes spatiales réalisées sur la Variable Régionalisée et les espérances mathématiques appliquées à la Fonction Aléatoire. Naturellement, les modèles classiques incluent cette propriété mais, une fois encore, il peut arriver en cours d'étude que cette hypothèse soit réfutée : il faut alors avoir recours à des méthodes qui dépassent la Géostatistique "classique".

Il faut toutefois noter que la propriété d'ergodicité, comme d'ailleurs celle de stationnarité, est essentielle pour la mise en œuvre de méthodes de nature probabiliste. Et finalement, les développements d'une Géostatistique "non classique", dont quelques-uns seront évoqués ci-dessous, cherchent toujours à se ramener à "quelque chose" de stationnaire et ergodique : toute la différence entre les diverses techniques proposées tient à la nature de ce "quelque chose".

Extrapolation du modèle

Une dernière difficulté surgit, que l'on retrouve du reste quel que soit le degré de sophistication de la Géostatistique mise en œuvre.

Il est assez naturel de garder à l'esprit les limitations du modèle vers les grandes distances : on sait bien que, même si le domaine théorique de validité du modèle est infini, il serait déraisonnable par exemple d'utiliser dans les calculs des valeurs du variogramme pour des distances supérieures à celles qui ont servi à choisir le modèle. Il s'agit là d'ailleurs d'une question de bon sens, car le modèle lui-même ne "se défend" pas toujours correctement : ainsi peut-on imaginer des cas où une extrapolation à l'évidence irréaliste serait assortie d'une variance de Krigeage en apparence acceptable...

On oublie peut-être trop facilement que ce type de problème se pose également pour les courtes distances. Une des caractéristiques décisives de la fonction structurale est son comportement mathématique à l'origine (caractères de dérivabilité). Or, fatalement, cette fonction structurale est calée à partir de données en nombre fini, qui ne permettent donc pas le passage aux distances infinitésimales. Cela ne condamne pas la méthode, mais cela signifie que lorsqu'on ajuste un modèle de variogramme, **on ajoute une information qui n'est pas présente dans les données disponibles**. Cette situation, qui est incontournable, est d'ailleurs une bonne chose : si une part d'initiative n'intervenait pas à certaines étapes, le travail géostatistique se résumerait à une simple trituration tautologique des données.

Ceci est un exemple de situation où l'expérience est décisive, le géostatisticien disposant d'une inévitable liberté de choix. Tant que l'on ne dispose pas d'information supplémentaire qui pourra venir réfuter ce choix, il est donc légitime d'orienter le calage du modèle, par exemple en fonction de la nature du problème posé : selon que l'on souhaite une carte esthétique, ou sûre, ou précise, ou ressemblant à la réalité, on n'adoptera pas le même variogramme. Il n'y a là aucune malhonnêteté, sous réserve que la part d'arbitraire dans l'Analyse Variographique — nul jugement de valeur dans le mot "arbitraire" — soit clairement circonscrite, et incluse dans le commentaire des résultats.

Les limites de l'outil linéaire

L'outil "Géostatistique Linéaire" est d'utilisation facile, et exige peu de prérequis. Mais du fait que la loi de la Fonction Aléatoire n'est prise en compte que par l'intermédiaire du variogramme, des phénomènes de natures très différentes mais possédant la même fonction structurale seront traités de la même façon, par exemple par le Krigeage. Une telle confusion peut évidemment ne pas être souhaitable, mais c'est le prix à payer de la simplicité.

Il est de la responsabilité du géostatisticien d'évaluer la nécessité de recourir à des techniques plus élaborées. Examinons ainsi à titre d'exemple le Krigeage. C'est le **meilleur** (au sens de la variance minimum) estimateur **linéaire**. Mais la gamme des estimateurs linéaires est-elle un "bon" choix ? Et le critère de la variance minimale est-il adéquat ? On sait que, pour une Fonction Aléatoire gaussienne, les régressions sont linéaires ; on peut alors montrer facilement que l'estimateur de Krigeage s'identifie à l'Espérance Conditionnelle : on est donc assuré que le Krigeage est le meilleur estimateur mesurable possible. Mais que se passe-t-il lorsqu'on "s'éloigne" de la loi gaussienne ? L'expérience montre bien que pour des lois très dissymétriques, les estimateurs linéaires ne sont plus adaptés et que même pour résoudre de simples problèmes d'estimation ponctuelle, il faut recourir aux techniques de la Géostatistique Non Linéaire.

De même, il n'est pas assuré que le critère de la variance soit toujours suffisant. Il ne "voit" pas par exemple les dissymétries, les multimodalités, etc... La symétrie en v et w de la Variance d'Extension $\sigma_E^2(v, w)$ peut légitimement être considérée comme non satisfaisante. De plus, la variance que l'on minimise dans le krigeage n'est pas conditionnelle, de sorte que les pondérateurs seront les mêmes que l'on se trouve dans un zone de faibles ou de fortes valeurs des données...

Ces inconvénients sont réels, et il faut les garder à l'esprit. Mais ils sont probablement peu de chose vis-à-vis de la tentation de mettre en œuvre des modèles irréalistes, totalement déconnectés des données. Par exemple, il serait tentant d'utiliser des médianes plutôt que des espérances, pour des raisons de robustesse. Malheureusement, le moindre calcul mettant en jeu des informations de supports différents devient immédiatement inextricable si l'on utilise des médianes, et l'on est amené soit à faire des hypothèses totalement incontrôlables sur le modèle, soit à proposer des approximations *ad hoc* qui dénaturent totalement le travail. L'intérêt de l'opération a complètement disparu.

En conclusion à cette introduction à la Géostatistique Linéaire, le **sens critique** et l'**initiative** du géostatisticien sont des éléments essentiels de la conduite à bien d'une étude. Rien n'est plus étranger à l'utilisation de "boîtes noires" que la démarche géostatistique.

Développements

La Géostatistique évoquée jusqu'ici est élémentaire, et ses indispensables développements sont nombreux. Nous n'en proposons ci-dessous qu'une revue très rapide, et non exhaustive.

Ces extensions de la Géostatistique "de base" sont de deux natures, suivant qu'il s'agit de modéliser des phénomènes qui échappent aux outils de base (Géostatistique Non Stationnaire, Géostatistique Multivariable) ou au contraire qu'il s'agit de répondre à des problèmes méthodologiques nouveaux (Géostatistique Non Linéaire, Ensembles Aléatoires, Simulations Conditionnelles, etc...).

Géostatistique Non Stationnaire

Il arrive fréquemment qu'aucun modèle de variogramme stationnaire ne soit compatible avec les données étudiées. Cet effet peut être dû à l'existence d'une **tendance**, au sens intuitif, c'est-à-dire à un comportement régulier, d'échelle comparable à la taille du domaine considéré (phénomène "basse fréquence"). L'idée directrice est alors de se ramener aux méthodes classiques, en mettant en évidence "quelque chose" de stationnaire au sein de ce phénomène plus complexe.

Une première approche, très intuitive, consiste à proposer une **dichotomie** : on cherche à décrire la variable globale comme superposition de deux composantes, dont l'une pourra être traitée par les méthodes de Géostatistique Linéaire, et l'autre sera considérée comme une **dérive** (ou tendance). Cette démarche est intéressante, en ce qu'elle convient en général aux naturalistes ; c'est d'ailleurs également la démarche adoptée dans l'étude des Séries Temporelles. Ceci explique sans doute que, historiquement, c'est cette approche qui a été formalisée en premier, sous le nom de Krigeage Universel.

Il faut toutefois être prudent dans cette approche faussement naturelle. Car il n'y a en général pas de raison pour que convergent les exigences du naturaliste, qui souhaite une explication physique à la dichotomie, et celles du géostatisticien, qui souhaite faire apparaître une composante stationnaire. De plus, la pratique de l'Analyse Variographique en vue du Krigeage Universel se révèle assez délicate. C'est pourquoi une seconde approche alternative a été proposée, la théorie des Fonctions Aléatoires Intrinsèques d'ordre k (FAI-k). L'optique est cette fois de proposer une **transformation** des données, qui ramène à "quelque chose" de stationnaire. Cette transformation s'apparente à une dérivation d'ordre k, et a donc pour effet de **filtrer** les composantes les plus régulières du phénomène.

Cette seconde approche a pour premier avantage de donner droit de cité à une gamme de fonctions structurales bien plus étendue que la famille des variogrammes, les **Covariances Généralisées**. On est ainsi à même de modéliser des variables bien plus complexes que celles qui relèvent de la Géostatistique Stationnaire. Si l'on désigne par K la Covariance Généralisée et par f^l la famille des fonctions de base décrivant la dérive que l'on filtre, le Krigeage en FAI-k, ou Krigeage Intrinsèque, s'écrit

$$z^*(x) = \sum_{\alpha=1}^{N} \lambda^\alpha z(x_\alpha)$$

où les λ^α sont solution du système

$$\begin{cases} \sum_{\beta=1}^{N} \lambda^\beta K(x_\alpha, x_\beta) + \sum_l \mu_l f^l(x_\alpha) = K(x, x_\alpha) & \forall\alpha \\ \sum_{\beta=1}^{N} \lambda^\beta f^l(x_\beta) = f^l(x) & \forall l \end{cases}$$

Dans ce système, les inconnues μ_l sont des paramètres de Lagrange : il y en a autant que de fonctions de base f^l à filtrer. A ceci près, le système de Krigeage Intrinsèque est identique à celui obtenu en Krigeage avec variogramme, si ce n'est qu'il faut remplacer la fonction $-\gamma$ par K. Par suite, les propriétés du Krigeage Intrinsèque sont similaires à celle du Krigeage avec variogramme : linéarité, interpolation exacte, non biais, etc...

L'approche des phénomènes non stationnaires dans l'optique FAI-k présente de nombreux avantages :

- L'automatisation de l'Analyse Variographique est beaucoup plus aisée que dans l'optique Krigeage Universel ; elle évite en particulier les **biais** inhérents au traitement par dichotomie. Cependant, les remarques précédentes subsistent : une étude variographique à l'aveugle est strictement à déconseiller.
- On peut déduire facilement du système de Krigeage Intrinsèque une **présentation duale**, soit

$$z^*(x) = \sum_{\alpha=1}^{N} b^\alpha K(x, x_\alpha) + \sum_l c_l f^l(x)$$

  où les coefficients b^α et c_l, qui ne dépendent pas du point à estimer x, sont solutions d'un système linéaire de même structure que le système de Krigeage Intrinsèque. Ainsi, en stockant ces quelques coefficients, il est facile de proposer une estimation extrêmement rapide en autant de points qu'on le souhaite.
- Le formalisme des FAI-k permet d'établir un théorème énonçant l'**équivalence formelle des Splines et du Krigeage.**

Géostatistique Multivariable

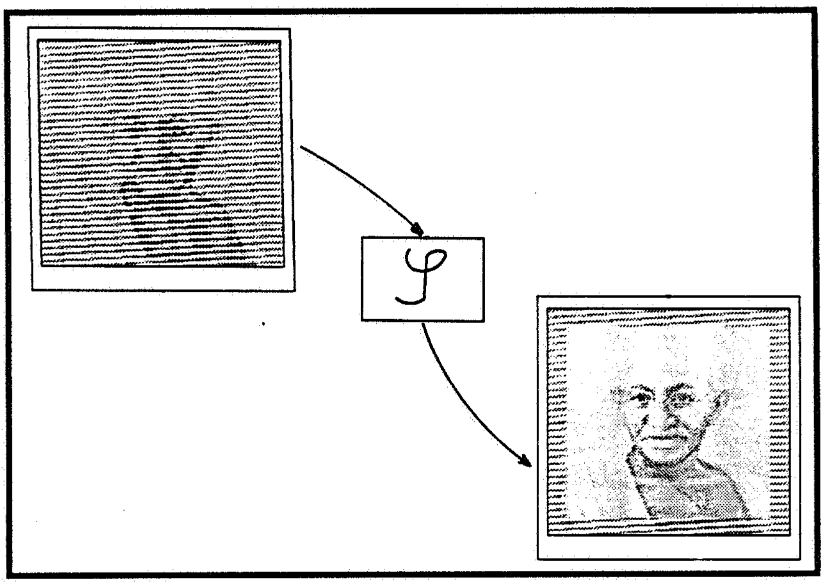
C'est évidemment un problème important en pratique que de pouvoir traiter plusieurs variables simultanément. Dans le cadre de la Géostatistique Linéaire, on sait manipuler des paires de variables, dont la structure conjointe est décrite par les **Variogrammes Croisés** définis par

$$\gamma_{YZ}(x,y) = \frac{1}{2}\,\mathbf{E}\left[(Y(x) - Y(y))(Z(x) - Z(y))\right]$$

où Y et Z désignent les deux Fonctions Aléatoires à étudier. Lors de l'Analyse Variographique, il faut naturellement s'assurer de la cohérence théorique de l'ensemble constitué de tous les Variogrammes Croisés et des variogrammes monovariables : la variographie peut ainsi se révéler délicate...

Les applications de la Géostatistique Multivariable sont nombreuses. Citons :

- Le **Cokrigeage**. Il s'agit d'estimer linéairement une variable à l'aide de données fournies éventuellement sur d'autres variables. Une observation de bon sens permet de prévoir que le Cokrigeage est peu intéressant pour des variables non corrélées (l'une n'apporte aucune information sur l'autre), ou au contraire en trop forte corrélation linéaire (les données sont redondantes). Cela étant, les applications du Cokrigeage sont multiples : on peut par exemple vouloir cokriger un signal à l'aide de données bruitées (**filtrage**). Le Cokrigeage apparaît là comme une généralisation du Traitement du Signal, avec la propriété de n'être contraint ni par la dimension de l'espace, ni par la disposition des données. Notons que rien n'interdit que le bruit soit structuré ou d'espérance non nulle, ni même qu'il soit corrélé au signal. La figure ci-dessous donne un exemple de filtrage par Cokrigeage, appliqué à une image.



Filtrage d'un bruit par Cokrigeage

Comme autre cas d'application du Cokrigeage, on peut encore citer le traitement de données entachées d'**erreurs de localisation**. Ce problème se rencontre tout particulièrement dans le traitement des résultats de campagnes à la mer. En général, les données issues d'un même profil de navigation sont bien situées en position relative, mais il peut y avoir une imprécision notable sur la position relative de deux profils distincts. Par ailleurs, l'incertitude de localisation a une importante répercussion sur l'Analyse Variographique : ainsi, lorsque deux profils se recoupent, il peut arriver que deux mesures réputées implantées au même point aient des valeurs différentes, ce qui se traduit au niveau du Variogramme Expérimental par un **effet de pépite apparent**. La modélisation d'une Variable Régionalisée affectée d'une incertitude de localisation devient très délicate dans le cadre non-stationnaire.

- L'étude de **variables liées par des équations** linéaires (équations aux dérivées partielles, en particulier). Les exemples de tels problèmes foisonnent : gisements d'Uranium où l'on traite simultanément des données de teneurs et de radioactivité (laquelle est approximativement une convoluée de la teneur), météorologie où vents et pression sont reliés par des équations, phénomènes d'écoulement en environnement, etc... C'est d'abord au niveau du modèle structural global que les équations doivent être incorporées. Mais la stratégie de l'utilisation des données (voisinage de Cokrigeage) est également considérablement compliquée par les liens entre variables.
- **Analyse Krigeante**. L'idée est de faire apparaître dans le modèle structural global des composantes élémentaires d'échelles différentes (de fréquences différentes), et de considérer les variables étudiées comme construites par superposition de ces composantes. L'étape suivante consiste à étudier séparément ces composantes élémentaires, en espérant pouvoir leur attribuer une signification physique. Il s'agit en quelque sorte d'une fusion entre Analyse des Données et Analyse de Fourier. On cherche à donner un statut précis aux notions de composantes et d'anomalies. Notons toutefois que l'on exige ici beaucoup du modèle, et que le dialogue avec les naturalistes et l'exigence de réalisme ont par suite une importance cruciale.

- **Dérive Externe.** Il s'agit ici d'utiliser deux variables de natures différentes : la variable d'intérêt est connue par un petit nombre de données très fiables, tandis que les données sur la seconde variable sont beaucoup plus nombreuses, et censées donner une information sur la structure générale de la variable d'intérêt. Comme exemple, on peut citer la prospection pétrolière, où les rares données aux puits sont complétées par une très riche information sismique. L'hétérogénéité des variables, et le petit nombre de données sur la variable d'intérêt, interdisent de faire un Cokrigeage dans de bonnes conditions. Aussi doit-on se contenter d'utiliser la seconde variable comme un guide, qui dessine les grandes lignes (rôle de **dérive**) de la variable d'intérêt. Cette technique est appelée à se développer avec la multiplication des images satellites, qui viennent en complément des données au sol (en pédologie, agronomie, bathymétrie, etc...).

Simulations

A l'occasion de l'exemple introductif, nous avons évoqué comme utilisation possible d'une carte bathymétrique le calcul de la longueur d'un câble sous-marin qui reposerait sur le fond. Pour un tel problème, le Krigeage n'est pas l'outil approprié, parce qu'une carte estimée de donne pas une image fidèle de la structure de la variable. Le processus d'estimation (ici, de simple interpolation) a pour effet de **lisser** les formes : en minimisant la variance, on "gomme" les aspérités et les fluctuations de haute fréquence. Ainsi, dans l'exemple du câble sous-marin, l'utilisation d'une carte krigée conduirait à **sous-estimer** — probablement de façon considérable — la longueur nécessaire. Notons en passant que ce problème est essentiellement **non linéaire**, et qu'il n'est par conséquent guère surprenant que le Krigeage s'y révèle inadapté.

L'idée est donc de construire à partir des données une image plausible de la réalité, c'est-à-dire un **modèle numérique** qui restituera les traits structuraux de la variable simulée. Rien, de plus, n'interdit de réaliser un grand nombre de tels modèles. L'usage de ces modèles est de visualiser des structures, mais aussi de permettre de résoudre des problèmes (par exemple non linéaires) que l'on ne sait pas aborder théoriquement. Il est aussi possible de réaliser des simulations de processus industriels (par exemple d'exploitation minière) sur ces modèles numériques. Mais il faut noter par contre que cette technique ne se substitue pas à une estimation : on peut montrer qu'une simulation ponctuelle constitue un mauvais estimateur (au regard du critère usuel...), dont la variance est double de la variance de Krigeage.

Une **Simulation Non Conditionnelle** est un modèle numérique qui restitue la distribution et la structure (variogramme) de la variable. Il existe de nombreuses techniques de Simulation Non Conditionnelle, parmi lesquelles la méthode des **Bandes Tournantes**; l'idée est de proposer d'abord des simulations à 1 dimension, par des moyens s'apparentant par exemple aux Séries Chronologiques (processus autorégressifs, moyennes mobiles), puis de plonger les résultats obtenus dans $\mathbb{R}^n$ par intégration sur toutes les directions de l'espace : la Géostatistique fournit en effet les formules de passage de $\mathbb{R}^1$ à $\mathbb{R}^n$, formules particulièrement simples pour $n = 3$. Notons que cette méthode des Bandes Tournantes n'est pas fondamentalement différente de certaines techniques de reconstruction en Tomographie rayons X.

Mais il n'y a pas de raison que les valeurs d'une Simulation Non Conditionnelle soient proches de celles des données : il n'y a similitude qu'au niveau statistique, pas au niveau des valeurs particulières. Avec les **Simulations Conditionnelles**, on impose de plus au modèle numérique de respecter les valeurs des données. Le simulation est ainsi "tenue" par des contraintes, d'autant plus fortes qu'il y aura davantage de données conditionnantes. On dispose cette fois d'un modèle qui non seulement ressemble statistiquement à la réalité, mais qui en plus est proche de cette réalité aux points voisins des données.

L'intérêt des Simulations Conditionnelles est évident pour résoudre des questions que l'on ne sait pas attaquer au niveau théorique. On peut par exemple, en utilisant un grand nombre de Simulations Conditionnelles, évaluer numériquement des intervalles de confiance d'estimateur — ce que l'on ne sait pas faire théoriquement avec les seuls outils de la Géostatistique Linéaire. Mais il faut être prudent : en Simulation Conditionnelle, le modèle structural est très fortement sollicité, et il faut toujours veiller à ce que les conclusions proposées soient significatives, et pas seulement l'effet d'artefacts opératoires (problèmes de discrétisation, de troncatures, etc...). Par ailleurs, il est toujours demandé la simulation d'un nombre considérable de valeurs : la recherche d'algorithmes rapides est ici cruciale. Le domaine des Simulations Conditionnelles est ainsi un champ de recherche particulièrement actif.

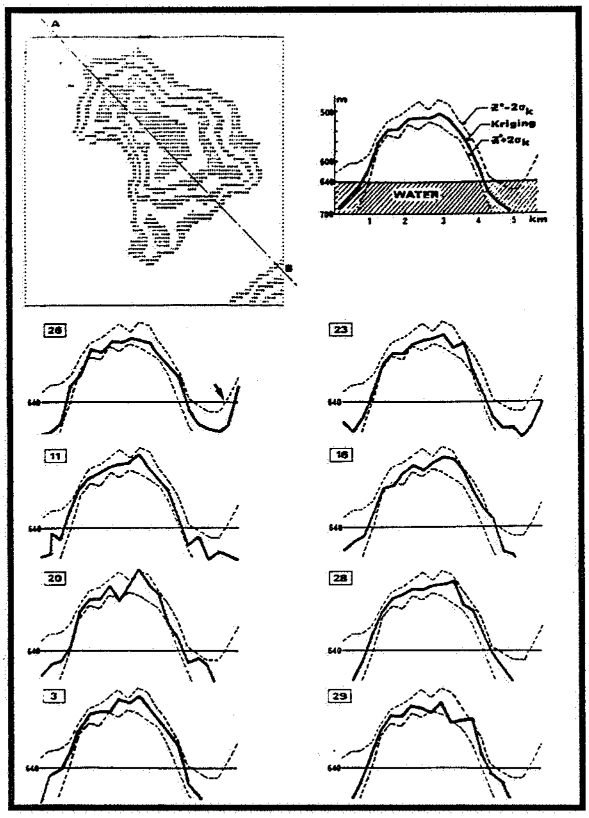
Comme exemple de Simulation Conditionnelle, nous proposons ci-dessous une section verticale d'un dôme pétrolier. La carte représente les isobathes (*) du toit de la structure, obtenus par Krigeage, et la première section représente la coupe selon A-B de ce toit estimé. On a également fait figurer sur cette coupe l'intervalle $Z^* \pm 2\sigma_K$, où Z^* est la valeur krigée et σ_K l'écart-type de Krigeage.

Le problème posé est typiquement non linéaire : il s'agit d'estimer le volume de pétrole piégé, c'est-à-dire le volume compris entre le toit de la structure et le niveau d'eau, supposé connu. Une difficulté importante, et insurmontable théoriquement, est en particulier de déterminer les limites de ce volume, c'est-à-dire l'intersection du toit de la structure et du plan d'eau.

Une première remarque est le caractère en général bien plus irrégulier des Simulations Conditionnelles, par comparaison au Krigeage. Pour visualiser cet effet, les valeurs simulées sont représentées par référence à l'intervalle $Z^* \pm 2\sigma_K$.

On constate, par exemple sur la simulation 20, que les valeurs simulées peuvent sortir de cet intervalle. C'est bien confirmation que, vis-à-vis du critère de variance minimum en tout cas, la Simulation Conditionnelle peut se révéler un bien mauvais estimateur.

Par ailleurs, l'exemple de la simulation 26 attire l'attention sur les risques qu'il y aurait à se limiter à une seule simulation. On voit en effet le toit simulé remonter sur la partie droite de la coupe, et repasser au-dessus du niveau d'eau. Dans cette simulation, la structure susceptible de contenir du pétrole est donc en deux parties ce qui, si cela se révélait conforme à la réalité, serait d'une importance considérable pour l'estimation du volume. En réalité, si l'on examine un grand nombre de simulations, on se rend compte que cette circonstance est exceptionnelle, et qu'il serait par conséquent très dangereux de fonder toute une stratégie sur les propriétés très particulières de cette simulation 26, d'autant plus qu'il s'agit d'une zone simulée en extrapolation et affectée d'une très forte variance théorique.



Simulation d'une structure pétrolière

Approfondissements méthodologiques

La dialectique estimation/simulation permet de compléter la réflexion sur le rôle et l'importance du modèle dans la démarche géostatistique. En simulant conditionnellement une structure donnée, on peut mettre en évidence l'impact du choix du type de modèle structural, ou visualiser l'effet d'un changement de paramètre (portée, effet de pépite, etc...). Dans le même temps, en réalisant un Krigeage à partir des données conditionnantes, on illustre la différence de comportements entre Estimation et Simulation Conditionnelle.

Cette comparaison n'est pas purement théorique. Elle fournit aussi des informations sur la manière dont les modèles réagissent à certaines situations (lacunes dans les données, par exemple), et sur la stabilité des résultats. Ainsi est-il possible d'enrichir l'expérience du géostatisticien par des informations non quantitatives, fort importantes au moment de faire des choix : par exemple, décider entre deux modèles également plausibles lors de l'Analyse Variographique, ou fixer une stratégie de recherche de voisinage lors du Krigeage.

A titre d'illustration, nous proposons, en fin de ce document, cinq planches mettant en évidence quelques notions importantes :

- Importance de la **portée**. Dans un modèle stationnaire, la portée est la distance à partir de laquelle deux données ponctuelles ne sont plus significativement corrélées. Ce paramètre rejoint la notion

(*) Tracé des courbes d'égales profondeurs.

intuitive de zone d'influence. Toutes choses égales par ailleurs, plus la portée sera grande, et plus le phénomène sera structuré à grande échelle. A l'opposé, si la portée tend vers 0, on se rapproche du bruit blanc, et de l'absence totale de structure spatiale.

La première planche représente quatre Simulations Non Conditionnelles, construites avec un modèle de variogramme sphérique, pour quatre valeurs croissantes de la portée. Comme prévu, on voit se dessiner pour les portées 50 et 75 des structures importantes à l'échelle du domaine, et qui étaient totalement absentes en particulier pour la portée 10. Il est d'ailleurs à prévoir, bien que le modèle utilisé pour ces simulations soit stationnaire, que si l'on réalisait dans les règles de l'art une Analyse Variographique sur les simulations de portées 50 et 75, on concluerait — **à l'échelle du domaine** — à un phénomène non stationnaire...

Il est par ailleurs clair que si ces simulations étaient des données à partir desquelles on souhaite réaliser un Krigeage, la politique de recherche de voisinage ne devrait pas être la même pour les portées 10 et 25 d'une part, 50 et 75 d'autre part. Dans le premier cas en effet, l'absence de structure globale marquante fait que l'on ne commet jamais une grosse Erreur d'Estimation si l'estimateur est proche de la moyenne globale; aussi a-t-on intérêt à utiliser un maximum d'information, dont la moyenne sera un estimateur convenable. Au contraire, pour les longues portées, le phénomène met en évidence une géométrie globale fortement structurée, et il est donc primordial pour faire une estimation d'avoir de l'information convenablement placée, c'est-à-dire surtout proche du point à estimer.

Cet exemple permet d'illustrer une règle empirique (ce n'est pas un théorème, et il peut exister des exceptions !) : lorsqu'un phénomène est faiblement structuré, il vaut mieux en général privilégier la **quantité** d'information (Statistiques classiques) ; lorsque le phénomène est fortement structuré, c'est la **qualité** de l'information qui prévaut, c'est-à-dire son implantation.

- Importance du **type de modèle**. La seconde planche propose quatre simulations de modèles ayant des portées pratiques équivalentes. La différence entre ces modèles tient essentiellement aux comportements à l'origine. Les modèles sphérique et exponentiel ont un comportement **linéaire** à l'origine, ce qui correspond à une Fonction Aléatoire **continue**. Le modèle "cubique" est deux fois dérivable à l'origine, ce qui correspond à une Fonction Aléatoire une fois dérivable. Enfin, une Fonction Aléatoire de variogramme gaussien est indéfiniment dérivable.

  La portée étant assez grande par rapport au champ simulé, on voit se former des structures sur ces quatre simulations. Mais dans le cas des modèles à comportement linéaire, ces structures ont des frontières extrêmement déchiquetées, en particulier en ce qui concerne le modèle exponentiel. Pour le modèle cubique, les frontières sont déjà beaucoup plus nettes, et les zones qui se dessinent sont fortement structurées. Dans le cas gaussien enfin, les courbes de niveau sont très lisses, et il est vraisemblable qu'une Analyse Variographique à l'échelle du demi-champ concluerait à l'existence d'une dérive — qui, notons-le au passage, est absente du modèle sous-jacent.

- Importance de l'**effet de pépite**, planche 3. Pour un phénomène de variance globale fixée, on fait croître dans le modèle la part due à l'effet de pépite. L'effet de déstructuration se passe de commentaires. Pour un effet de pépite couvrant 100% de la variance, on a affaire à un bruit blanc, totalement non-structuré.

- Les planches 4 et 5 suivent une présentation commune. Une première figure propose la "réalité", en fait une Simulation Non Conditionnelle réalisée avec des modèles de portées pratiques équivalentes. On retrouve en l'occurrence la différence entre structures sphérique et gaussienne, déjà mentionnée. Puis, sur ces simulations considérées comme la réalité, on effectue un échantillonnage (le même pour les deux modèles — seconde figure). A partir de cet échantillonnage, on réalise le Krigeage en voisinage unique — c'est-à-dire à l'aide de la totalité des données — (troisième figure). Ce résultat est complété par la carte des écarts-types (quatrième figure).

  On a déjà indiqué que le Krigeage lisse, et qu'une carte krigée est plus régulière que la réalité. Cet effet est bien sûr plus marqué pour le modèle sphérique, pour lequel la situation de départ était beaucoup plus erratique. Pour le modèle gaussien au contraire, l'effet est très atténué, et la carte krigée est très proche de la réalité. Autrement dit, dans les conditions présentes, il serait peu dangereux de confondre la réalité avec l'estimateur de Krigeage. Il faut cependant noter que le nombre de points de données est assez important (100), et que le modèle gaussien risquerait de susciter des artefacts si le Krigeage était contraint par un nombre trop faible de données.

Géostatistique Non Linéaire

Les simulations sont une première voie pour résoudre des problèmes non linéaires. Elles présentent cependant le risque que le modèle soit trop sollicité, c'est-à-dire que l'on cherche à faire dire plus qu'il n'est réaliste aux valeurs numériques simulées. C'est pourquoi il est souhaitable de chercher une approche théorique aux problèmes non linéaires, lorsque c'est possible. Trois types de questions sont ainsi traités en Géostatistique Non Linéaire :

- **Effet de support** : la formule fondamentale de la Variance d'Extension permet de calculer le variogramme d'une Fonction Aléatoire régularisée. Mais cette formule est insuffisante si l'on veut construire un **modèle de changement de support**, c'est-à-dire un modèle pour la loi bivariable $\big(Z(x), \bar{Z}(v)\big)$. Le problème se pose en particulier en Géostatistique minière, où les échantillons et les unités de sélection ont des volumes très différents : il importe bien sûr de connaître la loi des unités de sélection **conditionnellement** à la loi des échantillons.
- Toujours dans l'exemple de la Géostatistique minière, on fait souvent des choix (coupures, sélections) sur la base d'estimateurs Z^*, non de vraies valeurs Z. L'estimateur est une régularisée de la valeur vraie, et il est donc clair qu'il n'a pas même loi qu'elle : cet effet est appelé **effet d'information**. Il faut donc pouvoir modéliser les lois bivariables $(Z(x), Z^*(x))$.
- Enfin, on est souvent amené à estimer non pas des valeurs (ponctuelles ou régularisées), mais des **probabilités de dépassement de seuil**. En mine en particulier, la notion de **coupure** est capitale, par exemple pour la construction des courbes tonnage-teneur ; c'est une notion fondamentalement non linéaire.

C'est pour répondre à ces questions de Géostatistique Non Linéaire qu'a été élaboré le Krigeage Disjonctif, outil intermédiaire entre la régression linéaire (le Krigeage) et la régression (c'est-à-dire l'Espérance Conditionnelle). Plus puissant que la Géostatistique Linéaire, le Krigeage Disjonctif a besoin d'une modélisation des lois bivariables, et ne se contente plus des moments d'ordre 2 (variogramme). On voit donc réapparaître la notion de loi, absente de la Géostatistique Linéaire. Et comme en général on ne sait pas manipuler théoriquement des lois quelconques, un outil capital de la Géostatistique Non Linéaire est l'**anamorphose**, transformation qui a pour objet de convertir la Fonction Aléatoire de départ en Fonction Aléatoire de loi connue (en général gaussienne).

Notons enfin que le Krigeage Disjonctif exige une hypothèse de stationnarité très stricte, pour pouvoir réaliser l'inférence des lois bivariables. Ainsi, si la gamme des questions que l'on sait aborder théoriquement est considérablement élargie par rapport à la Géostatistique Linéaire, en revanche, il y a restriction de la gamme des phénomènes que l'on peut traiter. Le développement d'une Géostatistique Non Linéaire-Non Stationnaire est une question entièrement ouverte...

Perspectives

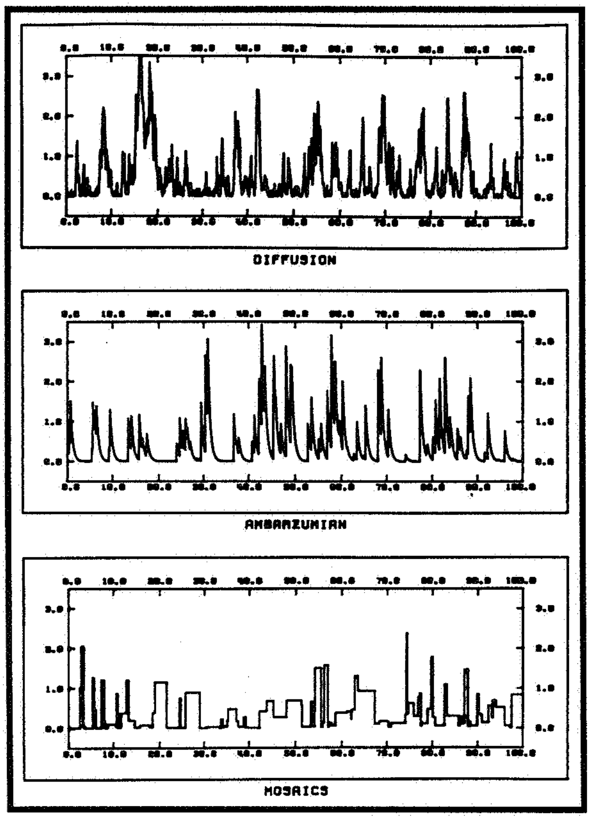
La Géostatistique dispose d'une panoplie d'outils qui s'est considérablement enrichie au cours des années. Il faut cependant être conscient que ces outils peuvent se révéler notoirement insuffisants pour modéliser les traits essentiels de certains phénomènes complexes, et que par suite la recherche d'instruments nouveaux est toujours à l'ordre du jour.

Dans les domaines classiques qui ont fait jusqu'ici l'objet de cette présentation, évoquons quelques-unes des perspectives de développements :

- Il est clair par exemple que le Variogramme Croisé, qui est une fonction paire par construction, ne "voit" pas les décalages systématiques entre variables (qui existent par exemple dans des phénomènes d'écoulement : météorologie, pollution, gisements alluvionnaires, etc...). Dans ce cas, il existe une parade, au moins dans le cadre stationnaire : c'est de travailler en Covariance Croisée plutôt qu'en variogramme. Dans cet exemple, la base théorique est établie, c'est une pratique qu'il faut développer.
- Autre domaine de recherche : le **changement d'échelle**. En particulier, l'utilisation de plus en plus fréquente d'images satellites comme complément d'informations au sol (agriculture, pédologie, océanographie, etc...) oblige à travailler avec des données extrêmement hétérogènes, représentant des domaines d'extensions très différentes. Il se pose donc des problèmes de conversion d'échelles, de surcroît sur des variables qui souvent ne sont pas additives (par exemple des données de nature géométrique).
- Un autre domaine extrêmement important : la modélisation des **phénomènes spatio-temporels**. Il ne faut plus se contenter de "photographier" une Variable Régionalisée, mais il faut décrire son évolution dans le temps. Le problème ne se réduit pas à l'ajout d'une dimension supplémentaire, l'évolution obéissant à des lois physiques, ou étant à tout le moins limitée par des contraintes

d'inégalité. La Géostatistique Multivariable est donc appelée à se développer dans l'étude des variables reliées par des équations.

- Autre sujet de réflexion : quelle information apporte finalement sur la structure d'une variable la connaissance simultanée de la loi de distribution et de la covariance ? La figure ci-dessous représente trois exemples de réalisations de processus ayant même covariance (en l'occurrence exponentielle) et même histogramme.



Trois simulations de mêmes loi et covariance

La première simulation est un schéma de diffusion.

La seconde est un processus d'Ambarzumian (décroissance exponentielle sur un schéma de Poisson). Ce second exemple est particulièrement intéressant, parce qu'il est dissymétrique : il est évident que la loi de $Z(x+h)|Z(x)$ — loi de $Z(x+h)$ **conditionnellement** à $Z(x)$ — n'est pas identique à la loi de $Z(x)|Z(x+h)$. Or, cette propriété n'est pas "vue" par le variogramme...

Enfin, le troisième exemple est un schéma mosaïque : le processus est constant par morceaux.

Un outil qui n'arrive pas à distinguer trois phénomènes aussi différents est bien sûr peu satisfaisant. C'est pourquoi il faut s'orienter vers des méthodes plus puissantes que la Géostatistique d'ordre 2, qui se borne à l'utilisation des **moments**. Un outil d'investigation prometteur est le **nuage de corrélation différée**, nuage de corrélation entre les Variables Aléatoires $Z(x)$ et $Z(x+h)$. Mais la mise en œuvre de tels outils, notablement plus sophistiqués que ceux de la Géostatistique Linéaire, requiert une bien meilleure qualité des données, et en particulier de très fortes conditions de stationnarité.

Et la liste des développement n'est pas close, la Géostatistique étant essentiellement une discipline progressive... En particulier, nous n'avons pas parlé ici du champ extrêmement vaste et prometteur des **Ensembles Aléatoires**, actuellement en plein développement. Il s'agit d'une branche de la Géostatistique qui établit le pont avec les techniques de la Morphologie Mathématique, en accordant une importance privilégiée à l'aspect **géométrique** des données. Les outils mathématiques mis en œuvre, plutôt que numériques, sont de nature **ensembliste**. Les applications sont nombreuses, non seulement pour restituer des images ressemblant davantage à la réalité (avec par exemple en stratigraphie des contraintes d'inclusion), mais aussi pour aborder des problèmes typiquement non linéaires comme par exemple les questions de connexité, capitales pour les phénomènes d'écoulement (réservoirs pétroliers, propagation de polluants). Riche de perspectives d'applications, le domaine des Ensembles Aléatoires est aussi un domaine de recherche plus fondamentale, d'une part en ce qui concerne le problème de l'inférence des modèles, d'autre part en ce qui concerne le conditionnement des simulations.

Lien avec d'autres méthodes

La Géostatistique est par nature une discipline de rencontre entre des méthodes mathématiques et des domaines d'application. Certaines de ces méthodes ont constitué des innovations théoriques (Ensembles Aléatoires par exemple), mais beaucoup d'autres sont tout-à-fait classiques, et se trouvent de ce fait à l'œuvre dans des problèmes bien différents de ceux rencontrés en Géostatistique. Par ailleurs, les praticiens des Sciences de la Nature disposent dans leurs domaines d'outils mathématiques adaptés, souvent de longue date, à leurs situations particulières. Ainsi, la Géostatistique n'est en aucune manière une discipline isolée ; il est donc indispensable de la situer par rapport à d'autres méthodes voisines, et de proposer une revue de ses champs d'application.

Au niveau méthodologique, la Géostatistique propose un choix fondamental :

- la **prise en compte de la structure spatiale des données**,

et sa mise en œuvre se distingue par trois facteurs :

- on travaille à partir d'un **échantillonnage**, c'est-à-dire à partir d'une connaissance partielle, souvent très incomplète, du domaine à étudier ;
- on travaille dans un espace métrique, mais *a priori* **de dimension quelconque**. En particulier, on n'est pas astreint à se ramener à $\mathbb{R}$.
- enfin, il n'est pas nécessaire de se ramener à un maillage régulier. Les procédures proposées sont applicables **quelle que soit la répartition des données**.

Ces quatre aspects de la Géostatistique vont guider la comparaison aux méthodes voisines.

Méthodes Probabilistes

A l'exception des méthodes transitives, la Géostatistique travaille sur des **Modèles Probabilistes**. Le formalisme se fondant sur un critère de variance, c'est donc principalement sur les **deux premiers moments** que l'on est amené à travailler. L'outil approprié, en hypothèse stationnaire, est ainsi la **fonction de covariance** dont les propriétés mathématiques sont définies par le **théorème de Bochner**. Notons que ce théorème, par le lien qu'il fait entre la fonction structurale et sa transformée de Fourier, permet le parallèle avec les **Méthodes Spectrales**. Lorsque les hypothèses de stationnarité doivent être élargies, il faut avoir recours à une généralisation du théorème de Bochner, mais les complications techniques que cela implique ne changent rien à ce parallèle. Il est donc intéressant, dans un enseignement de Géostatistique théorique, de présenter aussi les Fonctions Aléatoires sous l'angle de l'**Analyse Harmonique** afin de permettre la comparaison avec les techniques usuelles par exemple en **Traitement du Signal** : il apparaît ainsi que le Cokrigeage d'erreurs n'est autre qu'un **Filtrage**, et que l'Analyse Krigeante s'apparente à une **Analyse Fréquentielle**. La différence — pratique — tient à ce que la Géostatistique s'accommode aisément du travail à plus de une dimension, et que surtout il n'est absolument pas nécessaire d'avoir des données régulièrement espacées.

Quelles que soient les hypothèses de stationnarité, on se ramène toujours à la manipulation de quantités aléatoires admettant une variance. Le cadre mathématique de travail est donc un espace $\mathbf{L}^2$ de Variables Aléatoires, et la minimisation d'une variance peut donc être présentée comme une **projection** sur un **Espace de Hilbert** de Variables Aléatoires. On dispose ainsi d'une présentation géométrique du Krigeage. Les propriétés du Krigeage sont alors une conséquence immédiate du théorème des projections : par exemple, le **théorème d'additivité** établi dans le cadre du Krigeage Universel n'est autre que le classique théorème des trois perpendiculaires...

On peut enfin examiner le lien entre la Géostatistique et les méthodes de **Processus Stochastiques** — autrement dit de Fonctions Aléatoires. La Géostatistique, travaillant en général dans des espaces à plusieurs dimensions, fait peu usage des formalismes de type **Chaînes de Markov**, **files d'attente**, etc... parce qu'elle ne bénéficie pas d'une structure d'ordre dans son domaine de travail. Plus générale, la Géostatistique peut donc sembler moins adaptée au traitement particulier des **Séries Chronologiques**, surtout parce que le passage à $\mathbb{R}^n$ d'outils spécifiquement monodimensionnels semble extrêmement difficile. Mais des raffinements à l'étude géostatistique de $\mathbb{R}^1$ pourraient apporter un éclairage nouveau dans un domaine où, il est vrai, existent déjà d'autres méthodes parfaitement adéquates. En tout cas, des méthodes de type **Processus de Diffusion** ont déjà leur application en Géostatistique, non pas dans l'espace "géographique" de travail, mais dans l'espace des états pris par la variable (transition de faciès stratigraphiques, par exemple). Par ailleurs, la méthode de simulation des **Bandes Tournantes**, qui commence par simuler des processus monodimensionnels, fait parfois usages d'outils issus des Séries Chronologiques : **Moyennes Mobiles**, **Processus Autorégressifs**, ou plus généralement processus **ARIMA**.

Statistiques

Il est naturel, ne serait-ce qu'à cause du vocable "Géostatistique", de s'interroger sur les liens entre Géostatistique et Statistiques "classiques". La question est, en fait assez complexe.

Une première observation est que, historiquement, la Géostatistique s'est assez vite dégagée des interrogations sur la **loi** de la Variable Régionalisée étudiée pour privilégier l'aspect "structure spatiale". En réalité, cet éloignement est assez trompeur puisque l'on sait que les choix adoptés — recherche d'un estimateur linéaire, choix de la variance comme critère de qualité — sont optimaux dans le cas d'une Fonction Aléatoire gaussienne, et deviennent moins opportuns à mesure que l'on s'"éloigne" du cas gaussien. Autrement dit, l'importance de la loi sous-jacente est grande, même si cette loi n'intervient pas explicitement dans le formalisme de la Géostatistique Linéaire. Du reste, par le jeu des anamorphoses, la Géostatistique Non Linéaire reprend indirectement en compte la loi sous-jacente. Et les développements théoriques actuels vont effectivement dans le sens d'une analyse et d'une modélisation des lois bivariables.

Le vrai problème des relations entre Statistiques et Géostatistique tient probablement au rôle attribué au modèle. Pour le géostatisticien, le modèle probabiliste est fondamentalement un **outil** qu'il se forge pour les besoins d'un problème concret. Il n'y a pas dans la nature de "vrai" modèle caché derrière les échantillons, il n'y a pas un "vrai" jeu de paramètres à trouver. Plus qu'une méthode mathématique, la Géostatistique appliquée doit être considérée comme une **Science Physique**, dans laquelle la sanction de l'expérience prend le pas sur le modèle. C'est pourquoi l'**inférence statistique**, ou ajustement du modèle, est simplement comprise comme une **approximation** que l'on est obligée de faire faute de connaître exhaustivement la Variable Régionalisée, et tout complément d'information survenant en cours d'étude **doit** servir à contrôler le modèle. Ce rôle purement opératoire que le géostatisticien attribue au modèle, explique qu'il n'est pas d'usage de faire des tests statistiques pour évaluer la qualité de l'Analyse Variographique, le meilleur test étant la confrontation à la réalité. Par suite, un paramètre du modèle qui n'a pas de signification physique est au mieux un intermédiaire de calcul, au pire un parasite introduisant des artefacts.

Cette optique explique les difficultés de communication avec le monde des Statistiques. L'exemple le plus classique de divergence de point de vue concerne l'ajustement du Variogramme Modélisé, et plus particulièrement de son comportement à l'origine. Il est possible de concevoir des tests statistiques mesurant la distance d'un modèle au Variogramme Expérimental. Le problème est que justement, plusieurs tests sont envisageables, qui requièrent à leur tour des hypothèses sur le modèle, et dont les conclusions peuvent être fondamentalement divergentes : mentionnons ici que cette approche est proposée dans le logiciel Bluepack, et qu'effectivement les réponses ne sont pas toujours dépourvues d'ambiguïté. Or, on a vu combien les résultats (simulation ou estimation) peuvent être tributaires du modèle. Faute alors de critères incontestables, le géostatisticien préfère s'en remettre à son expérience, et aussi orienter ses choix — tant qu'il en a la liberté — en fonction du problème posé, charge à lui de corriger le modèle lorsqu'un supplément de données le permet (et *ipso facto* l'exige...).

Une approche véritablement statistique n'est possible que lorsqu'il existe réellement un modèle sous-jacent, c'est-à-dire lorsqu'on travaille sur un modèle numérique. Par contre, lorsqu'on travaille sur des échantillons de variables réelles, la philosophie d'une étude géostatistique pourrait avec un peu de provocation s'exprimer ainsi : dès lors que les algorithmes sont mathématiquement corrects, il n'y a pas de résultats vrais ou faux, il n'y a que des résultats adéquats ou non, un résultat n'ayant de signification qu'accompagné du descriptif du mode opératoire (les "**conditions de laboratoire**" en Physique).

Autres méthodes

Surtout dans sa présentation Transitive (non probabiliste), la Géostatistique peut servir à étudier des formes géométriques. Ce qui la différencie des méthodes de type **Analyse d'Image** et **Morphologie Mathématique** est qu'elle utilise des données fragmentaires, échantillonnées : elle présente donc un aspect **interpolatoire** qui n'existe pas lorsque l'on travaille sur une image totalement connue. Cette distinction a cependant tendance à s'atténuer, avec le traitement des images satellites qui de plus en plus sont utilisées comme complément des données au sol. Par ailleurs, il était d'usage de faire entre Morphologie Mathématique et Géostatistique une séparation portant sur la nature des outils structuraux employés : ainsi, en Morphologie Mathématique, on utilise des objets continus (segments, disques) pour réaliser les opérations d'ouverture, érosion, etc... alors que la Géostatistique utilise seulement des objets finis (une paire de points pour les variogrammes, quelques points pour les Covariances Généralisées). Mais ici aussi, la distinction disparaît, et la Géostatistique des Ensembles Aléatoires s'apparente de très près à la Morphologie Mathématique.

Dans sa version sans changement de support, le Krigeage est un interpolateur. On peut donc s'interroger sur ses liens avec les méthodes d'**Interpolation Numérique**. Une première remarque est bien sûr que, contrairement aux méthodes frustes de type "inverse des distances", "inverse du carré des distances", etc..., contrairement aux ajustements polynomiaux par moindres carrés, la Géostatistique a pour vocation de tenir compte de la structure intrinsèque de la variable à interpoler, et ne propose donc pas un estimateur "passe-partout". Une deuxième remarque est que, pour la quasi-totalité des variables naturelles, il n'y a pas de contrainte "aval", par exemple sur l'expression analytique de l'estimateur : ce n'est pas la même optique d'interpoler une carrosserie automobile ou un aile d'avion d'une part, une teneur en polluant ou une pression atmosphérique d'autre part. La Géostatistique, elle, est essentiellement assujettie à des contraintes "amont". C'est donc un intéressant sujet de réflexion que le **théorème** qui exprime l'**équivalence formelle** entre l'interpolation par **Splines** et le Krigeage (notons que les **Splines d'Ajustement** correspondent à un Krigeage avec effet de pépite). Ainsi, bien que les points de vue d'utilisation soient bien souvent diamétralement opposés, Splines et Krigeage sont une seule et même chose — ce qui ne signifie pas d'ailleurs que l'explicitation d'un Krigeage donné en termes de Splines, ou l'inverse, soit facile !

L'Analyse Krigeante s'apparente à une Analyse Fréquentielle, en cela qu'elle sépare dans un phénomène global des composantes d'échelles différentes. Mais dans sa version multivariable, elle est aussi à rapprocher de l'**Analyse des Données**, parce qu'elle vise à mettre en évidence les similitudes et les disparités des variables étudiées simultanément. Mais plutôt que d'une similitude statistique, il s'agit en Analyse Krigeante d'une similitude structurelle, établie au niveau des Variogrammes Croisés. Il faut d'ailleurs noter que la caractérisation, en vue d'une Analyse Krigeante, d'un modèle multivariable global, repose sur une **Analyse en Composantes Principales** réalisée non pas sur les Variables Régionalisées, mais sur les paramètres des fonctions structurales.

Enfin se pose la question du lien de la Géostatistique avec les **Fractales**. Le caractère fractal ou non des Fonctions Aléatoires modélisées en Géostatistique (ou de leurs réalisations) dépend du comportement à l'origine du variogramme. Toutefois, ce paramètre n'est pas caractéristique : ainsi une Fonction Aléatoire gaussienne sur $\mathbb{R}$, de covariance stationnaire exponentielle, sera fractale de dimension $1,5$ (*), alors qu'une Chaîne de Markov à 2 états de même covariance ne le sera pas (dimension 1). Il convient de noter que nous parlons ici de **dimension fractale**, et non pas d'**autohomothétie** : une Fonction Aléatoire stationnaire (non constante...) ne peut être autohomothétique. Par contre, lorsque l'hypothèse de stationnarité stricte est levée (Fonctions Aléatoires Intrinsèques d'ordre k), la Géostatistique permet de modéliser aussi bien des phénomènes fractals (ou non...) que des phénomènes autohomothétiques (ou non...). Cela dit, ces propriétés sont à prendre en compte avec précautions, très exactement au même titre que le comportement à l'origine du variogramme, parce qu'elles ne peuvent être analysées strictement à partir d'échantillons finis et procèdent donc d'une extrapolation vers les échelles infiniment petites. Et, pour la même raison, il n'est pas sûr que le caractère fractal ou non d'une Fonction Aléatoire soit une propriété décisive, au moins en ce qui concerne les Simulations : car, *in fine*, c'est toujours un nombre fini de valeurs que l'on construit...

(*) Dimension de Hausdorff-Besicovitch

Champs d'application

La Géostatistique est avant tout un recueil de **méthodes**, et donc n'est en théorie tributaire d'aucun domaine particulier. C'est un avantage bien sûr, dans la mesure où les algorithmes et la pratique peuvent être testés sur les variables les plus diverses. Mais c'est aussi un handicap : chaque discipline ayant ses formalismes, ses habitudes, son vocabulaire, il est parfois difficile de promouvoir une approche jugée trop généraliste. C'est pourquoi la Géostatistique s'est historiquement appuyée sur quelques domaines privilégiés qui en l'occurrence s'étaient en leur temps montrés plus que d'autres réceptifs à ses nouvelles méthodes.

Domaines porteurs

Historiquement, les premiers développements de la Géostatistique concernent le **monde minier**. Il s'agit d'une véritable symbiose, qui se manifeste même dans le vocabulaire ("effet de pépite", appelé dans d'autres domaines "bruit blanc"...). La situation (peu fréquente...) était favorable, puisque c'était le monde minier lui-même (ou en tous cas certains de ses représentants) qui étaient demandeur de méthodes nouvelles. Le problème était, au milieu des années 50, de nature statistique (correction de biais) ; mais, une fois les contacts pris, il est intéressant de constater que les recherches en Géostatistique ont fini par parcourir toutes les étapes d'un projet minier : depuis la **prospection** jusqu'à l'**optimisation de carrière**, en passant bien sûr par l'**estimation**. Ainsi par exemple, l'élaboration de la Géostatistique Non Linéaire au début des années 70 s'est justifiée initialement uniquement par des problèmes miniers. Pour les non-spécialistes, il ne faut d'ailleurs surtout pas imaginer le monde minier comme un tout monolithique, et même une discipline généraliste comme la Géostatistique doit adopter des approches différentes pour l'or, le nickel, l'uranium ou le charbon. Ainsi, grâce à l'extrême variété des études entreprises, la Géostatistique a pu enrichir considérablement son capital, tant au niveau des applications qu'au niveau des développements théoriques. A l'heure actuelle encore, et en dépit d'un contexte mondial moins favorable, la Géostatistique Minière demeure vivace.

Pourtant, malgré une situation en apparence favorable, il faut relever l'échelle de temps qui a été nécessaire pour faire admettre la Géostatistique comme outil standard. Entre la publication du premier traité de Géostatistique et les premières utilisations en routine de logiciels d'estimation géostatistique, il a fallu compter pas loin d'une vingtaine d'années. C'est un temps à peine inférieur qu'il a fallu pour introduire la Géostatistique dans l'industrie du **pétrole**, actuellement domaine d'application privilégié. Cette fois encore, une des raisons d'une telle durée se situe au niveau de la communication : le monde pétrolier a sa propre approche des problèmes, son propre vocabulaire, bien éloignés de ceux que pratique la Géostatistique dans le contexte minier. Une autre difficulté tient à l'existence, dans les compagnies pétrolières, d'environnements informatiques cohérents et particulièrement développés, auxquels les nouveaux venus géostatistiques doivent savoir s'adapter. Mais, une fois surmontées ces difficultés, le domaine pétrolier s'est révélé extrêmement fructueux, renouvelant la panoplie de problèmes méthodologiques posés à la Géostatistique. Une première particularité des réservoirs pétroliers tient au caractère non stationnaire des structures géologiques rencontrées. Une seconde est l'importance considérable de l'aspect géométrique : discontinuités (failles, contacts), hétérogénéité des terrains, problèmes de séquencement des faciès, problèmes de connexité. Par ailleurs, les données pétrolières de qualité sont en général très coûteuses, donc rares, et il importe de pouvoir travailler sur des modèles numériques fiables. Il a fallu ainsi d'une part développer l'aspect **Géostatistique Non Stationnaire**, en particulier en ce qui concerne les développements de logiciels d'**Analyse Variographique** et d'**estimation**, et d'autre part mettre au point des méthodes ensemblistes pour modéliser les géométries. C'est ainsi que la théorie des **Ensembles Aléatoires**, pourtant ancienne de près de 20 ans, connaît actuellement un notable regain d'applications. Enfin, une étude dans le domaine pétrolier a presque toujours à manipuler des données disparates : rares données aux puits, et multitude de données géophysiques. C'est pourquoi il faut également développer les outils de la Géostatistique Multivariable, en particulier la **Dérive Externe** qui d'ailleurs a été initialement conçue pour l'intégration des données de sismique à la cartographie des réservoirs.

Aperçu sur quelques autres applications

Quelques exemples — non exhaustifs — illustreront le large éventail des études géostatistiques qui ont effectivement été réalisées sur des problèmes réels.

- **Météorologie** : analyse et cartographie du géopotentiel (*). Historiquement, ce problème a été un des tous premiers cas d'application en routine du **Cokrigeage**. Il est à noter que le problème de la cartographie des équipotentielles est compliqué par la contrainte d'avoir des cartes cohérentes pour les différentes valeurs de la pression atmosphérique. Par ailleurs, ces données météorologiques avaient la particularité d'être très hétérogènes, les océans étant évidemment beaucoup moins informés que les continents : la **recherche des voisinages** de cokrigeage a donc été un problème crucial.

  Un point d'histoire : la météorologie est historiquement un des domaines privilégiés du traitement des données à support spatial. En effet, sous le nom d'**Analyse Objective**, **L. S. Gandin** (U.R.S.S.) a proposé en 1965 une méthode d'interpolation équivalente au Krigeage Universel.

- **Estimation Forestière** : ce domaine est le second exemple pour lequel une méthode d'estimation équivalente, mais indépendante de la Géostatistique, ait été proposée (**B. Matérn**, Suède, 1959). Mais par leur caractère discret, les décomptes d'arbres peuvent avantageusement être décrits par des modèles mieux appropriés, comme par exemple les **processus ponctuels**.

- **Agriculture, Pédologie** : à rapprocher de l'estimation forestière, ces domaines bénéficient actuellement de l'apport des données aériennes, et surtout des images satellites. La Géostatistique est ainsi confrontée au problème des **données hétérogènes**, compliqué par la quantité d'information, et aussi au problème du **changement d'échelle**, les différentes variables ayant des supports de plusieurs ordres de grandeurs différents.

- **Science des Matériaux** : la Géostatistique est employée par exemple en **Statistiques des ruptures**, et utilise en particulier les modèles d'**Ensembles Aléatoires**. L'étude géostatistique des lames minces rejoint les techniques de Morphologie Mathématique et d'Analyse d'Image.

- La **Géochimie** a été le domaine privilégié d'application de l'**Analyse Krigeante multivariable**, pour donner un sens précis à la notion d'**anomalie** et de **régionale**.

- En **Géophysique**, les méthodes géostatistiques sont en mesure de répondre à des problèmes importants, comme par exemple la **déconvolution** des valeurs de gravimétrie, ou le **filtrage** des anomalies diurnes du magnétisme. De manière générale, la Géostatistique est une approche originale du célèbre "problème inverse".

- La **Cartographie** ou la **Bathymétrie** (cartographie des fonds marins) constituent naturellement un domaine privilégié de mise en œuvre de la Géostatistique : **interpolation**, **compactage des données**, utilisation des données caractéristiques (sommets, lignes de crête, thalwegs, ...), prise en compte d'**erreurs de mesure** ou d'**incertitudes de localisation**.

- La **Climatologie** peut bénéficier de l'approche de la Géostatistique Multivariable. C'est ainsi que les données de **pluviométrie** peuvent être reliées à la topographie prise comme **dérive externe**, ou à des facteurs plus géométriques comme par exemple la distance aux océans ou l'orientation des vents dominants. Des phénomènes ainsi orientés sont par essence dissymétriques, et requièrent donc des outils d'investigation plus fins que les simples variogrammes (**lois bivariables**).

- La **pêche** constitue un dernier exemple qui ne vient peut-être pas spontanément à l'esprit : l'évaluation des stocks de poissons relève des méthodes d'**estimation globale**, compliquées par les conditions particulières d'échantillonnage, et surtout le fait que les populations étudiées sont mouvantes, de caractéristiques spatiales variant au cours de la journée et des saisons... Notons que l'apport de la Géostatistique a été jugé suffisamment novateur pour qu'une directive du CIEM (**) préconise explicitement l'usage des outils géostatistiques pour l'estimation des stocks de poissons.

Perspectives

- Nous avons évoqué l'utilisation d'images satellites ou de données de lames minces. De manière générale, le **Traitement d'Image** peut être abordé par les outils géostatistiques : **filtrage** d'un bruit, **compactage** de l'information, **analyse fréquentielle**. Il est de fait qu'actuellement, ce type

---

(*) Surface d'iso-valeurs de la pression atmosphérique.

(**) Conseil International pour l'Exploration de la Mer.

de problèmes est plutôt étudié en Traitement du Signal. Par ailleurs, les logiciels courants de Géostatistique ne sont peut-être pas en l'état actuel adaptés à des volumes de données comparables à ceux rencontrés en Traitement d'Image. Par contre, les outils géostatistiques d'**Analyse Variographique fine** devraient apporter un surcroît de précision dans la description des images, les **modèles de changement de support** devraient permettre l'utilisation d'images conjointement avec des échantillons d'autres origines, et l'**Analyse Krigeante** semble un outil adapté pour aider à la compréhension des phénomènes physiques sous-jacents qui décident de l'allure générale d'une image.

- L'étude géostatistique de **modèles numériques** est peu pratiquée actuellement, si ce n'est bien sûr le cas d'école de l'examen *a posteriori* des simulations (conditionnelles ou non). Il est à supposer qu'elle soulèvera des problèmes informatiques — éventuellement dans le sens d'une simplification des algorithmes — liés à la répartition particulière des données dans l'espace. Mais le plus important sera évidemment de prendre en compte les équations qui ont présidé à la fabrication de ces données. Mentionnons une étude récente dans ce sens, qui a abouti à une curiosité : un travail géostatistique sur une **variable complexe**.

  Par ailleurs, si le modèle n'est pas construit sur un espace euclidien, il faudra veiller à y définir une **distance**, puis, au niveau du modèle probabiliste, à donner un sens aux notions de **stationnarité** et **ergodicité**. Cette question n'est pas triviale, et se rencontre déjà dans les problèmes posés à grande échelle à la surface de la Terre (météorologie à l'échelle des hémisphères, par exemple).

- Un autre domaine où il semble y avoir des développements possibles est le **Génie Civil**, et plus précisément l'étude de la **stabilité** ou de la **déformation** de sols sous des ouvrages. Une difficulté théorique majeure à prévoir pour ce type de problèmes est la modélisation des **changements de support**, étant entendu que les variables concernées (par exemple résistance à la compression, à la pénétration, etc...) sont essentiellement **non additives**, et donc ne se prêtent pas à une approche linéaire.

- Les perspectives les plus prometteuses apparaissent aussi dans le domaine de l'**Environnement**. C'est un sujet qui non seulement est très actuel, mais qui *a priori* soulève des questions géostatistiques extrêmement intéressantes. Le trait commun aux données d'environnement, qu'elles soient d'origine naturelle ou humaine (polluants), est que leur répartition spatiale est gouvernée par des équations physiques. C'est ainsi que les phénomènes d'écoulement obéissent à des équations aux dérivées partielles, qui doivent impérativement être prises en compte par le modèle structural, faute de quoi les estimations ou simulations seront dépourvues de vraisemblance. Le traitement de **variables reliées par des équations** est donc au cœur d'une Géostatistique de l'Environnement. Les travaux effectués jusqu'ici dans cette direction, ont montré à quel point le mode d'attaque du problème est tributaire de la forme particulière des équations régissant le phénomène. C'est dire à quel point le champ de recherches est ouvert — sans même aller jusqu'à évoquer des équations qui ne seraient pas linéaires...

  Les équations interviennent comme contrainte de cohérence sur les résultats, mais elles se manifestent aussi au niveau des données. C'est ainsi que l'on peut disposer comme information initiale à la fois d'une variable et de son gradient (ou de son laplacien) qu'il faut intégrer dans un estimateur de **Cokrigeage**. Souvent d'ailleurs, les données de gradient interviennent comme **conditions aux limites**, ce qui accroît l'importance de la **géométrie** du problème. Enfin, si les conditions aux limites concernent les **comportements à l'infini**, le problème se complique encore de questions d'approximation, convergence, troncature...

  Comme développement naturel des questions précédentes, on est amené à modéliser l'évolution dans le temps des structures spatiales. Les **modèles spatio-temporels** ne se bornent pas, tant s'en faut, au rajout d'une dimension : il faut tenir compte de la forme des **équations d'évolution**, et sans doute aussi de **contraintes d'inégalité** qui limitent la vitesse d'évolution. Il s'agit là d'un domaine de recherche entièrement ouvert, et sans doute particulièrement riche.

  Notons enfin que le terme "Environnement" doit être compris dans un sens large, et qu'en particulier la Météorologie devrait pouvoir être de nouveau à la pointe des développements méthodologiques de la Géostatistique.

Trois exemples d'utilisation

Le but de cette brève présentation est seulement d'illustrer, sur deux cas réels, quelques aspects de la conduite d'une étude géostatistique. Le troisième cas, très partiel, est un exemple d'interprétation de fonctions structurales.

Etude de la bathymétrie sur le site du "Titanic"(*)

L'objet de cette étude était de fournir une image du fond sous-marin à proximité immédiate de l'épave du Titanic. Les planches proposées ici veulent attirer l'attention sur les étapes d'une Analyse Variographique réelle, ainsi que sur l'importance du choix du voisinage de Krigeage.

Comme presque toujours pour des données à la mer, les informations disponibles sont réparties sur des **profils** (planche 6). Les données sont très denses sur les profils (ici, une information tous les 50 mètres en général), tandis que les profils sont beaucoup plus espacés. On se trouve d'emblée confronté à une **anisotropie**, non pas de la Variable Régionalisée — on ne dispose actuellement d'aucune information à ce sujet — mais de l'information. A chaque étape donc, il faudra s'assurer que cette situation ne provoque pas d'artefact. En particulier, il faut évidemment à réception des données demander la raison de la maille de reconnaissance : pourquoi cette orientation précisément, pourquoi n'avoir pas fait de quadrillage. Une des conditions essentielles à la bonne marche d'une Analyse Variographique est en effet de disposer d'une **information non préférentielle** : les données doivent pouvoir avoir une signification statistique non biaisée. Cela ne signifie naturellement pas que la Géostatistique ne peut travailler que sur des données aveugles, mais que les informations qualitatives *a priori* doivent être soigneusement prises en compte **avant** tout traitement statistique, soit en délimitant des sous-zones homogènes, soit en incorporant explicitement ces informations au modèle.

En l'occurrence, le choix de la campagne obéissait à des contraintes étrangères à la variable bathymétrique à étudier : il n'y avait donc pas de biais à redouter *a priori*. Cela dit, à réception des données, force est de constater que les données privilégient une direction. En fait, à l'échelle de travail demandée (200 ou 300 mètres), il n'est pas possible d'avoir une information sur la structure de la bathymétrie dans la direction perpendiculaire aux profils. Aussi, nous sommes contraints de nous limiter aux variogrammes dans la direction des profils et, faute d'information supplémentaire, nous sommes **obligés** de chercher un modèle isotrope. En effet, nous de disposons d'aucun élément pour envisager quelque modélisation que ce soit d'une éventuelle anisotropie.

Cette situation est typique en Géostatistique. Faute d'arguments de décision, nous nous rabattons sur le "modèle minimum". Tout autre choix serait arbitraire, et les libertés qu'il laisserait dans le choix de paramètres seraient illusoires. Certes, le "modèle minimum" aussi est arbitraire, mais c'est celui qui fait courir le moins de risques, et celui qui sera le plus facile à corriger en cas de réfutation ultérieure. Pour des raisons de prudence, pour des raisons de réalisme aussi, il est bon d'adopter à chaque étape d'une modélisation un **principe d'économie** : éviter la prolifération d'hypothèses ou de paramètres qui ne peuvent être soumis à aucun contrôle.

Ainsi, les variogrammes sont calculés à une dimension, dans la direction des profils. La planche 7 présente ces variogrammes aux pas 50, 100 et 500 mètres. La dernière figure regroupe ces trois résultats, et confirme leur cohérence. La conclusion qui prévaut est clairement qu'**un modèle stationnaire n'est pas acceptable**. Sur les deux premières figures en effet, on ne voit pas apparaître de palier, ce qui semble indiquer qu'il n'y a pas de stationnarité jusqu'à l'échelle 5000 mètres au moins. Au pas 500 mètres, on pourrait admettre à la rigueur un phénomène de palier vers 20 kilomètres, mais ce palier serait de l'ordre

(*) Exemple proposé avec l'autorisation de TAURUS INTERNATIONAL / TITANIC VENTURES. Pour raison de confidentialité, les coordonnées et les profondeurs ne seront pas mentionnées, non plus que les détails sur la stratégie de la campagne de reconnaissance.

de trois fois la variance des données, ce qui est incompatible avec un modèle stationnaire. Bien sûr, un tel effet de dépassement de la variance pourrait être dû à une forte anisotropie (et alors, le palier dans d'autres directions serait très en-dessous de la variance), mais il n'y a aucun moyen de contrôler ce modèle. Le principe d'économie nous suggère donc de rechercher un modèle non stationnaire, en espérant que s'il y a effectivement une anisotropie, elle sera prise en compte par la dérive. L'Analyse Variographique est donc entreprise en FAI-k, à l'aide du logiciel BLUEPACK.

Une première approche "à l'aveugle", utilisant les options par défaut, conclut à une FAI-1, de Covariance Généralisée linéaire + spline. La carte résultante, obtenue par Krigeage avec des voisinages de 12 points (valeur par défaut) est inacceptable — voir planche 8. La raison en est que rien dans cette estimation ne contraint le programme à utiliser des données provenant de profils différents. Ainsi, pour estimer des valeurs voisines d'un profil donné, le voisinage de Krigeage n'utilise que des données issues de ce profil. D'une part donc, on travaille en FAI-1 avec des données presque alignées — ce qui conduit à une **quasi-singularité du système de Krigeage**. D'autre part, vues du point à estimer, les données ne sont réparties au mieux que dans un angle de 180° — ce qui signifie que l'on est en extrapolation. Paradoxalement, ce n'est que lorsqu'on est assez loin entre deux profils que l'on retrouve des conditions moins inacceptables : le voisinage comporte des informations sur chacun des deux profils, le système est régulier, et les courbes de niveau deviennent plausibles.

Il est possible de contraindre le programme, tant pour l'Analyse Variographique que pour le Krigeage, à utiliser des données appartenant à au moins n profils distincts. En prenant la valeur minimum $n = 3$, le programme diagnostique cette fois une FAI-2, avec une Covariance Généralisée linéaire. Le Krigeage, réalisé avec le voisinage par défaut de 16 points, est présenté sur la planche 9. On note une certaine amélioration, en particulier dans les zones où les profils sont régulièrement espacés et informés. Cependant, il subsiste des anomalies dès que le programme a des difficultés à trouver des informations à la fois bien réparties et issues de 3 profils différents.

L'idée est donc d'augmenter le nombre minimum de profils à utiliser. En passant à $n = 5$, et avec des voisinages de 24 points, le résultat est catastrophique (planche 10). *A posteriori*, on constate que cela est dû à un modèle de Covariance Généralisée trop régulier (cubique). Ainsi, lors du Krigeage, la valeur estimée est extrêmement sensible à la donnée la plus proche, et traduit donc les moindres fluctuations des données. En revanche, lorsqu'on est assez loin de toute donnée, c'est plutôt sur la dérive que s'aligne l'estimateur.

Notons que, dans le cas de la planche 8, il était clair a *priori* que l'on obtiendrait des anomalies — et cette estimation n'a été réalisée que dans un but pédagogique. En revanche, dans le cas de la planche 10, **le résultat catastrophique est un enseignement** : il signifie en quelque sorte que le modèle de FAI-2 est trop instable numériquement compte-tenu des voisinages choisis et compte-tenu de la structure intrinsèque des données. Le programme d'Analyse Variographique "se défend" alors en adoptant une Covariance Généralisée trop régulière qui, jointe à une dérive du second degré, conduit à des estimateurs complètement erratiques.

Il faut tirer un enseignement de cette expérience. A l'évidence, imposer 3 comme minimum du nombre de profils utilisés est insuffisant, selon la planche 9. Par ailleurs, on vient de voir qu'une dérive quadratique déstabilise l'estimation. Par rapport à la planche 10, l'idée est donc d'imposer au programme une dérive linéaire, toutes choses égales par ailleurs. Dans ces conditions, le programme diagnostique un modèle linéaire + cubique (planche 11). Ce modèle est assez voisin, au point de vue régularité et degré de dérive, de celui de la planche 8 ; par contre, il est identique à celui de la planche 10 quant à la structure des voisinages.

Le résultat peut cette fois être considéré comme assez satisfaisant. Sans doute conviendrait-il d'effectuer un faible lissage des lignes de niveau, pour "gommer" les quelques irrégularités dues localement à des voisinages légèrement défectueux. Par contre, la structure transversale que sur les planches 8 et 9 on aurait pu attribuer à des artefacts, et qui était complètement cachée par les parasites sur la planche 10, est ici clairement visible. Il s'agit d'un cañon qui existe réellement, et dont la structure est d'ailleurs connue. La restitution de ce relief a été jugé satisfaisante par le client.

Naturellement, à l'apparition d'une telle hétérogénéité, **il conviendrait de reprendre une Analyse Variographique fine**, qui distinguerait les structures du cañon et du reste du champ. Les données disponibles dans ce cas ne le permettaient pas. Tout au plus, on aurait pu retirer de l'Analyse Variographique les données du cañon, et proposer un modèle structural pour le reste : la restitution de la bathymétrie aurait été un peu meilleure sur la majeure partie de la carte, et vraisemblablement sensiblement moins bonne sur le cañon...

La dernière manipulation que nous proposons sur ce cas d'étude vise à une économie de temps calcul. Nous cherchons l'impact du nombre de points du voisinage, toutes les autres conditions étant identiques à celles de l'estimation précédente. En prenant la valeur par défaut de 8 points, on obtient la carte de la planche 12, qui par rapport à l'estimation précédente n'est qu'assez faiblement dégradée. Or, dans le cas présent, le système de Krigeage n'est que de dimension 11×11, alors qu'il était 27×27 dans le calcul précédent. Pour un travail en routine — ce qui d'ailleurs n'était pas le cas ici — le gain de temps justifierait vraisemblablement la faible dégradation constatée.

Enfin, pour conclure cet exemple, la planche 13 propose la carte des écart-types de Krigeage pour cette dernière estimation. On y distingue évidemment les profils de reconnaissance. On observera également — ce qui était prévisible théoriquement — que le cañon n'est pas discernable sur cette carte : cela est naturel, puisque l'on a adopté un modèle structural global, et qu'alors, l'écart-type de Krigeage ne dépend que de la géométrie de l'information, et non des valeurs des données. Naturellement, une étude plus détaillée ne saurait se contenter de ce résultat...

La leçon à tirer de cet exemple pourtant simple, est la grande difficulté qu'il y aurait à élaborer une procédure purement automatique d'Analyse Variographique et d'estimation. Le dialogue entre le géostatisticien et les données (et celui qui les a fournies...) est bien un élément essentiel d'une étude réelle.

Par ailleurs, le recours à des critères de qualité purement statistiques risquerait de causer bien des désillusions. Ainsi, dans l'étude proposée ici, c'est de la planche 10 que le programme s'accommoderait le mieux : c'est là qu'il a eu le plus de liberté pour ajuster les paramètres au mieux de ses tests statistiques. Et effectivement, la restitution des valeurs à proximité immédiate des données est excellente. Malheureusement, ce qui intéresse l'utilisateur est précisément ce qui se passe ailleurs que près des données ; pour conduire à un résultat intéressant, le modèle doit donc contenir plus d'information qu'il n'en est contenu dans les données. Ce surcroît d'information (ici : choix de la taille du voisinage et degré imposé de la dérive) mesure la nécessaire **responsabilité** que le géostatisticien doit engager dans la conduite d'une étude pratique.

Simulations de réservoirs hétérogènes

Lors de l'exploitation de réservoirs souterrains, aquifères, pétroliers ou de stockage, il est nécessaire de disposer d'une cartographie tri-dimensionnelles des propriétés hydrodynamiques du réservoir (porosités, perméabilités, ...).

Ces variables réelles peuvent être simulées par des outils géostatistiques classiques. Mais le problème se complique quand la répartition de ces variables n'est plus homogène dans l'espace étudié, en particulier quand ces variables varient de plusieurs ordres de grandeur d'une zone à l'autre correspondant à des types de dépôts différents : c'est ce que nous identifions par réservoirs hétérogènes. La procédure de simulation proposée consiste alors à simuler d'abord la géologie du réservoir sous forme de lithofaciès (types de terrains).

La procédure de simulation mise au point au travers du logiciel HERESIM, réalisé conjointement par l'IFP et le Centre de Géostatistique, s'applique essentiellement aux milieux sédimentaires fluvio-deltaïques. La méthodologie repose sur une interprétation géologique du réservoir qui conduit au découpage vertical du réservoir en **unités stratigraphiques** homogènes du point de vue sédimentologique (type et proportion de faciès) et génétique (forme et dimensions des corps gréseux). Chaque unité est caractérisée par sa logique de dépôt, en particulier par une paléo-horizontale (surface considérée comme horizontale à l'époque du dépôt) qui servira de **niveau de référence**, et par sa juxtaposition avec les unités voisines.

Les données d'une telle simulation sont constituées de forages effectués au travers des différentes unités constituant le réservoir et pour lesquelles la succession des lithofaciès est connue soit par carottage (prélèvement et analyse d'échantillons de roche), soit après interprétation des différents logs de sondage disponibles (enregistrements en continu le long du sondage de variables géophysiques). La première étape consiste à délimiter les traversées des différentes unités par les sondages et à étendre ces limites à tout le réservoir, ensuite à découper l'information de chaque puits au pas de la grille de simulation.

A l'intérieur de chaque unité, les lithofaciès sont ordonnés suivant la séquence des dépôts (par exemple du plus argileux au plus gréseux), et les variables étudiées sont les indicatrices des lithofaciès. Une fois l'unité remise à l'horizontale (par rapport à son niveau de référence), plusieurs types d'outils d'investigation à partir des données de puits sont alors disponibles :

- les courbes de proportion verticales qui représentent les proportions des lithofaciès rencontrés dans une tranche horizontale. Ces courbes de proportion verticales résument bien la séquence géologique de l'unité. Ainsi la figure 1 de la planche 14 présente une courbe de proportion verticale qui traduit le remplissage de l'unité par des faciès gréseux majoritaires à la base et qui graduellement laissent la place à de l'argile. Cette répartition verticale est généralement non stationnaire.
- les courbes de proportion horizontales qui représentent les proportions des différents faciès lorsque l'on parcourt le réservoir le long d'une section horizontale quelconque. Horizontalement, l'hypothèse de stationnarité s'avère souvent acceptable : c'est le cas pour l'exemple présenté. Toutefois, suivant l'échelle de travail et dans le cas de dépôts en milieu deltaïque, il peut être légitime de considérer également une non-stationnarité horizontale.
- les variogrammes simples et croisés des indicatrices des différents faciès, calculés expérimentalement soit le long des sondages, soit parallèlement au niveau de référence.

Les courbes de proportion ne se modélisent pas. A l'inverse, les variogrammes expérimentaux doivent être ajustés à l'aide d'un modèle cohérent. En fait, on considère une seule variable aléatoire gaussienne, chaque lithofaciès correspondant à un intervalle de valeurs de cette variable. Les seuils de ces intervalles sont liés aux valeurs expérimentales des proportions des différents lithofaciès et évoluent donc pour chaque tranche horizontale de l'unité comme indiqué sur la courbe de proportion verticale. Ils peuvent aussi varier horizontalement dans le cas de non stationnarité horizontale.

La structure spatiale est ajustée pour chaque unité stratigraphique à l'aide d'un modèle unique ayant un nombre de paramètres faible (un par direction de l'espace) : un exemple d'ajustement est proposé sur la figure 2 de la planche 14. La méthode consiste alors à simuler la variable gaussienne sous-jacente, conditionnée par les puits, et ensuite à convertir le résultat numérique obtenu en un type de faciès grâce aux courbes de proportion.

Un exemple de simulation est montré dans la planche 15 sous la forme d'une coupe verticale où l'on distingue nettement que chacune des unités présente des proportions et des répartitions de faciès différentes : en particulier la seconde unité (en partant du haut) ne contient pratiquement pas d'argile. La section horizontale à la cote où elle a été réalisée, intercepte les quatre unités.

Les variables pétrophysiques (porosité, composantes tensorielles de la perméabilité...) sont alors également simulées ou simplement affectées à chaque cellule constituant le réservoir conditionnellement au type du lithofaciès. Elles peuvent ensuite servir d'entrées à un programme de simulation d'écoulement de fluides.

Cette méthode donne déjà de nombreux résultats satisfaisants. Elle a de plus l'avantage de ne nécessiter que peu de paramètres (principe d'économie déjà cité) et d'être assez souple pour prendre en compte des informations supplémentaires quant à la répartition des lithofaciès par l'intermédiaire des courbes de proportion.

Eléments d'Analyse Structurale du couple Radioactivité-Teneur

Les données étudiées ici ont été obtenues au cours d'une campagne de sondages carottés sur un gisement d'Uranium (*). Sur chaque sondage, on dispose donc simultanément d'une analyse chimique — effectuée sur les carottes — et d'un enregistrement en continu de la radioactivité, réalisé en descendant une sonde dans le forage.

Les carottes apportent sur le gisement une information irremplaçable : en plus de fournir effectivement une valeur de la teneur (c'est-à-dire de la variable d'intérêt), elles permettent au géologue d'avoir une image qualitative de la minéralisation. Mais elles présentent un double inconvénient :

- Elles coûtent cher. Il est donc hors de question, lors de la phase finale de reconnaissance du gisement à maille resserrée, de carotter la totalité des sondages.
- Leur implantation verticale est imprécise, à cause du taux de récupération de la carotte qui n'est jamais de 100%. L'échantillon comporte ainsi des fragments qui ont pu se déplacer à l'intérieur du carottier, et même occasionner un coulissement global des morceaux sains de la carotte.

Les données de radiométrie, au contraire, sont économiques, nombreuses, et localisées avec une grande précision. Aussi, à mesure que la reconnaissance du gisement progresse, le pourcentage de données de

(*) Pour des raisons de confidentialité, les unités ne seront pas mentionnées ici. Nous travaillerons sur des covariances normées, c'est-à-dire sur des fonctions d'auto-corrélation.

radio-carottage devient-il considérable par rapport au pourcentage d'analyses chimiques. Mais bien sûr, la radiométrie présente un inconvénient majeur : il s'agit d'une **mesure indirecte**. Il est clair en effet que personne n'exploite de gisement de radioactivité ... En simplifiant, on peut dire que la variable Radioactivité s'apparente à une **convoluée** de la Teneur, les paramètres de la convolution dépendant essentiellement des conditions du forage (diamètre du trou, présence de boue, tubage) et de la géométrie de la minéralisation. C'est dire qu'il est irréaliste de proposer un modèle déterministe pertinent de cette convolution.

C'est pourquoi une phase préliminaire capitale de l'étude des gisements d'Uranium — et qui a été rodée dans les compagnies minières bien avant l'usage de la Géostatistique — consiste à ajuster un modèle de **régression** Radiométrie–Teneur qui permettra, lorsqu'on ne disposera plus que de données de Radioactivité, de convertir ces données en la variable d'intérêt. L'apport de la Géostatistique, par rapport à ces techniques statistiques, est d'abord la prise en compte de la structure spatiale, tant de la Teneur que de la Radioactivité. Nous nous bornerons ici à l'illustration géostatistique de l'effet de convolution, et seulement dans la direction verticale, mais il est bien entendu que le modèle de corégionalisation permet ultérieurement la mise en œuvre des méthodes de Géostatistique Multivariable, et en particulier du Cokrigeage.

La première figure de la planche 16 montre la covariance normée (la fonction d'**auto-corrélation**), le long des sondages, des données de Teneur. Cette covariance expérimentale est très satisfaisante : elle a été obtenue dans des conditions favorables (données abondantes, information régulièrement espacée), et le résultat fait apparaître une structure claire : un modèle $C(h)$ stationnaire est admissible, de comportement linéaire à l'origine et de portée de l'ordre de 1,20 mètres. Pour être plus précis, il faudrait observer que le variogramme, qui lui est associé par la formule classique $\gamma(h) = C(0) - C(h)$, **n'atteint pas** la valeur $C(0)$ pour $h = 1,2m$. Cet effet est dû à une anisotropie, les teneurs étant plus erratiques dans le plan horizontal que le long de la verticale — mais cette discussion concerne la modélisation tridimensionnelle, exclue de cette présentation.

La seconde figure montre la fonction d'auto-corrélation pour la Radioactivité. Afin de permettre la comparaison, on a fait également figurer en pointillés la courbe précédente. Cette fois encore, on observe une structure verticale prononcée, mais il y a d'importantes différences entre les deux courbes :

- Bien que selon toute vraisemblance un modèle stationnaire soit cette fois encore admissible, on note que la portée n'est pas encore atteinte à l'échelle de 3 mètres. A l'échelle de quelques mètres, la Radioactivité présente ainsi une structure plus marquée que la Teneur : la "zone d'influence" d'un échantillon est plus grande pour la Radioactivité que pour la Teneur.
- La courbe fait apparaître une concavité à l'origine, qui pourrait être modélisée par un comportement parabolique — alors que l'auto-corrélation des teneurs avait un comportement linéaire. C'est une illustration expérimentale de la relation de **régularisation** qui existe entre les deux variables. La Radioactivité présente aux petites distances une plus grande régularité que la Teneur : inutile de préciser qu'il s'agit là d'une vérité d'expérience connue de tous les praticiens de l'Uranium ...
- Entre 0 et 2,5 mètres, l'auto-corrélation de la Radioactivité est toujours plus forte que celle de la Teneur. L'effet de plus grande régularité de la Radioactivité s'observe à toutes ces distances.

Pour mémoire, nous présentons sur la troisième figure de la planche 16 le Variogramme Croisé expérimental (vertical) entre Radioactivité et Teneur. On peut penser *a priori* qu'il s'agit là d'un outil assez peu adapté de reconnaissance structurale. Le Variogramme Croisé est en effet par construction symétrique, c'est-à-dire que dans une direction donnée, il ne tient pas compte de l'**ordre** des variables. Or, dans les gisements d'Uranium, il peut y avoir des phénomènes de lixiviation (*), qui impliquent un déplacement systématique de l'Uranium $U(x)$ par rapport à ses traceurs radioactifs $Ra(x)$. Ainsi, la loi de $U(x)$ sachant $Ra(x+h)$ — classiquement notée $U(x) \mid Ra(x+h)$ — risque d'être notablement différente de celle de $U(x+h) \mid Ra(x)$, ce que ne "verra" pas le Variogramme Croisé. C'est pourquoi il semble opportun de calculer plutôt la **Covariance Croisée** expérimentale C_{URa}, dont la version théorique est définie par :

$$C_{URa}(x,y) \quad = \quad \mathbf{E}\left[U(x).Ra(y)\right]$$

ou, en modèle stationnaire comme c'est le cas ici :

$$C_{URa}(h) \quad = \quad \mathbf{E}\left[U(x).Ra(x+h)\right]$$

fonction qui, elle, n'est pas nécessairement symétrique en h.

(*) Dissolution du minerai, suivie d'un écoulement de la solution.

Notons au passage la formule reliant Variogramme Croisé et Covariance Croisée :

$$\gamma_{URa}(h) \quad = \quad C_{URa}(0) - \frac{C_{URa}(+h) + C_{URa}(-h)}{2}$$

Cette expression prouve que la connaissance de la Covariance Croisée implique la connaissance du Variogramme Croisé, alors que la réciproque est fausse : ce qui confirme que la Covariance Croisée est un outil d'analyse plus fin que le Variogramme Croisé.

En l'occurrence, une première constatation s'impose (quatrième figure planche 16) : la Covariance Croisée est très proche d'être symétrique par rapport à l'axe $h = 0$. Cela signifie qu'il n'y a pas, à l'échelle du gisement, de décalage systématique notable entre les mesures de Teneur et de Radioactivité, et ceci quelle qu'ait pu être la nature de ce décalage : il n'y a pas d'erreur **systématique** de mesure de position des carottes par rapport aux enregistrements de radioactivité, et il n'y a pas eu non plus depuis la formation du gisement de migration verticale de l'une des variables par rapport à l'autre. On est ainsi en mesure, par un contrôle purement numérique, d'énoncer ou de confirmer une conclusion de nature géologique, et de rassurer les sondeurs sur la qualité de leur travail.

Dans le cas du gisement étudié, il se trouve que cette première interprétation de la Covariance Croisée n'était pas une surprise pour les géologues. Par contre, cette même courbe présente un caractère beaucoup plus inattendu.

On observe en effet que la Covariance Croisée ne présente pas son maximum en $h = 0$. En fait, la corrélation entre deux mesures, l'une de Radioactivité et l'autre de Teneur, est plutôt meilleure lorsque ces mesures sont déclarées distantes verticalement de $\pm 10cm$ que lorsqu'elles sont réputées être implantées au même endroit. Cela signifie une incertitude de positionnement relatif des mesures de Teneur et de Radiométrie, de l'ordre précisément de $\pm 10cm$, et non systématique. Autrement dit, lorsque deux pics de radioactivité ne sont distants que de quelques centimètres, il est illusoire de vouloir attribuer à l'un plutôt qu'à l'autre un pic de teneurs isolé qui leur est vis-à-vis. Et de même, il est illusoire de vouloir "corréler horizontalement", entre des sondages distants de plusieurs dizaines de mètres, des passes minéralisées repérées sur les carottes, compte-tenu de l'imprécision sur la profondeur de ces carottes. Et, cette fois, ces conclusions étaient beaucoup moins prévues par les géologues ...

Confirmées par la suite du travail sur le gisement, ces remarques — entre autres — ont illustré de façon spectaculaire combien il serait erroné de voir dans le travail géostatistique un simple jeu d'esthète, et combien est important le dialogue entre spécialistes de compétences différentes.

Diffusion de la méthode

Il n'est pas aisé de dresser un tableau, même partiel, de l'extension des méthodes géostatistiques. Une première raison en est le caractère fondamental, pluridisciplinaire, de la Géostatistique, qui donc peut être utilisée dans les endroits les plus inattendus. Une seconde raison en est le danger qu'il y aurait à définir d'une façon dogmatique les frontières d'une Géostatistique "orthodoxe", distribuant les bons et mauvais points, accordant ou non le label "géostatistique" à telle ou telle méthode.

Il n'existe pas d'annuaire mondial (ni même français...) des géostatisticiens ou des Centres de Recherche en Géostatistique, et l'élaboration d'un tel document excéderait de beaucoup les ambitions de la présentation actuelle. C'est pourquoi il n'est possible de donner ici que quelques indications très fragmentaires, à titre de simple illustration.

Implantation des activités géostatistiques

Née de la rencontre entre les problèmes miniers posés par les gisements d'or d'Afrique du Sud et les formalismes probabilistes développés par les Ecoles française et soviétique, c'est initialement dans ces trois pays que la Géostatistique a connu ses premiers progrès. Au début des années 70, favorisée par des moyens de calcul incomparables, une Ecole américaine s'est développée à son tour.

A l'heure actuelle, nous disposons de peu d'information sur l'implantation de la Géostatistique dans l'ex-URSS, dont une caractéristique est la séparation nette qui existe entre les approches universitaire et industrielle. Quant aux développements géostatistiques en Afrique du Sud, ils sont surtout de nature industrielle, et sont le fait des entreprises minières. Au niveau Enseignement et Recherche, on peut donc *grosso modo* distinguer deux pôles principaux, européen et nord-américain. Il convient cependant de noter qu'un important pourcentage d'enseignants de Géostatistique aux USA ont reçu leur formation initiale en Europe.

Au niveau européen se trouve le berceau de la Géostatistique, le

- Centre de Géostatistique de l'ENSMP, qui forme des spécialistes depuis maintenant plus de 25 ans, et où sont passés et ont souvent longtemps séjourné la plupart des enseignants *seniors* de tous pays. En plus de ses activités de Recherche — à l'origine entre autre de la Géostatistique Non Linéaire, de la Géostatistique Non Stationnaire et des Ensemble Aléatoires — le Centre dispense un Enseignement de $2^{ème}$ et $3^{ème}$ Cycles (DEA et thèses), un Cycle Spécialisé plutôt destiné aux étudiants étrangers, et des enseignements "à la carte" à l'intention des industriels. Par ailleurs, les chercheurs du Centre effectuent de très fréquentes missions de formation, dans les entreprises ou les universités du monde entier. Le Centre enfin organise des Ecoles d'Eté d'initiation à la Géostatistique, ainsi que des journées de Géostatistique, davantage axées sur les développements nouveaux.

Ces différentes formes d'activités se retrouvent dans tous les organismes proposant un Enseignement de Géostatistique. A titre d'exemple, et toujours en Europe, citons :

- L'INPL, Institut National Polytechnique de Lorraine,
- L'Institut Polytechnique de Zürich,
- L'ETSECCP, Ecole Technique Supérieure de Chemins, Canaux et Ports de Barcelone (Université de Catalogne)
- Les Universités de Leeds, Rome, Lisbonne, ...

Il existe également un enseignement de Géostatistique dans le cadre du programme ERASMUS de la CEE. Par ailleurs, certain programmes de la CEE peuvent demander une formation géostatistique spécifique : c'est le cas par exemple dans le domaine de la pêche, où une directive européenne préconise explicitement l'estimation par les méthodes géostatistiques. Quant aux organismes de Recherche, ou aux laboratoires de Recherche des entreprises, qui travaillent **avec** ou **sur** la Géostatistique, ils sont nombreux même en se limitant à la France. Citons par exemple :

- le BRGM à Orléans,
- l'Institut Français du Pétrole,

- l'INRA à Avignon,
- la COGEMA,
- l'IRSID,
- l'IFREMER,
- les compagnies pétrolières (CFP, SNEA)
- l'ORSTOM ...

sans oublier les Universités : Nancy, Grenoble, Paris VI ...

Au niveau industriel, l'emploi systématique des méthodes géostatistiques se retrouve historiquement d'abord dans les compagnies minières : Pennaroya, COGEMA (pour l'anecdote, c'est au CEA qu'a été forgé le néologisme "Géostatistique"). Avec les développements de la Géostatistique Non Stationnaire et des Ensemble Aléatoires, ce sont les compagnies pétrolières qui ont assimilé les algorithmes géostatistiques dans leurs logiciels standard. Dans les autres branches industrielles, l'implantation en routine de la Géostatistique se fait souvent au coup par coup — l'espoir étant actuellement de suivre dans le domaine de l'Environnement une évolution semblable à celle qui a été connue en Mines et Pétrole. Mentionnons enfin l'existence d'une société, GEOVARIANCES, entièrement vouée à la prestation de service géostatistique.

La situation est similaire en ce qui concerne l'implantation industrielle de la Géostatistique en Amérique du Nord : compagnies minières et surtout pétrolières. Dans certains cas d'ailleurs, cette implantation est d'origine européenne : le programme BLUEPACK du Centre de Géostatistique connaît un beau succès dans le monde pétrolier. En matière d'Enseignement et de Recherche, parmi les probablement nombreuses Universités qui proposent de la Géostatistique, citons

- l'Université de Stanford (Californie),
- l'Université de Tucson (Arizona),
- l'Université de Lawrence (Kansas),
- Iowa State University ...

ainsi que, au Canada,

- l'Ecole Polytechnique de Montréal.

Dans le reste du monde, on peut noter une importante utilisation de la Géostatistique dans des pays de forte tradition minière : Afrique du Sud (or et pierres précieuses, charbon, platine), Australie (charbon), Afrique francophone (uranium, phosphate, manganèse,... mais aussi pétrole et forêts), Amérique du Sud : Pérou & Chili (cuivre), Brésil (pierres semi-précieuses, amiante, pétrole ...), Vénézuela (pétrole). Cette approche industrielle s'accompagne parfois d'activités d'Enseignement et de Recherche :

- GEOVAL en Australie (Sydney, Perth),
- Université de Santiago au Chili,
- Ecole des Mines d'Ouro Preto au Brésil , ...

et l'on sait qu'il existe un enseignement de la Géostatistique en Chine, un projet de création d'un Centre de Recherche en Inde, etc...

Communication

A l'initiative du Centre de Géostatistique, un premier **Congrès International de Géostatistique** s'est tenu en 1975 près de Rome. Après quelques années, l'intérêt s'est fait sentir de systématiser cette forme de communication et, au second Congrès (Lake Tahoe, Californie, 1984), il a été adopté une périodicité de quatre ans. Au Congrès d'Avignon en 1988 a été créée **IGeostA**, Association Internationale de Géostatistique, qui a vocation d'assurer la communication au sein de la communauté géostatistique. Une publication, *De Geostatisticis*, fait office de bulletin d'information entre les membres de l'Association. Il existe par ailleurs un bulletin de même nature, *Geostatistics*, en Amérique du Nord.

Il n'existe pas actuellement de revue de Géostatistique. Les communications orientées vers les Sciences de la Terre sont le plus souvent publiées dans *Mathematical Geology*, fascicule édité par le Kansas Geological Survey. Mais, les domaines d'application de la Géostatistique étant des plus variés, il est de fait que les publications peuvent être extrêmement dispersées : revues de Statistiques ou de Probabilités pour les recherches théoriques, de Stéréologie pour certains travaux sur les Ensemble Aléatoires, de Géophysique pour la Géostatistique pétrolière, etc...

Logiciels disponibles

De même que pour la diffusion des méthodes, il est difficile de dresser un inventaire des logiciels propres à la Géostatistique, ou l'utilisant partiellement. C'est pourquoi cette liste n'a pour propos que de donner quelques indications, parcellaires même au niveau seulement français.

Initiateur de l'utilisation de la Géostatistique en contexte industriel, le Centre de Géostatistique de l'ENSMP a dès le milieu des années 70 élaboré une bibliothèque de logiciels proprement géostatistiques, qui à l'heure actuelle s'articulent en deux corps de programmes :

- GEOSMINE, orienté vers la Géostatistique minière et non linéaire : Simulation de gisements, Krigeage Disjonctif, courbes de sélectivité.
- BLUEPACK, spécialisé dans la Géostatistique Non stationnaire : Analyse Variographique en FAI-k, cartographie, Simulations. Les études de structures pétrolières relèvent typiquement de BLUEPACK.

Ces programmes s'appuient sur une structure de base de données, également élaborée par le Centre de Géostatistique. Ils sont commercialisés conjointement par l'Ecole des Mines et la société GEOVARIANCES, qui de son côté commercialise

- OPMINE, logiciel de simulation d'exploitation à ciel ouvert.

Par ailleurs, conjointement avec l'IFP (Institut Français du Pétrole), l'Ecole des Mines a conçu et diffuse le programme

- HERESIM, logiciel de simulation conditionnelle de réservoirs hétérogènes, qui vise à modéliser les structures stratigraphiques complexes rencontrées en exploration pétrolière.

Le BRGM (Bureau de Recherches Géologiques et Minières) propose également un corps de programmes,

- GDM, orienté surtout vers la Géostatistique minière, mais incluant aussi des fonctions sortant du domaine de la Géostatistique. Comme GEOSMINE, ce programme s'appuie sur une structure propre de base de données.

La COGEMA, de son côté, a créé

- SERMINE, système de programmes d'estimation minière.

Avec la collaboration du Centre de Géostatistique, la SNEA(P) a construit et commercialisé le programme

- KRIGEPACK, orienté vers l'Analyse Variographique automatique et la cartographie par Krigeage de Variables Régionalisées non stationnaires.

A l'étranger, mentionnons :

- GEO-EAS, programme interactif d'Analyse Variographique et de Krigeage, proposé par l'EPA (Environment Protection Association) aux USA. Notons que ce programme est du domaine public.
- REGARDS, distribué par Trinity College à DUBLIN. Il s'agit d'un programme de "Géostatistique exploratoire", entièrement orienté vers l'interactivité et permettant des Analyses Structurales d'une grande précision.

Il est beaucoup plus difficile de faire un inventaire de logiciels pour lesquels la Géostatistique n'est qu'une composante, parfois minime. Ainsi, les programmes de la Météorologie Nationale utilisent des algorithmes dont certains — initialement installés avec la collaboration du Centre de Géostatistique — relèvent du Cokrigeage. Par ailleurs, si l'on sait que certaines Universités américaines (Arizona, par exemple) produisent et distribuent des programmes incluant de la Géostatistique, il arrive aussi de découvrir au hasard de déplacements des logiciels de Géostatistique dans des endroits plus inattendus

— par exemple des Universités peu importantes de l'ex-URSS. Enfin, la production de logiciels utilisant la Géostatistique peut être le fait d'entreprises industrielles (exemple : FLUOR aux USA).

A titre d'exemple de logiciels incluant de la Géostatistique, on peut citer :

- GEOCAD : programme de modélisation utilisant surtout les splines discrètes, mis au point conjointement par l'INPG de NANCY et l'Université de STANFORD.
- CARTOLAB, programme de cartographie automatique distribué par l'INPG et GEOVARIANCES.
- En Grande-Bretagne, un famille de logiciels graphiques utilisant accessoirement la Géostatistique est distribuée par la société GEOSTOKOS.
- Au Danemark, deux logiciels graphiques : UNIRAS et IRAP, contiennent une estimation par Krigeage.
- De même, le progiciel américain SAS propose explicitement le Krigeage.

La liste bien sûr est loin d'être exhaustive. On sait par exemple, par un retour de l'histoire, que des logiciels géostatistiques d'estimation minière sont écrits par des compagnies Sud-Africaines (pierres ou métaux précieux). Par ailleurs, avec un soupçon d'emphase, on pourrait noter que beaucoup font de la Géostatistique sans le savoir puisque la plupart des logiciels statistiques contiennent des méthodes de régression locale à base de splines cubiques, ce qui mathématiquement revient à un Krigeage en FAI-k — dépourvu il est vrai de la phase essentielle de l'Analyse Variographique...

Bibliographie

Cette liste comprend un nombre limité de références, choisies le plus souvent en raison de leur caractère synthétique, et privilégiant les documents de langue française. Les publications de l'Ecole de Mines de Paris (ENSMP) sont disponibles auprès de la bibliothèque du Centre de Géostatistique.

Références historiques

Matérn B. [1960] : Spatial Variation, in *Almaenna Fœrlaget* — Stockolm. Ouvrage réédité en [1986] par Springer Verlag, Berlin

Gandin L.S. [1965] : Objective analysis of meteorogical fields, *Israel program for scientific translations* — Jerusalem
(Deux textes qui introduisent le traitement des "Variables Régionalisées, respectivement dans le domaine forestier et en météorologie. L'Analyse Objective conduit à un formalisme identique au Krigeage Universel.)

Matheron G. [1963] : Traité de Géostatistique Appliquée — Editions du B.R.G.M., Paris
(Ouvrage de Géostatistique Linéaire, orienté vers les applications minières.)

Matheron G. [1965] : Les Variables Régionalisées et leur estimation — Masson, Paris
(Les bases mathématiques de la Géostatistique Linéaire, dans le cadre stationnaire ou intrinsèque. Une part importante est consacrée au formalisme transitif)

Matheron G. [1975] : Random sets and integral geometry — J. Wiley & Sons, New York
(A la frontière de la Morphologie Mathématique et de la Géostatistique. La création de la théorie des Ensembles Aléatoires.)

Ouvrages de présentation générale

David M. [1977] : Geostatistical ore reserve estimation — Elsevier Scientific Publishing Co., Amsterdam, Oxford, New York

Rendu J.M. [1978] : An Introduction to Geostatistical Methods of Mineral Evaluation, *South African Institute of Mining and Metallurgy Monograph Series* — Johannesburg
(Deux livres orientés vers la Géostatistique minière.)

Cressie N. [1991] : Statistics for spatial data — J. Wiley & Sons, New York
(L'état de l'art au début des années 90, présenté du point de vue statisticien. L'ouvrage aborde également le domaine de la Morphologie Mathématique, des Champs Markoviens, etc...)

Manuels de Géostatistique Linéaire

Matheron G. [1970] : La théorie des Variables Régionalisées et ses applications — ENSMP, Paris
(Existe aussi en version anglaise. Cours de base de la Géostatistique Linéaire jusqu'au Krigeage Universel inclus.)

Chauvet P. [1991] : Aide-mémoire de Géostatistique Linéaire — ENSMP, Paris
(Inclut la présentation des FAI-k.)

Variographie

Journel A. [1977] : Géostatistique Minière *(Thèse)* — ENSMP, Paris

Journel A. & Huijbregts Ch. [1978] : Mining Geostatistics — Academic Press, London
(L'ouvrage est la version anglaise de la thèse. Présentation générale de la Géostatistique, très orientée vers l'utilisation minière. Se caractérise par un volume important d'exemples pratiques, en particulier concernant l'Analyse Variographique.)

Variance d'extension

Formery Ph. & Matheron G. [1962] : Recherche d'optimum dans la reconnaissance et la mise en exploitation des gisements miniers, in *Annales des Mines, Novembre 1962* — Paris

Krigeage

Rivoirard J. [1984] : Le comportement des poids de Krigeage *(Thèse)* — ENSMP, Paris
(Analyse fine des pondérateurs du Krigeage en hypothèse stationnaire ou intrinsèque.)

Chauvet P. [1988] : Réflexions sur les pondérateurs négatifs du Krigeage, in *Sciences de la Terre* — Nancy
(Commentaires sur le lien entre modèle probabiliste et cartes résultantes, à partir d'un problème qui trouble souvent les utilisateurs : les pondérateurs négatifs.)

Géostatistique Non Stationnaire

Delfiner P. & Matheron G. [1980] : Les fonctions aléatoires intrinsèques d'ordre k — ENSMP, Paris
(Un cours d'introduction au formalisme des FAI-k, orienté vers la mise en pratique dans des logiciels d'Analyse Variographique automatique.)

Chilès J.P. [1977] : Géostatistique des phénomènes non stationnaires (dans le plan) *(Thèse)* — ENSMP, Paris
(Synthèse du traitement dans le plan de données non stationnaires, complétée par un examen des méthodes de simulation.)

Multivariable

(Pour le cokrigeage, voir le manuel de Géostatistique Linéaire de 1970)

Delfiner P., Delhomme J.P. & Pélissier-Combescure J. [1983] : Application of geostatistical analysis to the evaluation of petroleum reservoir with well logs, *Annual Logging Symposium of the SPWLA* — Calgary
(Introduction de la méthode de la Dérive Externe.)

Wackernagel H. [1985] : L'inférence d'un modèle linéaire en Géostatistique Multivariable *(Thèse)* — ENSMP, Paris
(Introduction à l'Analyse Krigeante.)

Dong A. [1990] : Estimation Géostatistique des phénomènes régis par des équations aux dérivées partielles *(Thèse)* — ENSMP, Paris

Daly C. [1991] : Applications de la Géostatistique à quelques problèmes de filtrage *(Thèse)* — ENSMP, Paris
(Deux thèses dont les titres parlent d'eux-mêmes...)

Simulations

Freulon X. [1992] : Conditionnement du modèle gaussien par des inégalités ou des randomisées *(Thèse)* — ENSMP, Paris
(Rappels commentés de la méthode des Bandes Tournantes, et développements sur la simulation de structures géométriques et les simulations sous contraintes.)

Géostatistique Non Linéaire

Matheron G. [1976] : A simple substitute for conditional expectation : the disjunctive kriging, in *Proceedings NATO ASI "Advanced Geostatistics in the Mining Industry", Octobre 1975* — Reidel, Dordrecht
(Le texte fondateur du Krigeage Disjonctif.)

Rivoirard J. [1991] : Introduction au Krigeage Disjonctif et à la Géostatistique Non Linéaire — ENSMP, Paris
(Un cours de synthèse sur la Géostatistique Non Linéaire.)

Méthodes diverses

Séguret S. [1991] : Géostatistique des phénomènes à tendance périodique (dans l'espace-temps) *(Thèse)* — ENSMP, Paris
(Comparaison au Traitement du Signal et aux techniques de Filtrage, appliquée à la Géophysique marine.)

Boulanger F. [1990] : Modélisation et Simulation des Variables Régionalisées par des Fonctions Aléatoires Stables *(Thèse)* — ENSMP, Paris
(Généralisation des modèles de simulation à un cadre non-gaussien, utilisant les méthodes ARMA des Séries Chronologiques.)

Dubrule O. [1981] : Krigeage et Splines en cartographie *(Thèse)* — ENSMP, Paris
(Le Krigeage considéré comme un interpolateur, et l'équivalence Splines-Krigeage.)

Langlais V. [1990] : Estimation sous contraintes d'inégalités *(Thèse)* — ENSMP, Paris
(Pour un Krigeage sous contrainte, on pourra également se référer à l'article de R.J. Barnes & T.B. Johnson *: "Positive Kriging", dans les actes du Congrès de 1984 cités ci-dessous.)*

Jeulin D. [1991] : Modèles morphologiques de structures aléatoires et de changement d'échelle *(Thèse de Docteur ès Sciences Physiques)* — Université de Caen
(Une approche synthétique des modèles d'Ensembles Aléatoires.)

Actes des Congrès Internationaux de Géostatistique

Guarascio M., David M. & Huijbregts Ch. ed. [1976] : Advanced Geostatistics in the Mining Industry — Reidel, Dordrecht

Verly G., David M, Journel A. & Maréchal A. ed. [1984] : Geostatistics for Natural Resources Characterization — Reidel, Dordrecht

Armstrong M. ed. [1989] : Geostatistics — Kluwer Academic Publishers, Dordrecht

Les actes des trois premiers Congrès Internationaux de Géostatistique (le quatrième s'est tenu en septembre 1992) contiennent aussi bien des innovations théoriques que des exemples d'applications. A ces trois documents, il convient d'ajouter

Armstrong M. & Matheron G. [1987] : Geostatistical case studies — Reidel, Dordrecht
(Recueil d'exemples particulièrement représentatifs d'applications.)

Notons également la présence de communications de Géostatistique dans les actes des symposium "Application of Computers and Operation Research in the Mining Industry" (APCOM).

Voir la Géostatistique avec du recul...

Mentionnons pour l'anecdote l'existence depuis une dizaine d'années d'une entrée "Géostatistique" dans le Petit Larousse... Mais plus instructif est l'article de l'*Encyclopedia Universalis*, qui met l'accent principalement sur la Géostatistique Non Linéaire. Enfin, pour prendre du recul par rapport à l'ensemble de la Géostatistique, citons pour conclure

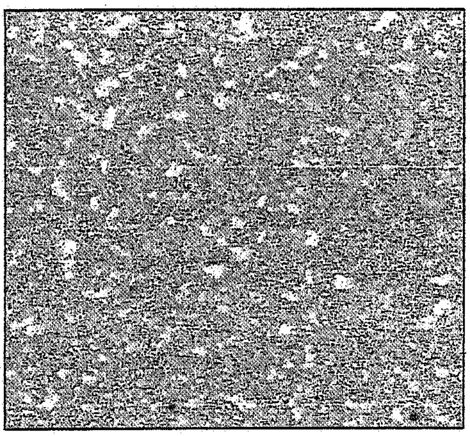
Matheron G. [1978] : Estimer et Choisir — ENSMP, Paris

ouvrage traduit en anglais :

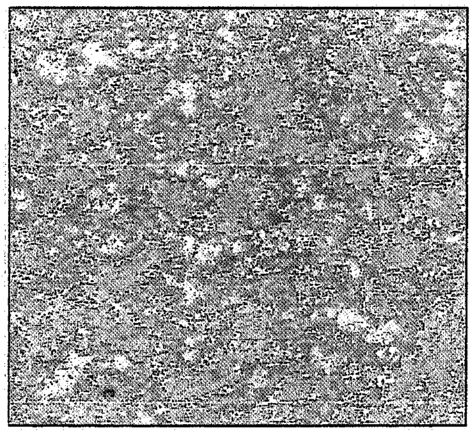
Matheron G. [1989] : Estimating and Choosing — Springer Verlag, Berlin

Ce texte propose un regard synthétique sur la Géostatistique et sur la signification des opérations classiques faites en routine en cours d'étude. Les notions capitales de stationnarité et d'ergodicité sont minutieusement analysées, ainsi que les problèmes essentiels de l'adéquation d'un modèle à la réalité et de la place réelle des probabilités dans la démarche Géostatistique.

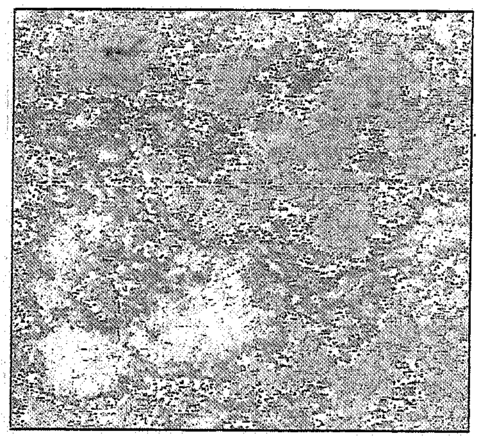
INFLUENCE DE LA PORTÉE



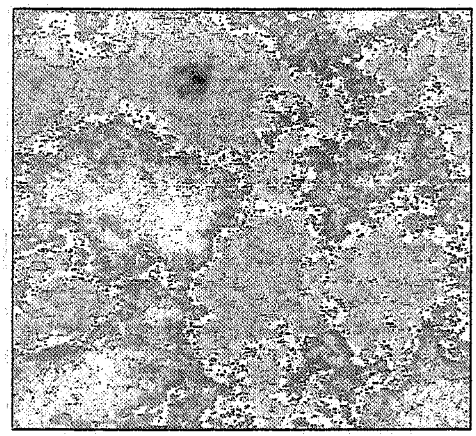
Portée 10



Portée 25



Portée 50

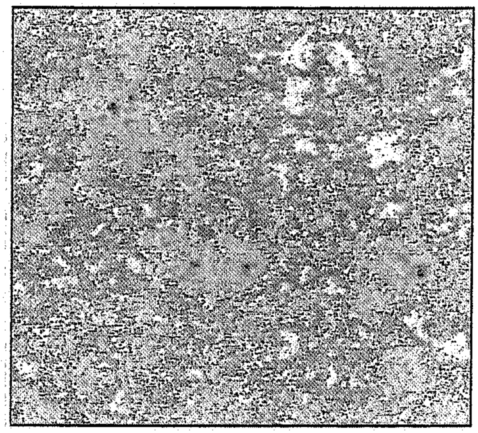


Portée 75

Planche 1

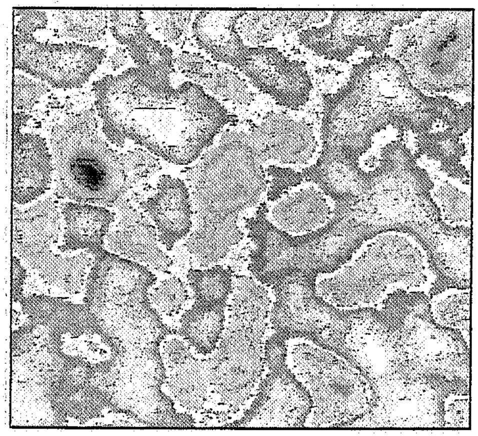
INFLUENCE DU MODÈLE

Sphérique (portée 40)



Exponentiel (portée 13.33)

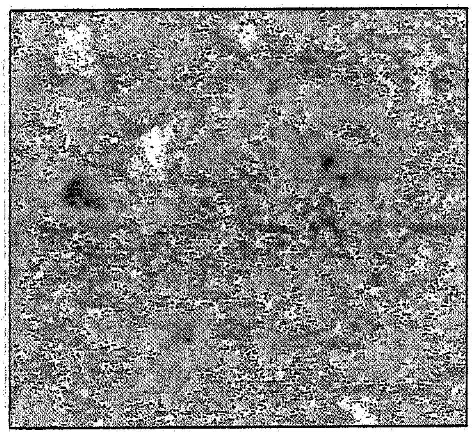
Gaussien (portée 23.12)



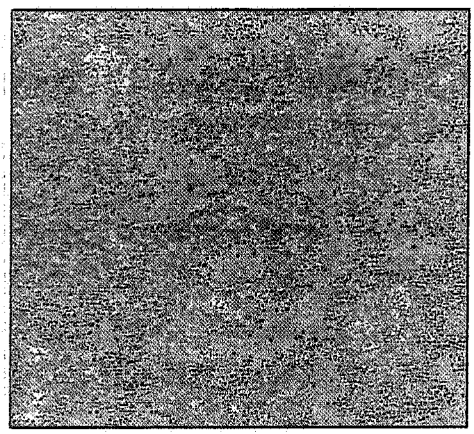
Cubique (portée 40)

Planche 2

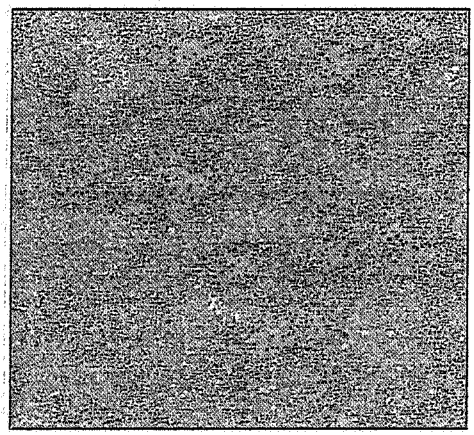
INFLUENCE DE L'EFFET DE PÉPITE



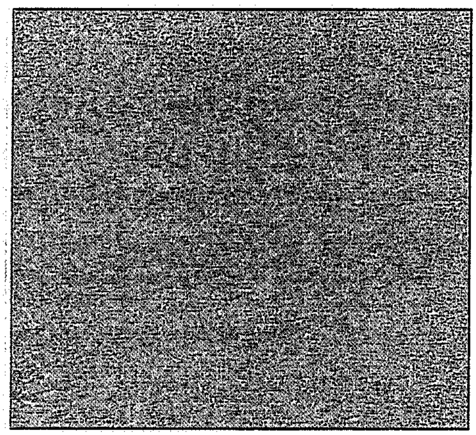
100% sphérique



2/3 sphérique – 1/3 pépite



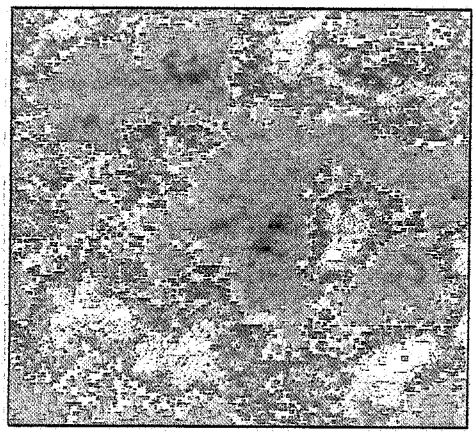
1/3 sphérique – 2/3 pépite



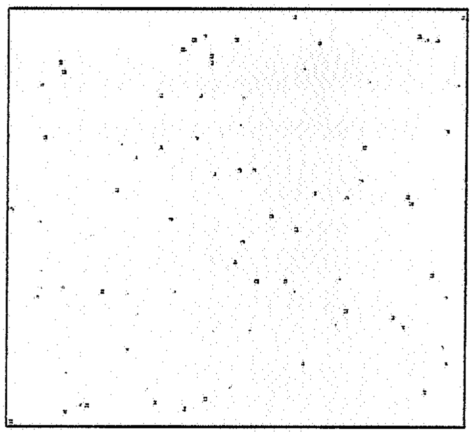
100% pépite

Planche 3

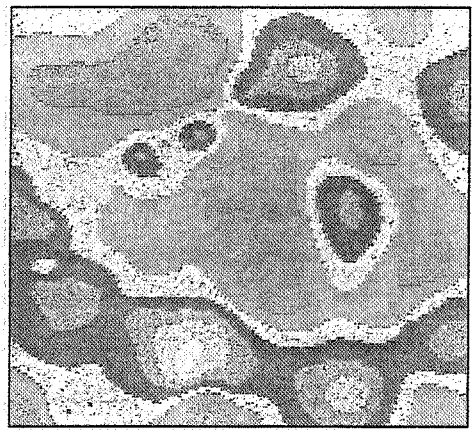
COMPARAISON RÉALITÉ–KRIGEAGE : MODÈLE SPHÉRIQUE



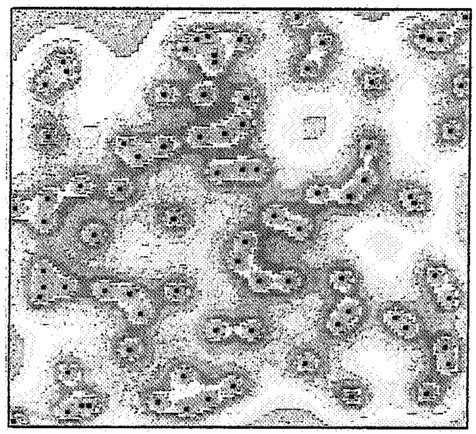
Réalité



Echantillonnage

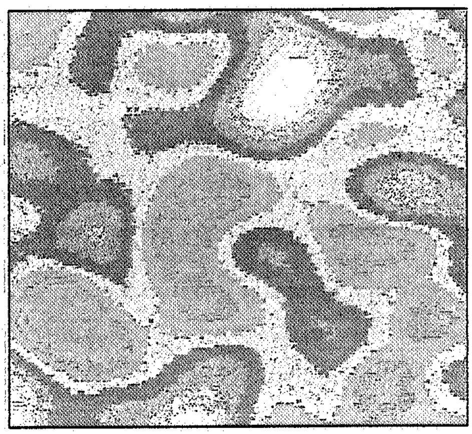


Krigeage

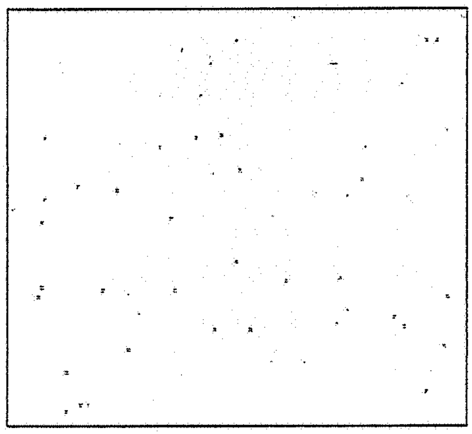


Ecart-type de Krigeage

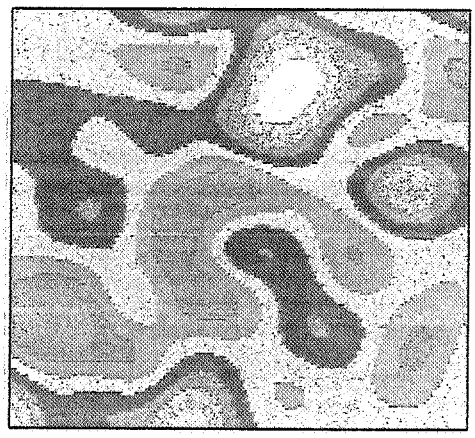
COMPARAISON RÉALITÉ–KRIGEAGE : MODÈLE GAUSSIEN



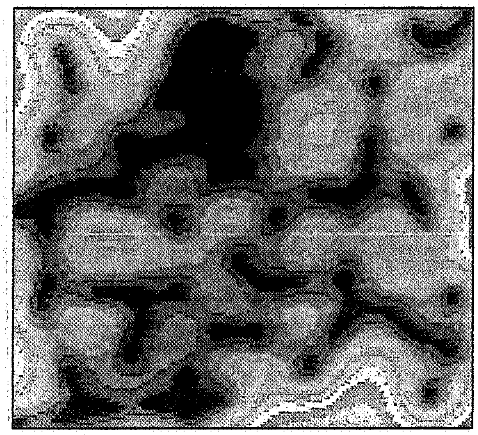
Réalité



Echantillonnage



Krigeage



Ecart-type de Krigeage

"TITANIC" : INFORMATION DISPONIBLE

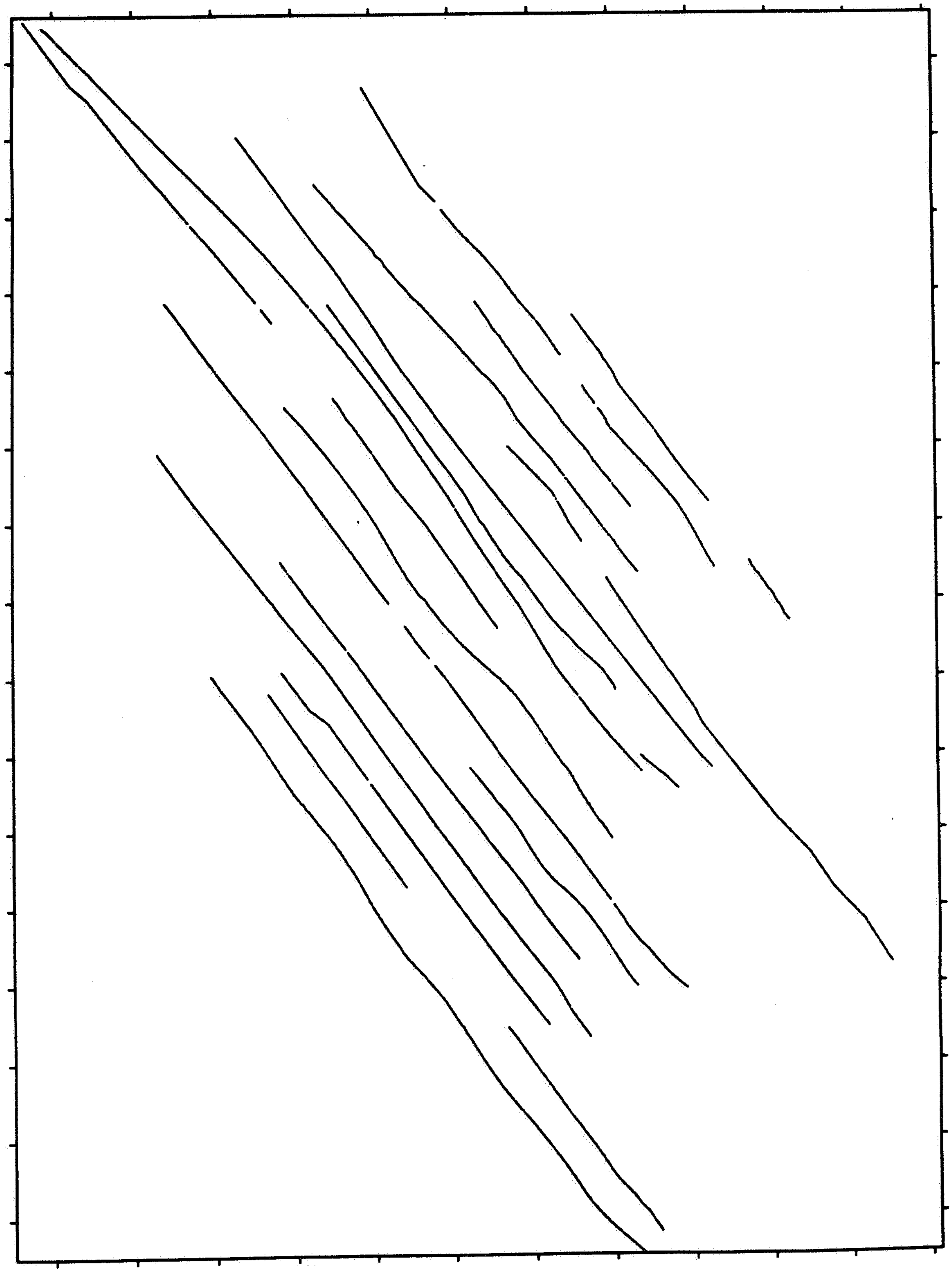
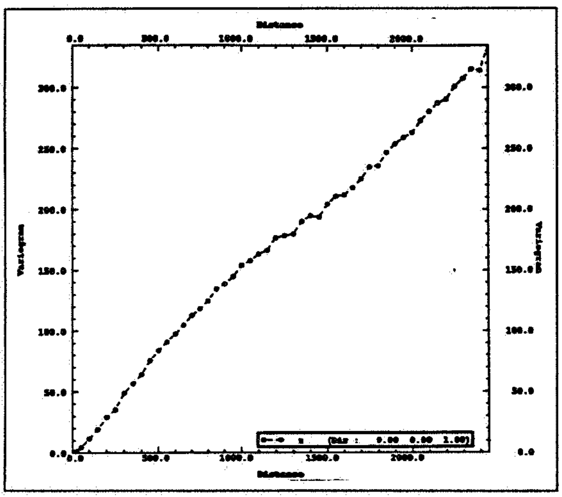
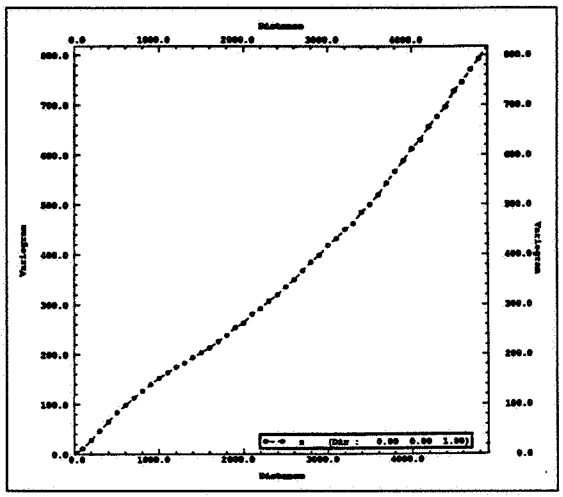


Planche 6

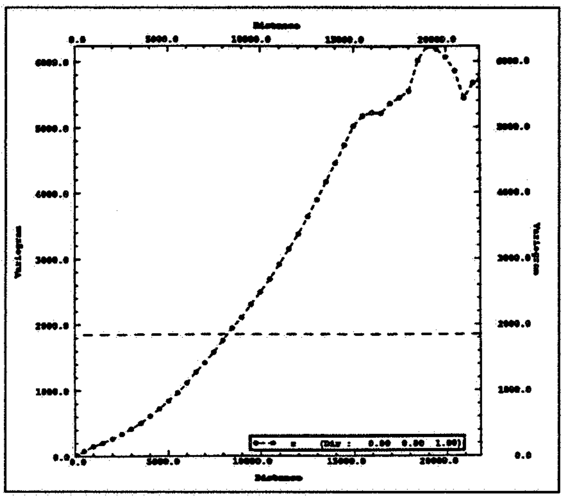
"TITANIC" : VARIOGRAMMES SUR PROFILS



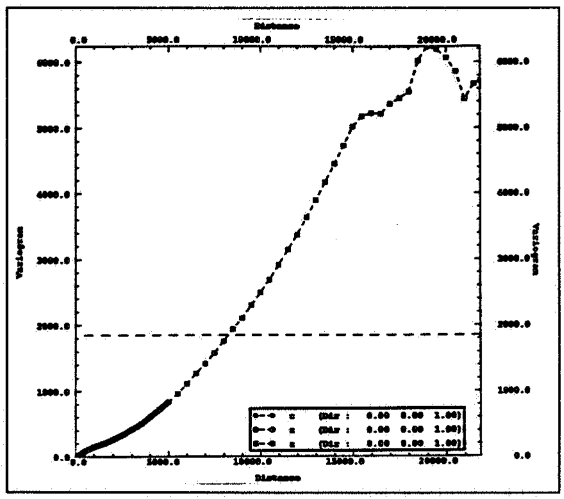
Pas : 50 mètres



Pas : 100 mètres



Pas : 500 mètres



Récapitulation des variogrammes

"TITANIC" : ESTIMATION PAR KRIGEAGE PONCTUEL

Voisinage :	12 points, pas de code	
Structure :	k = 1	
	Linéaire	$-0.3591\ 10^{-1}$
	Spline	$0.4383\ 10^{-4}$

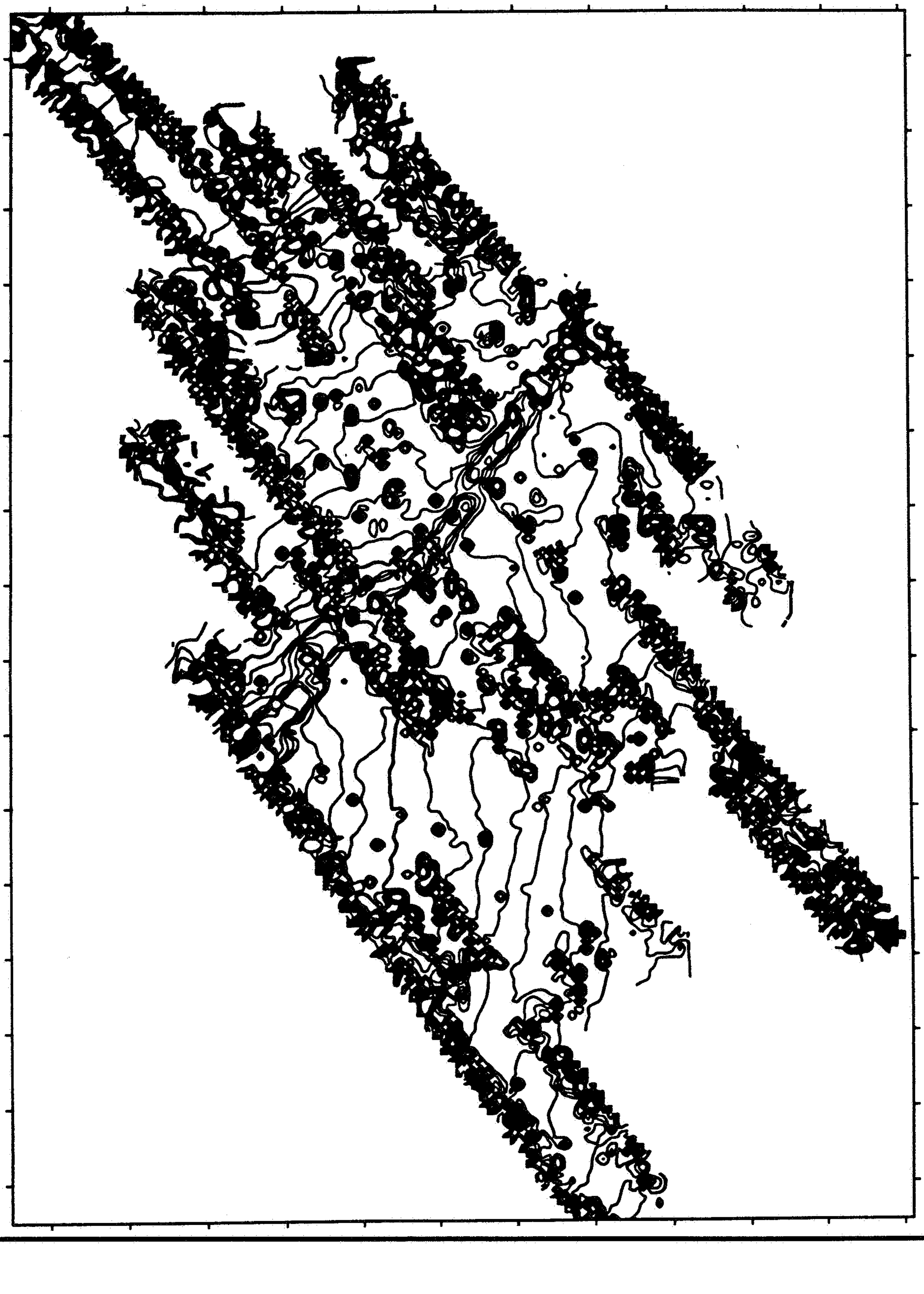


Planche 8

“TITANIC” : ESTIMATION PAR KRIGEAGE PONCTUEL

Voisinage : 16 points, code [3]
Structure : $k = 2$
Linéaire -0.3087

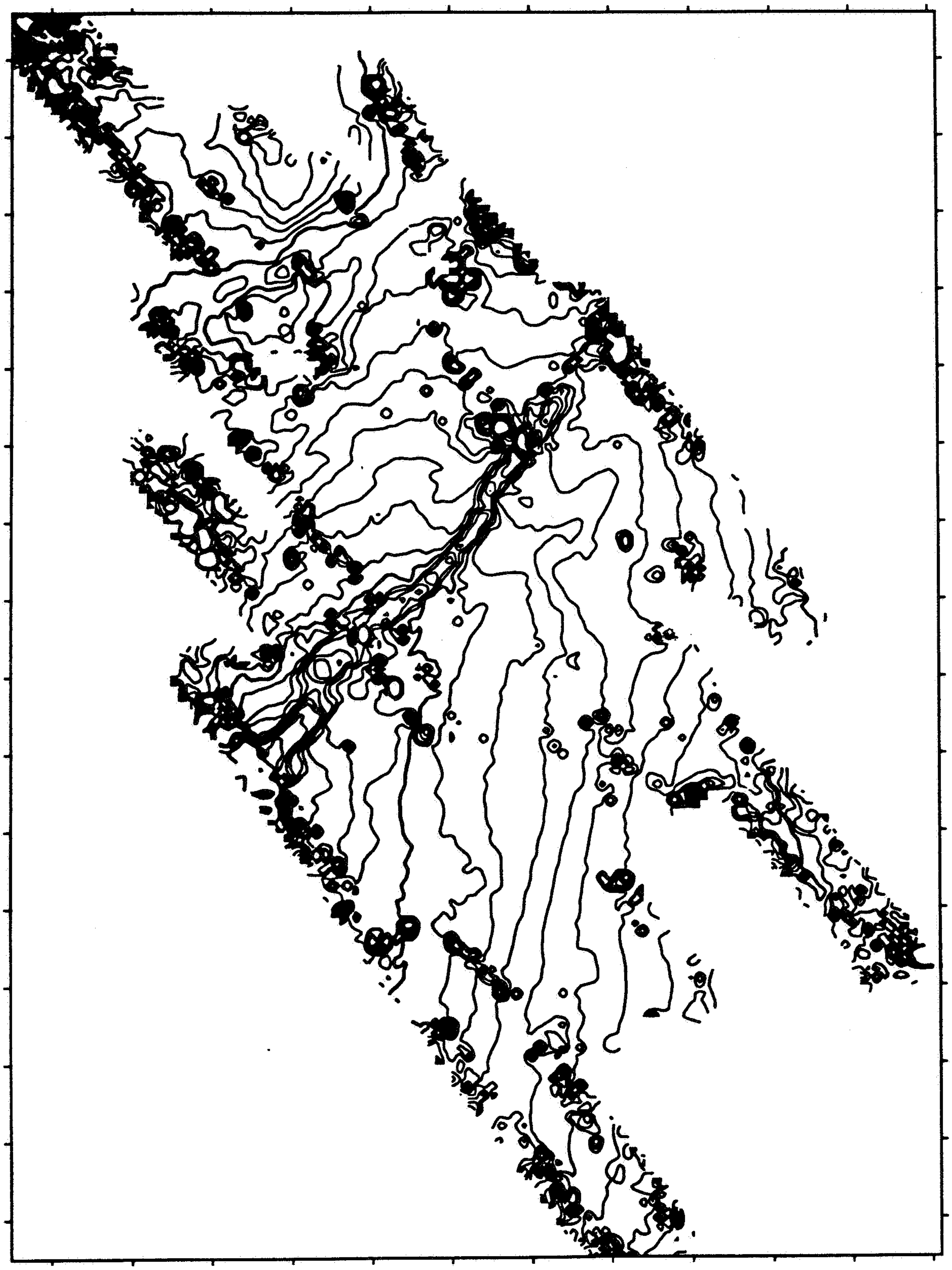


Planche 9

“TITANIC” : ESTIMATION PAR KRIGEAGE PONCTUEL

Voisinage : 24 points, code [5]
Structure : $k = 2$
Cubique $0.5056\ 10^{-7}$

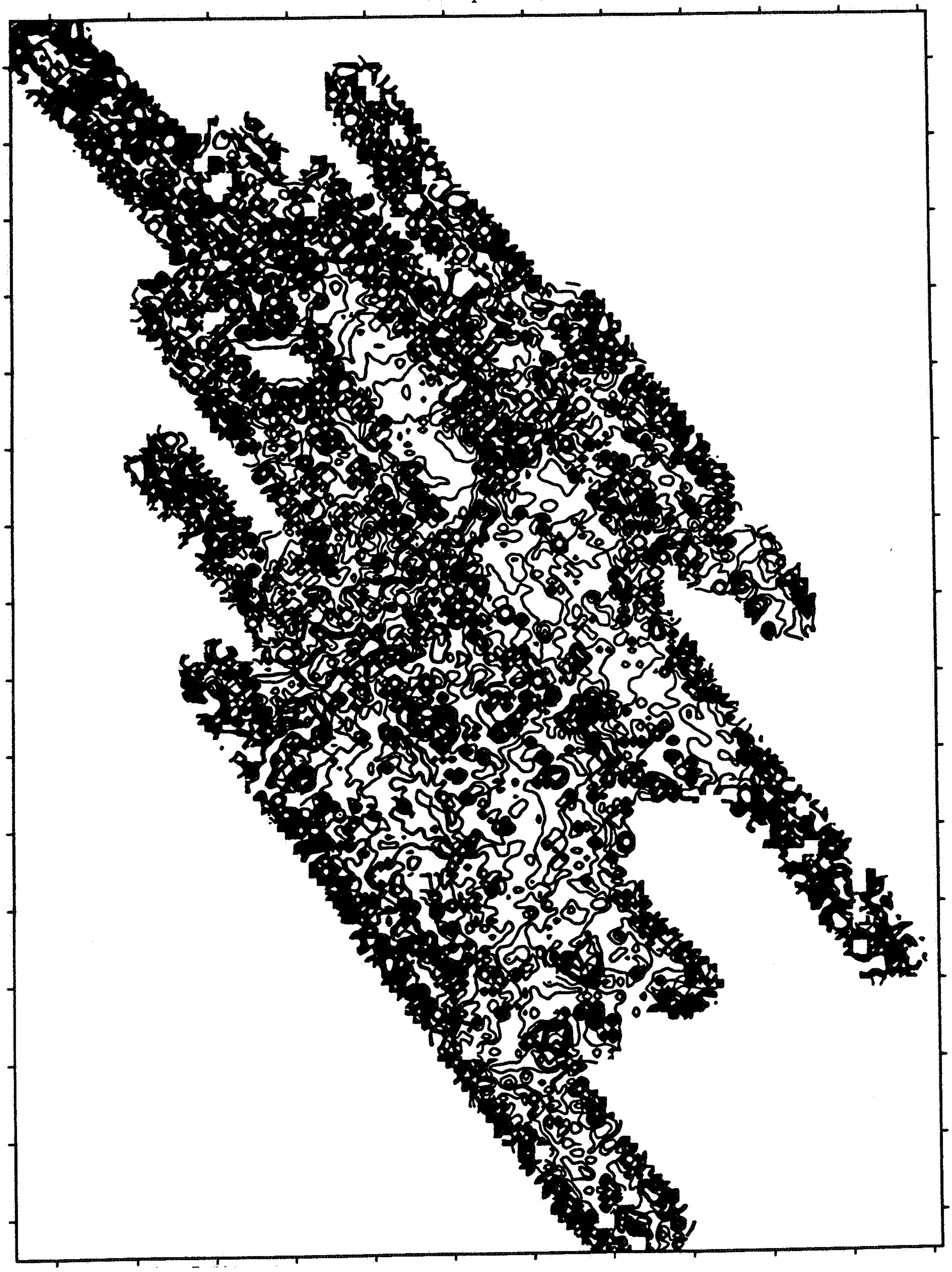


Planche 10

“TITANIC” : ESTIMATION PAR KRIGEAGE PONCTUEL

Voisinage : 24 points, code [5]
Structure : k = 1 (forcé)
Linéaire $-0.1360\ 10^{-1}$
Cubique $0.6797\ 10^{-8}$

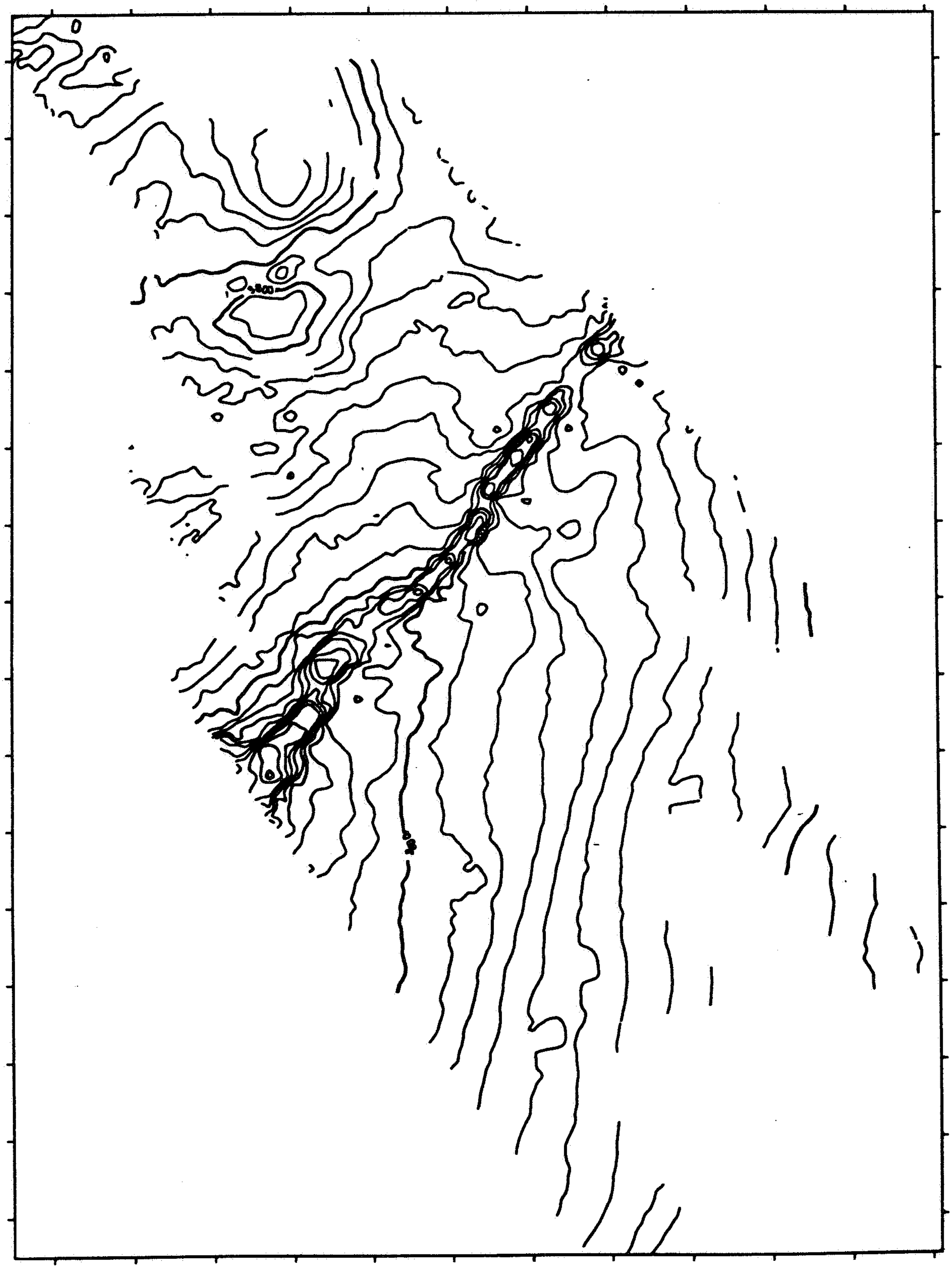


Planche 11

"TITANIC" : ESTIMATION PAR KRIGEAGE PONCTUEL

Voisinage : 8 points, code [5]
Structure : k = 1 (forcé)
Linéaire -0.2047 10^{-1}
Cubique 0.1869 10^{-7}

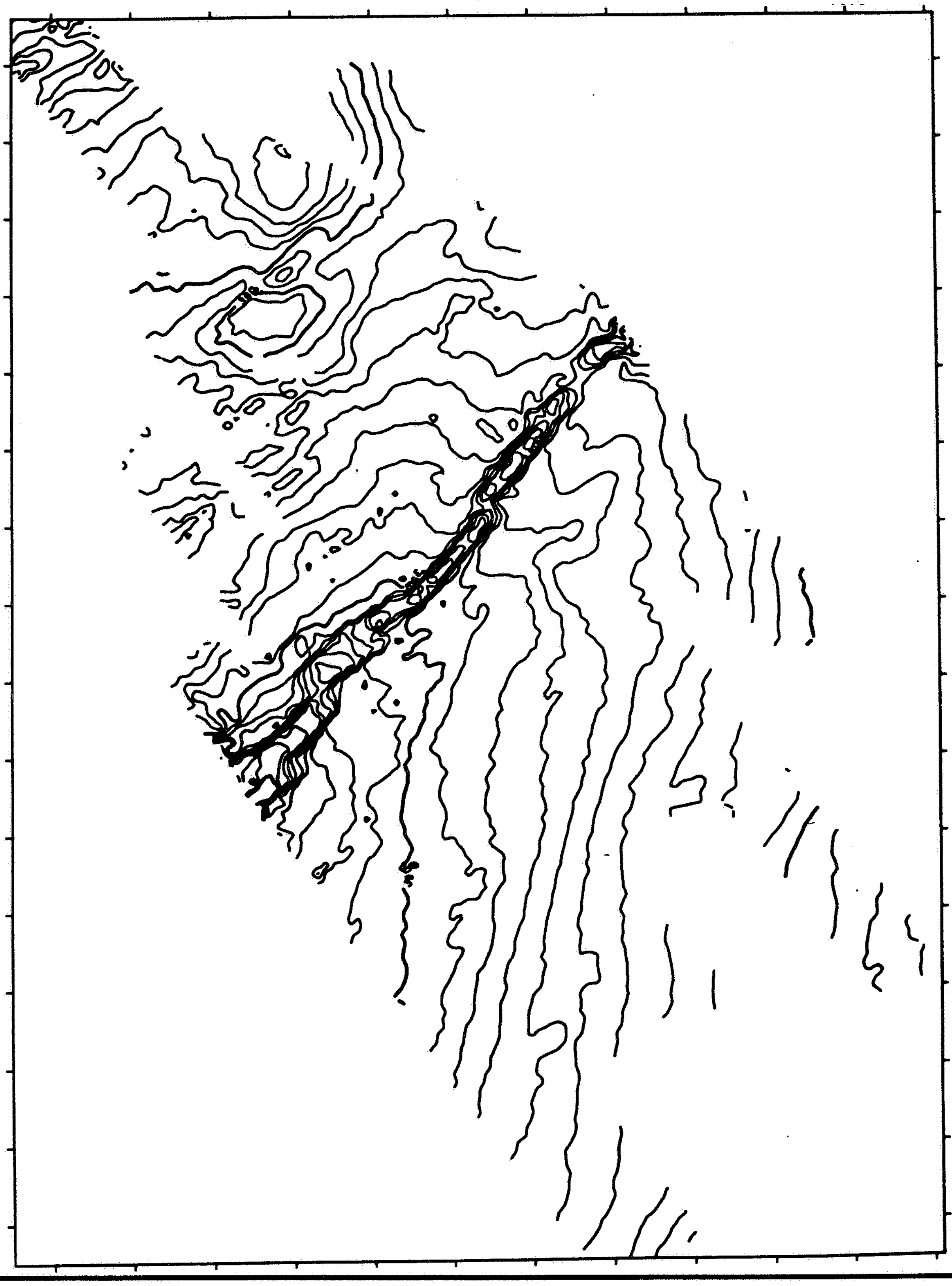
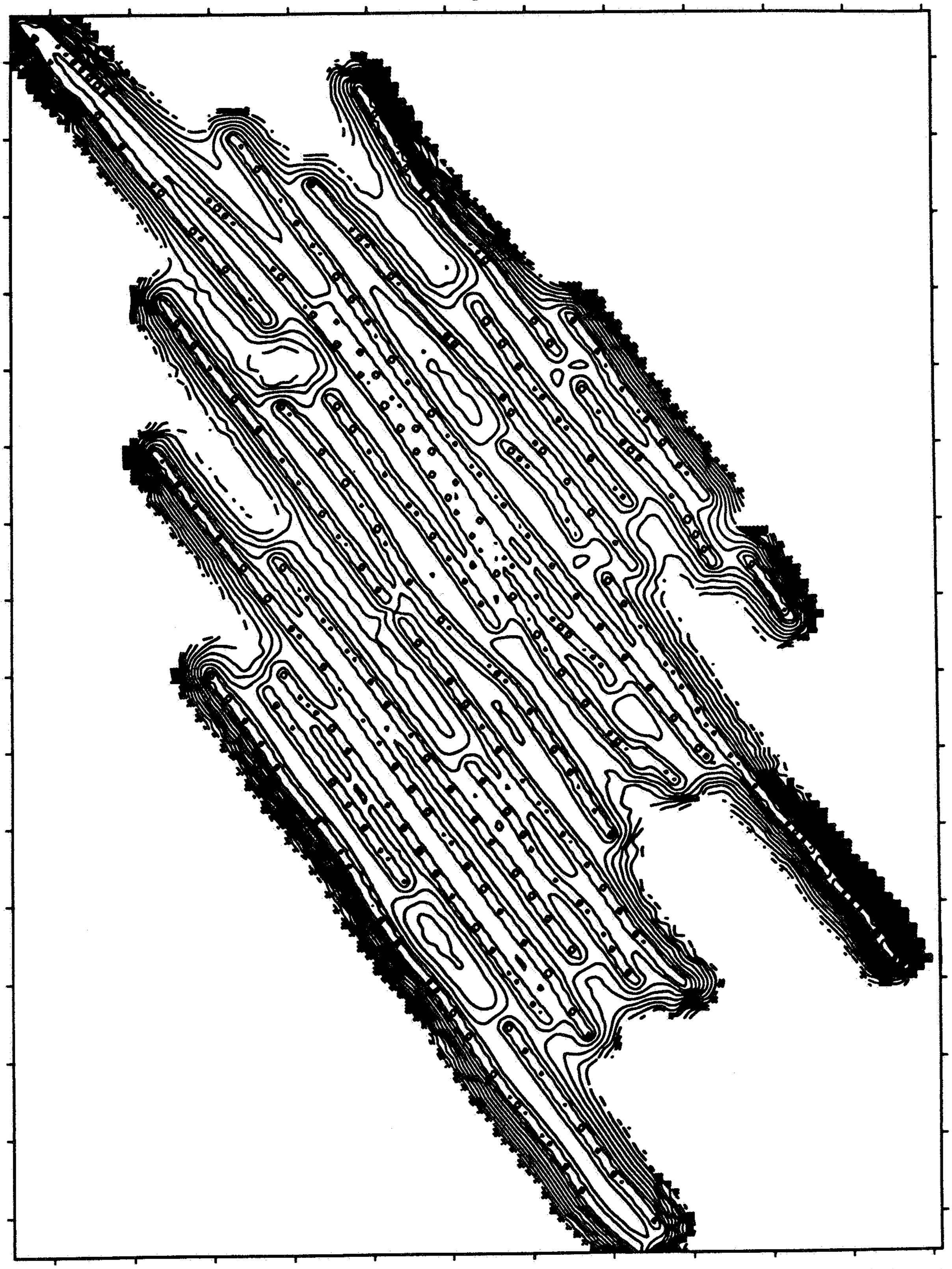


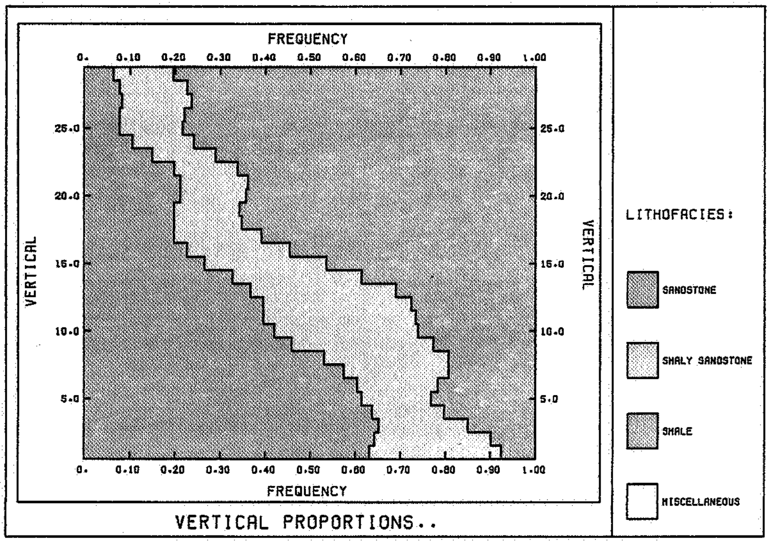
Planche 12

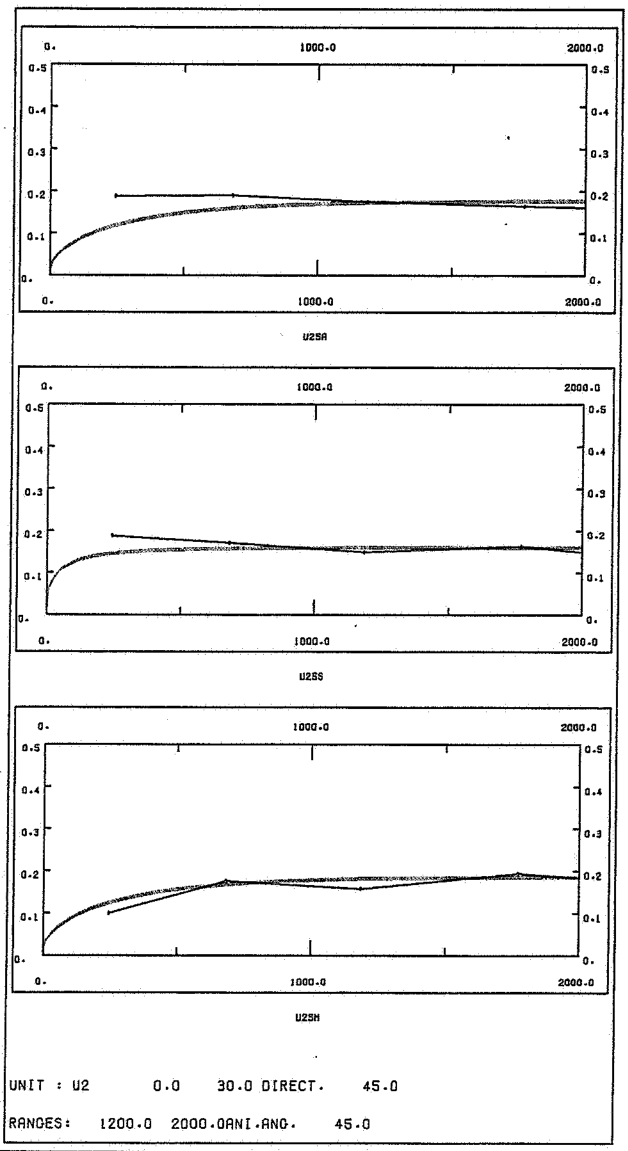
"TITANIC" : ÉCART-TYPE DE KRIGEAGE PONCTUEL

Voisinage : 8 points, code [5]
Structure : k = 1 (forcé)
Linéaire -0.2047 10^{-1}
Cubique 0.1869 10^{-7}



RESERVOIR HETEROGENE : ANALYSE STRUCTURALE





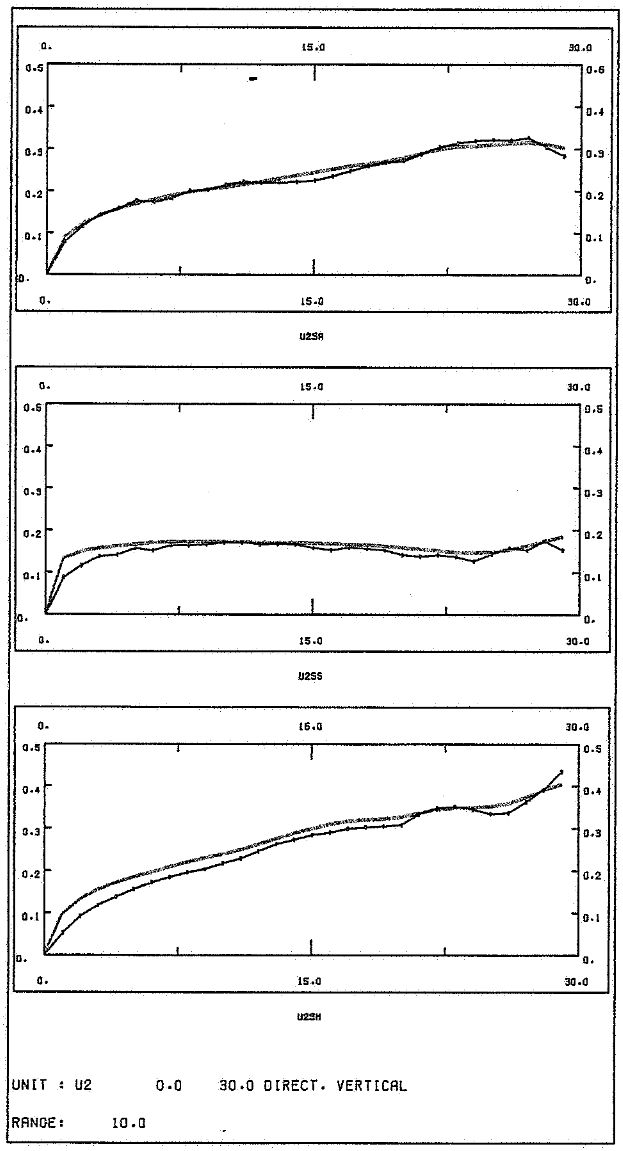


Planche 14

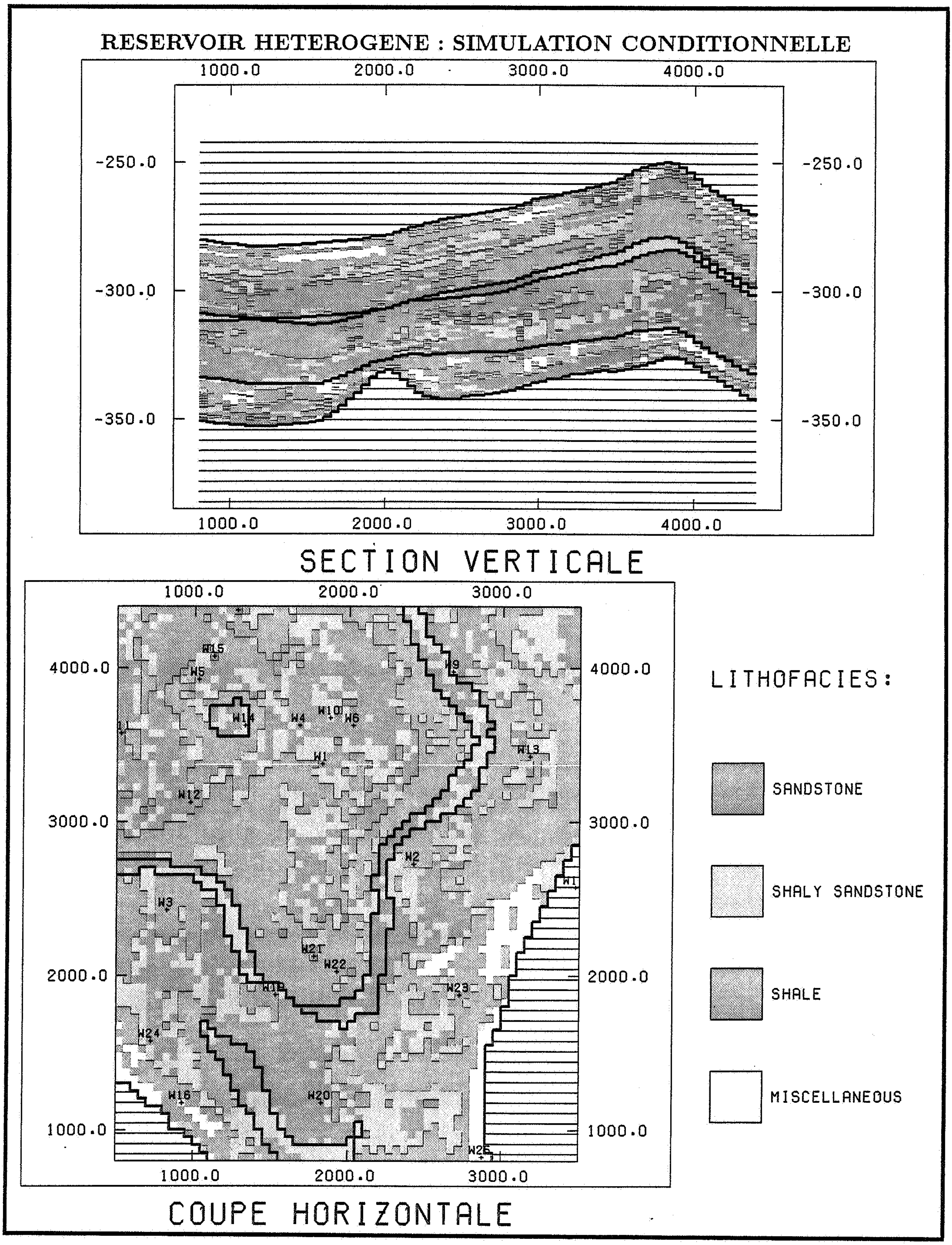
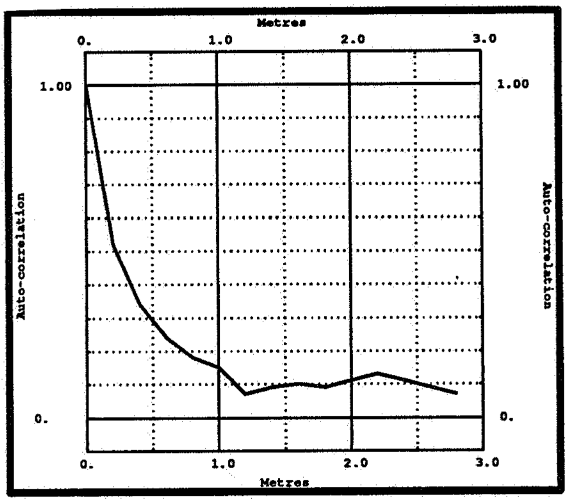
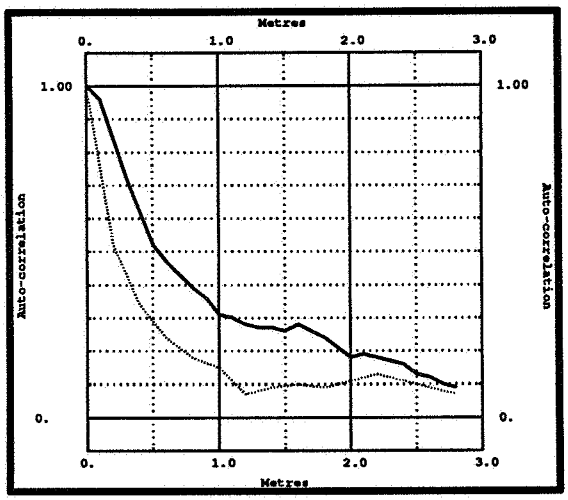

RESERVOIR HETEROGENE : SIMULATION CONDITIONNELLE
1000.0
2000.0
3000.0
4000.0
-250.0
-300.0
-350.0
SECTION VERTICALE
LITHOFACIES:
SANDSTONE
SHALY SANDSTONE
SHALE
MISCELLANEOUS
W15
W5
W9
W14
W4
W10
W6
W1
W13
W12
W2
W3
W21
W22
W23
W24
W16
W20
W25
COUPE HORIZONTALE

Planche 15

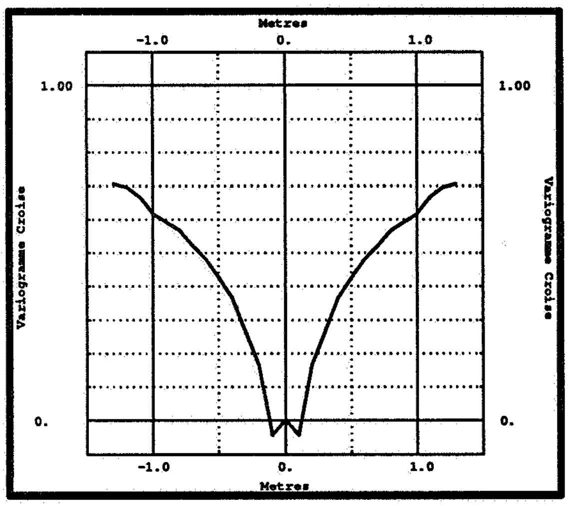
VARIOGRAPHIE RADIOACTIVITÉ–TENEUR



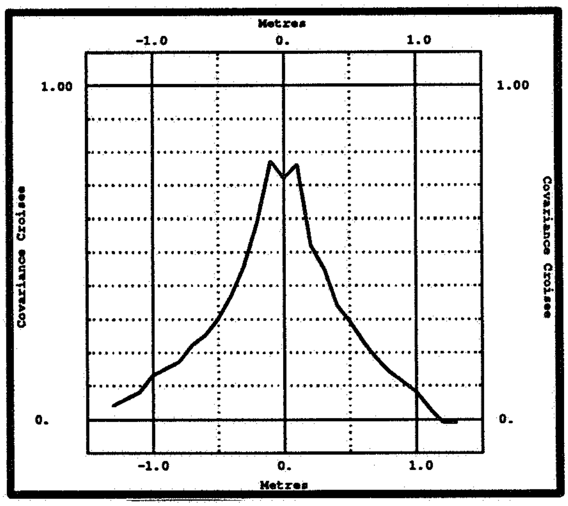
Covariance de la teneur



Covariance de la radioactivité



Variogramme croisé



Covariance croisée