

# REFLEXIONS SUR LES PONDERATEURS NEGATIFS DU KRIGEAGE

Pierre CHAUVET

## TABLE DES MATIERES

A - POSITION DU PROBLEME.....	66	2 - L'échelle de travail.....	85
B - QUELQUES AFFIRMATIONS CONTRADICTOIRES ET REMARQUES PRELIMINAIRES.....	66	3 - Régularité de la carte.....	85
C - LA THEORIE FACE AUX PONDERATEURS NEGATIFS.....	68	4 - Dualité: irrégularité du modèle - régularité de la carte.....	86
D - LA THEORIE FACE AUX ESTIMATEURS NEGATIFS.	70	5 - Des points de vue contradictoires...	87
E - GENESE ET INTERPRETATION DES PONDERATEURS NEGATIFS.....	71	6 - A la source d'un malentendu.....	88
1 - Quelques exemples simples.....	71	a - Précision et irrégularité.....	89
2 - Hypothèse de stationnarité d'ordre 2...	72	b - Estimation et simulation.....	89
3 - Point de vue algébrique.....	72	c - Une explication des divergences?..	90
4 - Interprétation en termes de tendance...	74	d - Un exemple typique.....	91
5 - Généralisation au cas non stationnaire.	74	e - Conclusion partielle.....	92
6 - La dualité "régularité-instabilité".... en hypothèse stationnaire.....	75	7 - Dualité: régularité-précision.....	92
7 - Deux réponses possibles.....	76	8 - Le point de vue des splines.....	93
8 - Première réponse: l'effet de pépite....	76	9 - Utilisation des cartes d'iso- variances.....	95
9 - Deuxième réponse: les conditions d'universalité.....	77	10 - Adéquation du modèle structural.....	96
10 - Conclusions provisoires.....	78	11 - Conclusion; retour au problème des pondérateurs négatifs.....	97
F - DES PONDERATEURS NEGATIFS AUX ESTIMATEURS NEGATIFS.....	79	H - QUELQUES REPONSES AU PROBLEME DES PONDERATEURS NEGATIFS.....	98
1 - Remarque méthodologique préalable.....	79	1 - Remarque préalable.....	99
2 - Remarque sur le changement de support..	80	2 - Le statu-quo.....	99
3 - Instabilité de l'estimateur selon les pondérateurs.....	80	3 - La mise à zéro.....	100
a - Pondérateurs positifs de somme égale à 1.....	81	4 - Réajustement du modèle structural...	102
b - Pondérateurs positifs de somme quelconque.....	81	5 - Modification du plan de krigeage....	102
c - Modèle global, caractéristiques locales.....	82	a - Nombre de données.....	102
d - Pondérateurs négatifs.....	82	b - Qualité de l'information.....	103
e - Conclusion provisoire sur l'insta- bilité.....	83	c - Choix d'un plan de krigeage.....	104
G - LE POINT DE VUE DE L'INTERPOLATION.....	83	d - Modification du plan de krigeage..	106
1 - Régularité du modèle et de la fonction d'interpolation.....	84	6 - Passage à la Géostatistique Non Linéaire.....	107
		7 - Choix d'algorithmes nouveaux.....	108
		a - Contraintes sur les pondérateurs..	109
		b - Contraintes sur les valeurs estimées.....	110
		c - Brève conclusion.....	111
		BIBLIOGRAPHIE.....	112

\* Centre de Géostatistique, ECOLE NATIONALE SUPERIEURE DES MINES DE PARIS, 35 rue Saint-Honoré, 77305 FONTAINEBLEAU (France).

*Etudes Géostatistiques V - Séminaire C.F.S.G. sur la Géostatistique 15-16 Juin 1987, Fontainebleau. Sci. de la Terre, Sér. Inf., Nancy, 1988, 28, pp. 65 à 113.*

## A – POSITION DU PROBLEME

La possibilité de voir surgir des pondérateurs négatifs au cours d'un krigeage, est un exemple typique de ces problèmes qui risquent de susciter l'incompréhension entre les géostatisticiens théoriques et les praticiens. On peut en effet imaginer les positions extrémistes suivantes:

- pour le théoricien, l'existence de pondérateurs négatifs ne pose pas de problème théorique. C'est un faux problème. Donc, il n'y a pas lieu d'examiner cette question.
- pour le praticien, l'apparition d'estimateurs négatifs peut être rédhibitoire. Donc, il convient de rejeter le krigeage dans son ensemble.

Ces points de vue sont évidemment caricaturaux. Cependant, ils illustrent bien comment, à partir de prémisses correctes, on peut générer des situations d'incompréhension - voire de polémique. Un tel état de fait est évidemment extrêmement regrettable dans une discipline scientifique, et dire que chacun a raison "de son point de vue" n'est pas une solution bien satisfaisante. Aussi, je me propose d'essayer de démystifier ici ces fameux pondérateurs négatifs: d'abord par quelques remarques sur l'approche théorique des poids, puis des estimateurs négatifs (Sections C et D); ensuite surtout par l'examen pratique de la genèse et la signification de ces pondérateurs (Section E). Pour compléter les commentaires sur les estimateurs négatifs (Section F), puis préciser la notion de régularité, j'adopterai le point de vue de l'interpolation (Section G), qui a l'avantage d'offrir des interprétations simples en termes de tracés de cartes.

Dans une dernière étape, j'essaierai d'apporter quelques éléments de réflexion aux solutions pratiques qui peuvent être proposées par divers auteurs. Naturellement, je supposerai que ces solutions s'inscrivent dans le cadre du krigeage: car invoquer des estimateurs de krigeage négatifs pour rejeter le krigeage dans son ensemble, ne saurait être qu'un prétexte venant d'utilisateurs décidés de toute façon à rejeter la géostatistique. Les difficultés de mise en oeuvre d'une théorie sont choses courantes, et ne suffisent pas à faire renoncer à une méthode: pensons par exemple aux praticiens de l'uranium qui obtiennent des teneurs négatives par "déconvolution" des mesures de radioactivité...

## B – QUELQUES AFFIRMATIONS CONTRADICTOIRES ET REMARQUES PRELIMINAIRES

L'inventaire des problèmes soulevés par les pondérateurs négatifs, est proposé de façon remarquablement concise dans (Barnes et Johnson, 1984, p. 231):

*"The presence of negative weights yields several unsatisfactory results.*

- 1) *Negative weights can produce negative estimated grades, an intolerable outcome.*

2) Negative weights can produce estimated grades greater than the highest sample value, a very dangerous outcome.

3) Also, negative weights can cause highly erratic estimated grades: slight changes in location can bring about substantial changes in estimate."

(Note: j'ai rajouté une numérotation à ces trois propositions, ainsi qu'aux prochaines, pour faciliter les références).

L'énoncé des faits est inattaquable. En ce qui concerne leur appréciation ("intolérable", "very dangerous"), une restriction s'impose: les auteurs se placent dans une optique minière. Il est significatif d'ailleurs de noter cette différence d'intensité entre leurs deux appréciations, alors que sur le plan théorique il ne s'agit que d'une seule et même propriété (il y a là une source possible d'incompréhension entre théoriciens et praticiens; c'est dans de tels cas qu'il faut accentuer les efforts pédagogiques).

Il est d'ailleurs amusant que d'autres auteurs attendent au contraire du krigeage qu'il puisse fournir des valeurs extérieures à l'intervalle des valeurs des données. Ainsi, (Journal, 1986, p. 272):

*"The fact that some kriging weights can be negative is extremely positive because it warrants the possibility of a nonconvex estimate lying outside the minimum and maximum data value."*

Cette possibilité est d'un intérêt évident, par exemple en cartographie, quand il s'agit de restituer des traits caractéristiques (lignes de crêtes, talwegs, etc...). Aussi, bien que cette quatrième proposition tirée de (Philip et Watson, 1986a, p. 100 puis p. 102):

4) "Kriging is constrained by the data range"

soit - simplement - fausse, elle contient un souhait implicite qui aurait été sans cela parfaitement respectable.

On ne peut bien sûr pas reprocher au krigeage de ne pas satisfaire simultanément aux vœux contradictoires de PHILIP et WATSON d'une part, de BARNES et JOHNSON d'autre part. Si souhaitées, des modifications du krigeage usuel devront donc témoigner d'un choix, et engageront la responsabilité de leurs auteurs. Ainsi, BARNES et JOHNSON ont choisi de résoudre explicitement le point 1 (considéré comme "intolérable"); et comme le Krigeage Positif est conçu comme un prolongement du Krigeage Ordinaire (KO), les points 2 et 3 sont ipso facto résolus. Mais si, tout aussi légitimement, le Krigeage Positif était élaboré dans une optique KS (Krigeage Simple: sans condition sur la somme des pondérateurs), le choix serait autre, puisque seul serait résolu le premier problème (et indirectement en partie le troisième).

Naturellement, le point de vue du Krigeage Positif va à l'encontre du vœu de PHILIP et WATSON. Rien d'anormal à cela: à l'évidence, les contextes ne sont pas les mêmes, ni les points de vue. Il est donc naturel que les choix proposés divergent. La seule règle est de faire ces choix en connaissance de cause, et d'évaluer si les coûts d'une nouvelle méthode (je veux dire: coûts en rigueur théorique, en précision, etc...) sont justifiés: ce dernier point sera évoqué ultérieurement (Section H).

Si l'on n'est pas a priori engagé dans un point de vue plutôt que dans un autre (ni minier, ni cartographique, ...), on est alors fondé à considérer que c'est la troisième critique de BARNES et JOHNSON qui est la plus grave. Car une instabilité numérique n'est jamais souhaitable, quel que soit le problème traité. De plus, il s'agit là d'un danger à la fois plus immédiat et plus caché que l'apparition de teneurs négatives:

- a) plus immédiat. Car, pour qu'apparaisse une valeur estimée négative, il faut qu'il y ait à la fois des pondérateurs négatifs et des données de forte valeur qui leur soient associées. L'instabilité numérique par contre est une propriété du seul système de krigeage: c'est donc une propriété qui ne dépend que de la structure de la variable et de la géométrie de l'information. C'est un danger permanent.
- b) plus caché, bien sûr. Car l'apparition d'une teneur négative est un phénomène spectaculaire, qui a par-là même valeur de signal d'alarme. Au contraire, des fluctuations importantes des valeurs estimées peuvent être prises à tort, hors de leur contexte, pour un regain de précision: il n'y a pas cette fois de mise en garde immédiate, mais bien au contraire piège pour l'utilisateur inattentif.

Un dernier point enfin, étroitement lié au problème des pondérateurs négatifs. On sait déjà en effet (par exemple, Matheron, 1971, pp. 130-131; repris ci-dessous, Section E-1) que, plus la structure du phénomène est régulière, plus les possibilités sont grandes de voir apparaître des pondérateurs négatifs. Les valeurs estimées, quoique très structurées, sont alors simultanément erratiques (au sens: elles peuvent prendre un gamme très large de valeurs), et instables: telle est la proposition (3) de BARNES et JOHNSON. Le facteur "instabilité" est de toutes façons un danger qu'il faut soigneusement prendre en compte.

Le double aspect: "structure forte" et "valeurs erratiques", est, quant à lui, l'illustration du mécanisme suivant: l'existence de pondérateurs négatifs a pour effet, alors même que le modèle de variable régionalisée est stationnaire, de générer au niveau de l'estimateur des comportements de dérive locale ("dérive" est pris ici au sens intuitif).

Ce phénomène est jugé "unsatisfactory" par BARNES et JOHNSON; il est au contraire nié ou méconnu par d'autres auteurs. Ainsi, on trouve dans (Philip et Watson, 1986a, p. 100) l'assertion suivante:

5) "Local trends within the data are ignored"

Une nouvelle contradiction à examiner...

## C – LA THEORIE FACE AUX PONDERATEURS NEGATIFS

Du point de vue du théoricien, l'existence de pondérateurs négatifs ne constitue pas un problème, car elle ne met nullement en cause la validité de la théorie. Plus précisément, repreneons brièvement les deux étapes du krigeage:

- 1) on cherche un estimateur linéaire qui soit non biaisé (i.e.: dont l'erreur d'estimation soit, dans le modèle, une variable aléatoire d'espérance nulle). En théorie des FAIk, les pondérateurs doivent satisfaire à un certain nombre de "conditions d'universalité"; en KO

(Krigeage Ordinaire), ces conditions se limitent à: "somme des pondérateurs égale 1"; et en KS (Krigeage Simple), il n'y a aucune contrainte. Mais de toutes façons, aucune de ces conditions n'impose aux pondérateurs d'être positifs: au niveau du KS, il s'agit même d'une liberté totale.

- 2) la seconde contrainte de l'estimateur de krigeage, est d'être de variance minimale. Là encore, cette condition n'a par elle-même aucune conséquence directe sur le signe des pondérateurs.

En résumé donc, en langage théorique dépourvu de jugement de valeur:

*"le formalisme de la géostatistique linéaire n'implique aucune contrainte sur le signe des pondérateurs de krigeage"*

et, en langage pratique plus engagé (correspondant par exemple à l'optique des utilisateurs miniers):

*"la pratique de la géostatistique linéaire ne garantit aucune protection contre les pondérateurs négatifs".*

Du point de vue théorique, il n'y a pas à décider si cette liberté sur les signes est un bien ou un mal; cette question est dépourvue de signification théorique. Quant au point de vue de l'application, on a vu que les auteurs ont des opinions contradictoires, ce qui est d'ailleurs parfaitement naturel; car il est bien normal qu'à des problèmes divergents correspondent des choix opératoires contradictoires.

Il me semble par contre essentiel, pour que la discussion méthodologique reste rigoureuse, de noter que cette propriété des pondérateurs du krigeage est une propriété globale. Autrement dit, on ne saurait critiquer telle valeur des pondérateurs (ou plus tard de l'estimateur) sans faire la critique globale de l'algorithme d'estimation lui-même. De façon plus concrète, cela signifie qu'il faut éviter les réajustements ad hoc, réalisés "dans le feu de l'action", qui perdent de vue l'ensemble du phénomène étudié. Lorsque l'analyse structurale est réalisée et lorsque la géométrie des données à utiliser est connue, on dispose de tous les éléments qui généreront ou non des pondérateurs négatifs: c'est donc à cette étape qu'il convient de réfléchir aux dispositions à prendre ou, si le krigeage a été effectivement réalisé et si des résultats observés ont été contestés, c'est à cette étape qu'il convient de revenir.

Une remarque de vocabulaire: j'emploierai toujours les mots "critique(r)" dans leur acception noble. Critiquer signifiera ici: "réfléchir sur", "commenter", "analyser". Donc, nulle nuance péjorative, bien au contraire. Au sens littéral, une proposition "criticable" est une proposition "digne d'intérêt".

Ainsi, la critique de notre problème repose sur deux facteurs:

- a) le modèle structural adopté,
- b) la géométrie de l'information,

et ce sont ces deux points qu'il faut examiner pour comprendre la genèse des pondérateurs négatifs. Mais pour le praticien, il ne suffit pas d'avoir correctement appréhendé l'origine de ses difficultés: il faut également pouvoir élaborer une solution satisfaisante. Deux attitudes (qui ne sont pas incompatibles) sont alors possibles:

- a) agir sur l'un des deux facteurs déterminants ci-dessus (et on peut alors rester dans le cadre de l'algorithme habituel du krigeage), ou
- b) changer d'algorithme (exemple: le Krigeage Positif).

Ces deux options seront examinées ultérieurement (Section H).

## D - LA THEORIE FACE AUX ESTIMATEURS NEGATIFS

C'est à dessein que je dissocie cette question de la précédente. Car l'apparition d'estimateurs négatifs met en jeu un facteur nouveau, à savoir les valeurs particulières prises par les données.

Toujours en restant au plan théorique, on sait que la géostatistique linéaire ne prend pas en compte la loi de la variable régionalisée. Ainsi, au niveau du modèle, il n'est jamais exprimé que la variable doit rester positive, ou inférieure à un seuil maximum. Ultérieurement, il n'est pas non plus imposé de contraintes d'inégalité à l'estimateur. L'apparition d'un estimateur négatif (pour une variable supposée bien sûr positive...) n'est en fait que la manifestation la plus spectaculaire d'une propriété plus générale de la géostatistique linéaire:

*"l'estimateur du Krigeage n'est pas astreint à être inclus dans l'intervalle de valeurs défini par les données expérimentales".*

Telle quelle, cette propriété (qui soit dit en passant est l'exacte négation de la proposition (4) ci-dessus) est une conséquence presque immédiate de l'existence possible de pondérateurs négatifs (à laquelle il conviendrait d'ajouter l'existence possible de pondérateurs supérieurs à 1). Ensuite, l'apparition effective d'une valeur estimée négative nécessite la coïncidence d'un des ces pondérateurs négatifs et d'une donnée forte.

L'approche théorique de cette simultanéité est probablement inextricable dans le cas général. Il faut cependant rappeler qu'une hypothèse de base de la géostatistique est le caractère non préférentiel de l'information: l'implantation des données ne doit pas être conditionnée par les valeurs espérées, présumées ou déjà observées de la variable (je suppose donc ici résolus les problèmes pratiques importants d'anisotropie ou d'hétérogénéité, qui peuvent légitimement - et même doivent - orienter une campagne de reconnaissance future). Il faut alors faire la distinction, suivant que l'on est ou non en hypothèse stationnaire.

Dans le cas stationnaire, pour une configuration d'information implantée au hasard de façon équiprobable dans le champ étudié, la valeur observée en quelque point de mesure que ce soit a toujours - dans le modèle - la même loi de probabilité. Cela s'applique bien sûr en particulier aux points de poids de krigeage négatif. Aussi, la probabilité d'observer une valeur forte affectée d'un poids négatif, ne dépend que de la loi (l'histogramme) des valeurs ponctuelles. (Remarque: je néglige ici les problèmes de frontières, qui méritent un examen particulier). Quant à la probabilité d'obtenir une valeur estimée négative, elle dépendra de la loi (stationnaire) à  $n$  variables, où " $n$ " est le nombre de points de la configuration de krigeage.

Un cas intermédiaire est celui du Krigeage Universel (KU), lorsque le modèle sous-jacent est stationnaire. L'optique du KU impose en effet une interprétation du phénomène, fondée sur la dichotomie. Un tel point de vue permet, dans l'examen des estimateurs du krigeage, de distinguer une composante due à la "dérive" et une composante due aux "résidus". Pour la critique des estimateurs négatifs, on pourrait donc imaginer une généralisation du travail présenté dans (Rivoirard, 1984, chapitre 4).

Le cas non stationnaire, enfin, est bien plus complexe. En effet, un poids de krigeage affecte une valeur de la variable régionalisée elle-même, alors que l'hypothèse stationnaire concerne soit les résidus (dans l'optique du Krigeage Universel, Matheron, 1971, chapitre 4), soit les combinaisons linéaires autorisées (dans l'optique des FAIk, Matheron, 1973, pp. 439-468). C'est donc la configuration de krigeage dans son ensemble qu'il faut examiner, ou plus précisément l'erreur d'estimation. On ne peut dans le modèle donner un sens à des énoncés comme: "il y a une probabilité  $p$  pour que la valeur estimée soit négative", parce que la valeur estimée n'est pas une combinaison linéaire autorisée. Une approche théorique de cette question devrait donc se faire dans l'optique des représentations glissantes locales (Matheron, 1978, pp. 147-162).

## E - GENESE ET INTERPRETATION DES PONDERATEURS NEGATIFS

### 1 - QUELQUES EXEMPLES SIMPLES

Rappelons d'abord l'exemple de (Matheron, 1971, pp. 130-131): soit, à 1 dimension, une configuration de quatre données régulièrement espacées, et soit à estimer la valeur moyenne du segment du milieu. Par raison de symétrie, les deux pondérateurs extérieurs seront égaux; il en sera de même pour les deux pondérateurs intérieurs. On examine les signes de ces pondérateurs lorsque le modèle de variogramme est un monôme de degré compris entre 0 et 2 (comme ce modèle n'est pas stationnaire, on doit pratiquer un Krigeage Ordinaire: la somme des poids est astreinte à être égale à 1).

Le résultat est bien connu: les pondérateurs extérieurs sont une fonction monotone décroissante de l'exposant.

- lorsque l'exposant est inférieur à 1, tous les pondérateurs sont positifs. En particulier, lorsque l'exposant décroît vers la valeur 0 - c'est-à-dire lorsque le modèle tend vers l'effet de pépite pur - les pondérateurs extérieurs croissent chacun vers la valeur  $1/4$ .
- à la limite, comme on le sait, les quatre pondérateurs sont tous égaux à  $1/4$  pour une variable totalement non structurée.
- pour un exposant égal à 1, on observe un effet d'écran: les pondérateurs extérieurs sont nuls. C'est la propriété de Markov du variogramme linéaire.
- enfin, lorsque l'exposant est supérieur à 1, les pondérateurs extérieurs deviennent négatifs.

D'autres exemples, à 2 dimensions cette fois, sont proposés et commentés dans (Rivoirard, 1984).

Il ne s'agit pas là de démonstrations, mais d'illustrations d'un résultat tout-à-fait général: l'apparition de pondérateurs négatifs, pour une configuration donnée, est liée à la plus ou moins grande régularité du phénomène. A la limite bien sûr, pour l'effet de pépite pur, tous les pondérateurs sont égaux, et positifs. A l'opposé, un modèle aussi régulier que la covariance gaussienne peut faire apparaître des pondérateurs très fortement négatifs.

Naturellement, le modèle structural n'explique pas à lui seul ce phénomène. Ne serait-ce que pour des raisons de symétrie, certaines configurations d'information ne généreront jamais de coefficients de krigeage négatifs: ainsi le krigeage d'un cercle par des données équiréparties sur sa circonférence. Par contre, d'autres figures peuvent être considérées comme particulièrement "sensibles": ainsi les doublets de points lorsque le domaine à estimer est approximativement dans leur alignement. On peut d'ailleurs démontrer qu'à 2 dimensions ou plus, et pour tout modèle de covariance continu à l'origine, on peut toujours trouver des configurations de krigeage qui généreront des pondérateurs négatifs (Matheron, 1986).

En résumé, en employant des termes excessivement anthropocentriques, on pourrait dire que la génération de pondérateurs négatifs est la réponse du système de krigeage à une situation qui lui convient mal: la confrontation d'une géométrie d'information et d'une structure sous-jacente difficiles à concilier.

## 2 - HYPOTHESES DE STATIONNARITE D'ORDRE 2

Dans toute la suite de ce paragraphe sur la genèse et l'interprétation des pondérateurs négatifs, je supposerai satisfaite l'hypothèse de stationnarité d'ordre 2: la variable régionalisée est supposée avoir une espérance nulle, et admettre une covariance stationnaire.

Cette restriction n'est pas indispensable, mais devrait rendre plus spectaculaires certaines remarques d'apparence assez paradoxale. Il faudra en particulier garder à l'esprit que, dans le modèle, la dérive est toujours supposée nulle. Sauf indication contraire, on sera donc dans les conditions d'un Krigeage Simple.

## 3 - POINT DE VUE ALGEBRIQUE

On peut proposer une approche plus mathématique. Soit, à titre d'illustration, cet exemple emprunté à (Matheron, cité dans Rivoirard, 1984, p. 55): le cas particulier d'une variable régionalisée à 1 dimension, de covariance gaussienne (donc: indéfiniment dérivable), et échantillonnée à maille régulière.

Remarque pour prévenir toute mauvaise querelle: il s'agit ici d'une hypothèse de travail. Ce n'est pas le lieu d'examiner comment on a pu "diagnostiquer" une propriété comme l'indéfinie dérivabilité à partir d'un échantillonnage discret.

Admettons donc qu'il s'agit d'un choix, et contentons-nous de remarquer que, s'il engage peut-être excessivement la responsabilité du géostatisticien, il n'introduit en revanche aucune incohérence théorique.

La fonction étudiée est donc analytique, et peut faire l'objet d'une approximation par une série de NEWTON (par ex., Jordan, 1965). Il est alors facile de voir que cette approximation est une combinaison linéaire des informations, de poids tantôt positifs, tantôt négatifs (voir aussi Matheron, 1971, p. 201). L'exemple est encore plus probant lorsqu'il s'agit d'une extrapolation: l'approximation, par un polynôme de degré  $n$ , du  $(n+1)$ ème point d'une grille régulière à l'aide des  $n$  points précédents, fait apparaître comme pondérateurs de ces  $n$  points les coefficients alternés du binôme. Ainsi, non seulement les coefficients sont de signes alternés, mais leur valeur absolue peut devenir considérable.

Une remarque préalable de vocabulaire: j'entends par "configuration de krigeage" l'union des données et du domaine à estimer. C'est très exactement l'ensemble des points qui servent à construire l'erreur d'estimation (et non pas l'estimateur seulement).

Revenons maintenant au krigeage. Si on suppose que la portée de la covariance gaussienne soit grande par rapport à la configuration de krigeage, on constate effectivement que le krigeage devient très proche de l'ajustement polynomial. De plus, pour une configuration donnée, on peut étudier l'évolution des poids de krigeage lorsque la portée varie:

- a) pour une portée croissant indéfiniment, l'estimateur du krigeage est donc à peu près équivalent à l'approximation par un polynôme de degré égal au nombre de points de données.
- b) pour une portée tendant vers 0 - c'est-à-dire pour un phénomène tendant vers l'effet de pépite pur - on retrouve au contraire des pondérateurs presque tous égaux, et tous positifs.

On retrouve ici au passage une propriété connue de la covariance gaussienne (Matheron, 1971, p. 135): par passage à la limite, si le domaine informé est continu (segment), le krigeage d'un point extérieur à ce segment ne peut pas s'exprimer par une mesure. Mais surtout, on met ainsi en évidence la genèse des pondérateurs négatifs, à configuration donnée, lorsque la régularité de la variable régionalisée augmente.

Inversement, supposons maintenant le modèle structural fixé. Si la configuration de krigeage est lâche par rapport à la portée de la covariance, tout se passera comme si on était en présence d'une structure très irrégulière (krigeage des grandes mailles: Matheron, 1971, pp. 105-110, pp. 137-138 et p. 200), et on observera une disparition des pondérateurs négatifs. Au contraire, si la configuration de krigeage est petite par rapport à la portée, on se trouve dans le cas où l'estimateur de krigeage est proche de l'ajustement polynomial.

On souligne ainsi le rôle capital de l'échelle dans l'apparition des pondérateurs négatifs. La "régularité" du modèle, les "dérives locales" ne sont pas des notions absolues: elles n'ont de signification que relatées à l'échelle de travail, c'est-à-dire, concrètement, à la maille d'information. On peut donc déjà prévoir que les critiques de nature géostatistique qui seront faites plus tard sur la régularité de l'estimateur (Section H), seront bien différentes de celles qui pourraient être portées d'un point de vue mathématique (par exemple: continuité, dérivabilité, etc...)

## 4 - INTERPRETATION EN TERMES DE TENDANCE

Le formalisme du KU montre que l'ajustement polynomial par moindres carrés est à comparer à l'estimateur optimal de la dérive (Matheron, 1969, pp. 70-77; Matheron, 1971, pp. 159-164). Dans l'exemple de la covariance gaussienne donc, lorsque la structure augmente et qu'apparaissent des pondérateurs négatifs, l'interpolateur de krigeage finit par avoir le comportement d'une "tendance". Cette remarque - il s'agit bien sûr d'une interprétation, pas d'une démonstration - peut s'étendre à d'autres modèles lorsque des pondérateurs négatifs importants sont mis en jeu.

Une importante remarque de vocabulaire: le concept de dérive est ici celui qui est défini rigoureusement en théorie du Krigeage Universel (Matheron, 1969, p. 3; Matheron, 1971, p. 141). Par opposition, le terme de tendance désigne une notion vague de "structure locale", "comportement régulier", etc... Rappelons que, pour simplifier, on a supposé dans ce paragraphe que la dérive est nulle.

Autrement dit, même dans le cas d'un Krigeage Simple (ce que nous avons supposé ici), le système cherche par le biais d'une "tendance" à introduire une dérive qui, par hypothèse, est absente du modèle. Mais cette "tendance" que le krigeage cherche à représenter, n'est pas maîtrisée par le géostatisticien; ou, plus exactement, sa complexité est liée essentiellement au nombre de données utilisées pour le krigeage.

Pour reprendre des termes anthropocentriques: faute d'avoir une modélisation a priori de cette tendance, le système est contraint de se reposer exclusivement sur les données elles-mêmes, avec bien sûr tous les risques que cela comporte. Les données sont très (vraisemblablement: trop) sollicitées par l'algorithme d'estimation.

On se trouve alors dans une situation comparable - il s'agit là d'une analogie, pas d'une preuve! - à un ajustement par moindres carrés d'un polynôme de degré très (vraisemblablement: beaucoup trop) élevé:

- la courbe (ou surface) estimée a de forts caractères de continuité et de dérivabilité; en contrepartie, les valeurs interpolées peuvent sortir notablement de l'intervalle des données utilisées;
- les écarts quadratiques sont très faibles; l'ajustement est donc en apparence d'excellente qualité;
- mais ces résultats sont extrêmement instables, et une toute petite modification de l'implantation des données peut conduire à un ajustement totalement différent, quoique tout aussi régulier, et considéré comme tout aussi précis par le modèle.

Les première et troisième remarques ci-dessus recouvrent la totalité de la proposition (3) de BARNES et JOHNSON.

## 5 - GENERALISATION AU CAS NON-STATIONNAIRE

L'hypothèse de stationnarité faite au paragraphe précédent, n'avait pour but que de rendre plus spectaculaire le sens de l'apparition des poids négatifs. Mais bien sûr, dans un

formalisme non stationnaire (KU ou FAIk), le mécanisme demeure exactement le même: à la dérive qui cette fois est présente dans le modèle (donc: contrôlée par le géostatisticien), s'ajoute donc une "tendance", que suscite la régularité du modèle structural par le biais des pondérateurs négatifs.

Le point de vue de l'interpolation permet de séparer, au niveau du comportement de l'estimateur, ce qui est dû au terme de dérive (les fonctions de base), et ce qui est dû à la composante aléatoire du modèle (covariance ou covariance généralisée) (Matheron, 1971, pp. 168-173). Sans préciser les détails théoriques, notons que cette seconde composante s'apparente de près à la réalisation d'une dérive aléatoire (Matheron, 1970; Matheron, 1971, pp. 173-183); or, on sait combien il est illusoire, lors de l'analyse structurale, d'"identifier" une composante déterministe et une composante aléatoire de la dérive. En réalité, il n'y a pas de "vraie" dichotomie cachée sous les données: la séparation procède d'un choix. C'est une initiative qui doit être prise par le géostatisticien, sous sa propre responsabilité, et en fonction bien sûr du problème qu'il a à résoudre.

On complète ainsi les observations faites en hypothèse stationnaire, avant de revenir à cette même hypothèse dans les paragraphes suivants pour raisons de simplicité. On justifie aussi a priori le principe des manipulations qui seront critiquées ultérieurement (Section H): le géostatisticien non seulement peut, mais doit, intervenir au niveau de l'analyse structurale et du plan de krigeage, pour adapter au mieux son modèle au problème rencontré.

## 6 - LA DUALITE "REGULARITE-INSTABILITE" EN HYPOTHESE STATIONNAIRE

On peut maintenant préciser le sens des pondérateurs négatifs:

*"les coefficients négatifs sont révélateurs de la difficulté d'un modèle unique à rendre compte A LA FOIS des propriétés de STATIONNARITE GLOBALE et de grande REGULARITE LOCALE".*

La covariance gaussienne ne représentait que le cas extrême de ces difficultés, qui apparaissent également souvent avec par exemple les modèles réguliers de type "polynomial" (Chilès, 1977). Plus que dans le cas d'un modèle irrégulier, l'utilisateur doit alors avoir une totale confiance dans ses données, et ceci pour deux raisons au moins:

- les comportements locaux, suscités par les pondérateurs négatifs, ne sont contenus que de façon implicite - voire cachée - dans le modèle. En quelque sorte, un modèle très régulier veut dire beaucoup plus que ce que l'utilisateur croit y avoir mis; mais ses conséquences ne sont pas maîtrisées a priori.
- d'autre part, même si les variances de krigeage sont contrastées, elles demeurent en général assez faibles, précisément parce que le modèle est stationnaire. En KS, sans conditions sur les poids de krigeage, la variance de krigeage ne sera jamais supérieure à la variance a priori, même en extrapolation. Il n'y a pas de "garde-fou" contre une utilisation trop confiante du modèle. (Ce point sera réexaminé à propos des remarques sur le krigeage comme interpolateur, Section G-10).

## 7 - DEUX REPNSES POSSIBLES

Remarquons d'abord que la genèse des pondérateurs négatifs aurait pu être abordée sous un angle complètement différent: par l'étude purement algébrique du conditionnement du système de krigeage. Mais le point de vue géostatistique a l'avantage de permettre d'insister sur l'interprétation des phénomènes rencontrés par le praticien.

Et de fait, on a vu que l'apparition des pondérateurs négatifs peut être associée à une (trop) grande régularité du modèle structural. Aussi, et c'est la première attitude possible, on peut vouloir agir sur ce modèle dans le sens d'une plus grande irrégularité, pour essayer au besoin de faire disparaître ces pondérateurs négatifs. Une seconde attitude possible consiste au contraire à chercher à maîtriser cette régularité, en spécifiant davantage le modèle, pour essayer cette fois de stabiliser le système de krigeage (sans pour autant s'affranchir des poids négatifs).

Remarque importante: la question, bien naturelle, de savoir si une telle manipulation peut être légitime, sera examinée au dernier paragraphe. Pour le moment, je m'intéresserai seulement au mécanisme par lequel on peut agir sur le signe des pondérateurs.

## 8 - PREMIERE REPNSE : L'EFFET DE PEPITE

La régularité d'un modèle est caractérisée par son comportement à l'origine. Une première idée serait donc d'infléchir l'analyse structurale afin de refuser les modèles jugés particulièrement dangereux. L'expérience montre, par exemple, qu'on peut facilement remplacer un modèle gaussien (donc: indéfiniment dérivable), par un modèle "polynomial" très voisin, mais de comportement seulement cubique à l'origine (Chilès, 1977).

Une autre idée, beaucoup plus radicale, consiste évidemment à rajouter un effet de pépité. On constate expérimentalement que, même faible, un tel effet de pépité suffit la plupart du temps à éliminer les pondérateurs négatifs.

Le théorème déjà cité de (Matheron, 1986; cf. ci-dessus Section E-1) montre le rôle privilégié joué par l'effet de pépité, qui introduit une discontinuité à l'origine. A contrario, ce théorème permet de prévoir que le choix entre deux modèles tous deux continus à l'origine, sera de moindre conséquence sur les signes des pondérateurs. Aussi, une première réponse pratique au problème des poids négatifs, est bien le rajout d'un effet de pépité.

On peut interpréter cette opération comme l'adjonction d'une erreur de mesure, et son effet sur les pondérateurs négatifs s'apparente à la levée de l'effet d'écran (Matheron, 1971, pp. 134-135). Pour employer une terminologie empruntée à la théorie des splines, on passe ce faisant du point de vue interpolation au point de vue ajustement (Laurent, 1972, chapitre 4).

C'est une démarche connue, mais naturellement, la valeur de l'effet de pépète reste presque totalement arbitraire.

## 9 - DEUXIEME REPONSE : LES CONDITIONS D'UNIVERSALITE

Une autre approche possible du problème consiste à retrancher du modèle stationnaire les comportements trop réguliers. Plaçons-nous pour commencer dans une optique KU. On demande donc à la dérive (dérive au sens précis du KU) de prendre explicitement en compte la partie régulière du phénomène; on n'a donc plus simultanément sur les résidus les exigences - si difficiles à concilier - de stationnarité et de régularité.

Algébriquement, cela signifie simplement qu'on rajoute au système des conditions d'universalité. La structure de la dérive est cette fois choisie en connaissance de cause. Bien sûr, il faut supposer que l'analyse structurale prend en compte ces hypothèses nouvelles; de sorte que le modèle sous-jacent doit être moins régulier que si l'on en était resté au Krigeage Simple.

Notons que cette "deuxième réponse", contrairement à la première, n'a pas pour effet l'élimination des pondérateurs négatifs: probablement, on aura au contraire beaucoup plus de chances de voir surgir de tels pondérateurs. Car cette fois, il y a explicitement une dérive dans le modèle, et le moyen de la faire apparaître au niveau des estimateurs est précisément l'emploi de poids de krigeage négatifs.

Par contre, la stabilité du système est accrue, et c'est ce qui est important. On s'en assure en utilisant le théorème d'additivité (Matheron, 1969, pp. 13-16; Matheron, 1971, pp. 166-168):

- l'estimation des résidus se fait par un Krigeage Simple, mais avec un modèle structural plus irrégulier que si l'on n'avait pas rajouté par ailleurs une dérive. Aussi, on est fondé à penser qu'il y a moins de chances d'observer des poids négatifs au cours de cette estimation. Purgés de leur composante régulière, les résidus interpolés ne doivent plus manifester de "tendance".
- par contre, l'estimation de la dérive doit assumer la totalité de la partie régulière du phénomène. On est donc presque assuré de rencontrer ici des poids de krigeage négatifs. Mais cette fois, il ne s'agit plus de l'ajustement d'un polynôme de degré très élevé: la forme et le degré de cette dérive ont été choisis explicitement et en connaissance de cause. Aussi peut-on espérer une stabilité satisfaisante de l'estimation.
- l'estimation de la variable elle-même (résidu + dérive), présente bien une possibilité accrue d'obtenir des poids de krigeage négatifs. Mais ces poids n'ont plus le même sens que dans le cas du Krigeage Simple, dans la mesure où le phénomène qu'ils tendent à décrire est cette fois maîtrisé par l'utilisateur. Aussi, la stabilité globale du système de krigeage devrait être assurée.

Naturellement, cette comparaison des stabilités entre les systèmes du KS et du KU pourrait être envisagée d'un point de vue purement algébrique, sans interprétations géostatistiques: par comparaison des conditionnements des matrices de krigeage. Ce n'est pas l'objet des présentes réflexions; mais une telle approche devrait également être envisagée si on voulait se livrer à une critique approfondie du krigeage.

L'addition de conditions d'universalité incite à adopter le point de vue des FAIk (Mathéron, 1973, pp. 439-468). L'analyse structurale permet alors l'accès à des modèles plus riches, et plus rigoureusement adaptés au formalisme non stationnaire: les covariances généralisées. Mais du point de vue purement algébrique, le mécanisme de stabilisation du système de krigeage est exactement le même que dans l'optique du KU.

Au niveau de l'interprétation, le point de vue des FAIk peut parfois apparaître déroutant au praticien; sans doute d'ailleurs s'agit-il là d'une impression injustifiée (Chauvet, 1986). Il est vrai que, dans la mise en oeuvre des FAIk, il n'est plus possible de répartir la variable étudiée entre une dérive et un résidu: cela nécessiterait de faire des hypothèses supplémentaires qui impliqueraient le choix d'une représentation.

Remarque: ce sont d'ailleurs ces hypothèses qui, sous une forme implicite et en général non critique, fondent le formalisme du KU et en orientent l'analyse structurale.

Mais en réalité, cette dichotomie n'est en rien nécessaire: on sait même qu'elle présente plus de risques que d'avantages (Chilès, 1979). Aussi, une interprétation "neutre" du phénomène de stabilisation du krigeage, sera:

*"Par le choix de l'ordre  $k$  de la FAI, on contrôle le degré de régularité de l'estimateur".*

et c'est cette fois le passage aux combinaisons linéaires autorisées d'ordre  $k$  (CLA) qui filtre la composante "trop" régulière de la variable régionalisée.

Naturellement, le choix de l'ordre  $k$  n'est pas entièrement libre:

- il existe d'abord surtout une limite inférieure, en dessous de laquelle les CLA ne peuvent plus être considérées comme stationnaires. C'est d'ailleurs dans la détermination de cette limite, que se manifeste le mieux la supériorité opératoire des FAIk sur le KU.
- il existe en fait également une limite supérieure, cette fois dictée par le jugement du praticien. La théorie en effet ne pose pas de limite supérieure à l'ordre  $k$ : toute FAIk peut être considérée comme une FAI( $k+p$ ) d'ordre supérieur. Mais le choix d'un ordre trop élevé revient à supposer à la variable régionalisée une régularité tout arbitraire; et le risque resurgit d'avoir des pondérateurs très contrastés (très négatifs ou très supérieurs à 1), et par suite des estimateurs très erratiques.

Mais, au niveau pratique, il sera parfaitement légitime de choisir l'ordre, à l'intérieur de cet intervalle, en fonction des propriétés que l'on souhaite pour l'estimateur. Pour le problème examiné ici, comme précédemment en KU, le but sera d'exclure les traits structuraux trop réguliers du modèle probabiliste (ici: de la covariance généralisée).

## 10 - CONCLUSIONS PROVISOIRES

De l'examen de la genèse des pondérateurs négatifs, deux conclusions au moins peuvent être déjà tirées:

- a) on a résolu la contradiction entre les propositions (3) de BARNES et JOHNSON, et (5) de PHILIP et WATSON. Il apparaît à l'évidence que les effets de dérive locale peuvent exister, du moins si le modèle structural s'y prête. Ces effets peuvent tout aussi légitimement être déplorés par BARNES et JOHNSON ("unsatisfactory"), qu'appréciés par JOURNEL ("very positive"); mais ils ne peuvent être niés.
- b) il y a, au plan pratique, possibilité de choix dès l'étape de caractérisation des modèles, suivant que l'on redoute ou au contraire souhaite des pondérateurs négatifs. Ce choix portera essentiellement sur les conditions d'universalité (ou encore: sur le degré de FAIk), et sur l'éventuelle présence d'un effet de pépète.

C'est précisément un des avantages du krigeage, que de faire précéder l'estimation elle-même d'une phase d'analyse structurale: on peut, avant même d'effectuer l'estimation, prévoir au vu de la géométrie de l'information si de tels effets sont probables compte-tenu du modèle structural. Si, comme BARNES et JOHNSON, on juge ce risque inacceptable, c'est à cette étape qu'il convient d'y remédier.

## F - DES PONDERATEURS NEGATIFS AUX ESTIMATEURS NEGATIFS

### 1 - REMARQUE METHODOLOGIQUE PREALABLE

Il est significatif que les remarques de BARNES et JOHNSON s'expriment en termes d'estimateurs: les pondérateurs de krigeage ne sont que des intermédiaires de calcul qui la plupart du temps ne sont pas explicités; au contraire, ce sont sur les valeurs estimées que vont porter les critiques du praticien. Toutefois, au risque de me répéter, je voudrais revenir sur un point qui me semble essentiel au plan méthodologique.

Un estimateur linéaire est construit par combinaison de deux facteurs:

- les données elles-mêmes, qui ont un caractère indiscutable.
- les pondérateurs, qui sont issus d'un modèle et d'un mode opératoire, et qui de ce fait procèdent de choix et sont donc criticables.

Si donc une contestation surgit à propos de la valeur particulière prise par un estimateur, c'est malgré tout au niveau des pondérateurs que doit porter la critique. L'apparition d'une teneur estimée négative peut bien sûr très légitimement être considérée comme un événement intolérable; elle peut tout aussi légitimement d'ailleurs être considérée comme un événement anodin: mais, au delà des jugements de valeur, elle n'est de toute façon que la manifestation spectaculaire - mais plutôt superficielle - d'un problème bien plus significatif.

Plus précisément: lorsque - disons - un estimateur négatif "intolérable" apparaît, il ne fait qu'attirer l'attention sur une difficulté qui existe de toute façon. Même sans la coïncidence d'un pondérateur négatif et d'une donnée forte, cette difficulté aurait été présente. Elle aurait pu être révélée par la forme des courbes de niveau; elle aurait même pu être totalement invisible, et ne se traduire que par un mauvais conditionnement sans conséquence du système de krigeage. Mais le problème de fond serait demeuré strictement le même.

Aussi, le praticien qui veut rester rigoureux a le choix entre deux attitudes, tout aussi respectables l'une que l'autre:

- ou bien, il considère ces difficultés comme bénignes, et il en accepte d'avance toutes les conséquences.
- ou bien au contraire, il les considère comme inacceptables, et il doit les prévenir sur l'ensemble de son étude.

## 2 - REMARQUE SUR LE CHANGEMENT DE SUPPORT

BARNES et JOHNSON se placent dans un contexte minier, et la formulation qu'ils proposent pour le Krigeage Positif correspond explicitement à une estimation de blocs. Cela signifie que leurs observations préalables (1) à (3) portent également sur des résultats de krigeage étendu (teneurs estimées sur des blocs).

Dans une telle optique, et en plus des phénomènes de lissage qui seront examinés ci-dessous (Section F-3), on doit donc tenir compte de l'effet de changement de support. Ce changement de support introduit toujours un lissage supplémentaire, souvent considérable. Rappelons d'ailleurs que c'est pour corriger les biais dus à ce lissage, que KRIGE a élaboré la technique du krigeage, dès le début des années 1950. Aussi, comme on sait que les teneurs de blocs doivent être à la fois plus structurées dans l'espace et moins dispersées statistiquement que les teneurs des données, l'existence de phénomènes indésirables (teneurs négatives) ou mal maîtrisés (valeurs erratiques) prend une signification bien plus forte que s'il s'agissait d'une estimation ponctuelle.

Cependant, je reste ici dans le cas d'une estimation ponctuelle. Il est ainsi possible de comparer directement les propriétés des données et celles des valeurs estimées. De plus, ces propriétés peuvent être facilement rattachées à des caractéristiques de tracés de cartes (cf. Section G). Si ultérieurement il fallait quand même réaliser un changement de support, on pourrait de toute façon passer simplement des estimations ponctuelles aux estimations étendues par linéarité du krigeage (Matheron, 1971, pp. 129-130). Mais cette opération n'ajoute rien au fond, elle ne fait que renforcer les conclusions; par exemple:

- du point de vue théorique, une teneur négative révèle des problèmes de stabilité d'autant plus importants qu'elle a été estimée sur un volume plus grand;
- du point de vue du mineur, une teneur estimée négative sera considérée comme d'autant moins acceptable qu'elle sera affectée à un volume plus grand.

## 3 - INSTABILITE DE L'ESTIMATEUR SELON LES PONDERATEURS

Quelle que soit la manifestation locale de l'existence de pondérateurs négatifs: estimateur négatif ou supérieur au maximum des données, courbes de niveau excessivement fluctuantes, etc., cet effet sur les estimateurs peut se résumer en un seul mot: instabilité. C'est, très exactement, le sens de la proposition (3) de BARNES et JOHNSON. Précisons maintenant les mécanismes de cette instabilité.

## a – Pondérateurs positifs de somme égale à 1

Examinons d'abord le cas le plus fréquent où la somme des pondérateurs est égale à 1: c'est par définition le cas du Krigeage Ordinaire, ainsi que du krigeage non stationnaire le plus courant. Si de plus tous ces pondérateurs sont positifs, l'estimateur se présente comme une régularisation des données (en termes intuitifs: une "moyenne pondérée", ou même: une "convolution" des données). C'est ce que l'on traduit en disant que le krigeage lisse le phénomène: les valeurs krigées ont un variogramme plus régulier que celui des données initiales, un histogramme moins dispersé, etc... Cet effet de lissage peut être considérable si on a utilisé un voisinage de krigeage contenant un grand nombre de données.

Dans l'hypothèse stationnaire, lorsque la variable est très peu structurée, les pondérateurs de krigeage sont alors à peu près tous égaux, et sont appliqués à des données par hypothèse peu corrélées: le lissage peut ainsi être interprété comme une conséquence du théorème de la limite centrale. Par suite, les valeurs krigées ont une distribution plus proche de la distribution normale que les données utilisées pour le krigeage. Cet effet de "normalisation" peut s'observer la plupart du temps, y compris sur des variables plus structurées, et peut prendre une ampleur considérable (Chauvet, 1987).

A propos de cet effet de lissage, notons une nouvelle fois au passage la relativité des jugements de valeur:

- en cartographie, ce lissage sera presque sûrement considéré comme chose souhaitable, car il conduira à de "belles" cartes.
- dans le domaine minier au contraire, ce lissage occulte les variabilités locales, et rend très hasardeuse la sélection sur krigeage (introduction de biais): il faut alors recourir à la technique des simulations conditionnelles (Journel, 1974).

## b – Pondérateurs positifs de somme quelconque

On peut imaginer des situations où un Krigeage Simple produit des pondérateurs tous positifs, mais de somme supérieure à 1. Dans ce cas, il n'y aura plus lissage, mais au contraire amplification des éventuelles fluctuations des données: en ce sens, on pourra dire que l'estimateur est plus irrégulier que ces données.

Il s'agit là d'un cas d'apparence paradoxale. On aurait envie de dire que l'estimateur est plus "variant" que les données; et pourtant, s'agissant d'un Krigeage Simple, la variance de krigeage est nécessairement inférieure à la variance de ces données.

Si cette contradiction apparente peut être jugée gênante en pratique, elle ne révèle pas d'incohérence théorique. Car une telle situation ne peut se présenter que sous certaines conditions, liées au caractère de convexité de la fonction de covariance dans le voisinage de krigeage. Concrètement, cela signifie que de tels pondérateurs ne peuvent surgir que si la structure est assez régulière à l'échelle de travail. Un tel modèle est alors extrêmement exigeant avec les données; ou, en termes moins anthropocentriques, il y a des chances accrues qu'un jeu de données observées localement, se révèle difficilement compatible avec le modèle global.

Notons que, pour gênante qu'elle puisse être au plan pratique, cette circonstance n'est pas la plus paradoxale que l'on puisse envisager. On peut en effet (Matheron, 1986) proposer des exemples où tous les pondérateurs de krigeage soient négatifs...

### c - Modèle global, caractéristiques locales

La discussion ci-dessus illustre bien une caractéristique de la géostatistique linéaire, qu'il convient peut-être de rappeler maintenant. Le modèle structural est ajusté globalement, fondé sur l'ensemble des données disponibles. Ne prenant pas en compte la distribution des variables, il ne peut a fortiori produire une estimation conditionnée par les caractéristiques locales.

Le cas le plus favorable est celui d'une variable stationnaire multigaussienne. La fonction de covariance à elle-seule caractérise l'ensemble du phénomène, et le krigeage s'identifie à l'estimateur idéal, c'est-à-dire l'espérance conditionnelle. Par contre, dès que l'on s'"éloigne" du cas stationnaire multigaussien, la classe des estimateurs linéaires ne contient plus cet estimateur idéal et, dans des cas de distributions extrêmes (atomes, dissymétries, etc...), il faudra avoir recours à des techniques non linéaires même pour résoudre des problèmes de simple estimation ponctuelle.

Ainsi, ce n'est que dans le cas gaussien que l'on peut donner une signification conditionnelle à l'estimateur du krigeage. En particulier, dans le cas général, l'erreur d'estimation n'est pas une erreur conditionnelle, et la variance d'estimation n'est pas une variance conditionnelle.

Autrement dit, il est parfaitement admissible au plan théorique qu'apparaissent des anomalies locales. Mais il ne s'agit pas de biais, au sens précis où cela a été défini en Krigeage Universel; en particulier, sur l'ensemble du champ, la moyenne des erreurs d'estimation sera proche de 0. Mais on peut bien sûr observer localement des effets indésirables, les teneurs estimées négatives n'en étant qu'un exemple. Le krigeage, outil global, ne saura pas toujours "se défendre" contre des particularités locales, qu'il s'agisse de données "anormales" ou d'irrégularités dans la géométrie de l'information.

Une leçon pratique peut être tirée de ces remarques. Car, si l'on n'a aucun moyen d'action sur les valeurs des données, on peut en revanche veiller à disposer d'une structure homogène de l'information - l'idéal étant bien sûr une maille régulière: celle-ci est en effet la configuration optimale tant pour des raisons théoriques (on est assuré que la moyenne expérimentale des erreurs d'estimation doit être très proche de 0) que pour des raisons informatiques (les pondérateurs sont calculés une fois pour toutes). Ces considérations devront être prises en compte lors des discussions des divers moyens d'éliminer les pondérateurs négatifs (Section H).

### d - Pondérateurs négatifs

Lorsque certains pondérateurs de krigeage sont négatifs, le phénomène de régularisation peut être atténué. A la limite, dans un cas extrême comme l'extrapolation en KS avec une covariance gaussienne, il s'agirait plutôt d'une véritable déstabilisation des données. Cet effet

s'interprète aisément: avec des pondérateurs alternés, le krigeage a plutôt cette fois le sens d'une différence finie.

Un exemple simple est celui des doublets de points: si deux données sont très voisines l'une de l'autre, et approximativement alignées avec le point à estimer, elles tendent à ne plus être distinguées par le processus d'estimation: elles interviennent grosso modo par leur valeur moyenne (facteur de stabilisation) et par leur différence (une approximation du gradient: facteur d'instabilité). Naturellement, au point de vue théorique, le système de krigeage reste régulier tant que les deux données sont distinctes; par contre, dans la pratique, sa stabilité est considérablement compromise.

## e - Conclusion provisoire sur l'instabilité

Les remarques des paragraphes précédents ne sont évidemment que des interprétations, et doivent dans la pratique être examinées en détail, compte-tenu du modèle structural adopté et du choix du voisinage de krigeage. Mais elles aident à comprendre un effet très général, qui est bien connu des cartographes, et qui semble paradoxal aux néophytes:

*"pour un jeu de données fixé, et pour un choix fixé de voisinage de krigeage, le tracé des courbes de niveau est plus erratique pour un modèle structural régulier, et plus stable pour une modèle structural irrégulier."*

Il s'agit bien sûr là d'un résultat empirique qui connaît des exceptions, non d'un théorème! Afin maintenant de préciser ce phénomène en lui gardant son caractère le plus spectaculaire, je vais me placer dans le prochain paragraphe du point de vue de l'interpolation, c'est-à-dire que le krigeage sera là encore supposé ponctuel. Pour interpréter les résultats en termes d'estimation minière (donc de krigeage étendu), il suffirait de se remémorer la remarque faite ci-dessus (Section F-2) sur l'effet de changement de support.

## G - LE POINT DE VUE DE L'INTERPOLATION

Remarque préalable: j'abandonne ici la restriction de stationnarité d'ordre 2 adoptée précédemment. L'interpolation qui est commentée ci-dessous, peut être indifféremment un krigeage avec ou sans conditions d'universalité. Je suppose cependant, pour simplifier la discussion, que les fonctions de base éventuelles sont des monômes - ce qui est d'ailleurs le cas le plus courant. Les méthodes plus particulières telles que Fonctions Aléatoires Intrinsèques Généralisées (Matheron, 1979), ou surtout krigeage en présence d'une Dérive Externe (Delfiner, Delhomme et Pelissier-Combescure, 1983), pourraient faire l'objet d'un examen particulier.

## 1 - REGULARITE DU MODELE ET DE LA FONCTION D'INTERPOLATION

Le concept de régularité est sans ambiguïté au niveau théorique. La régularité est en effet définie rigoureusement par les caractères de continuité ou de dérivabilité du modèle structural (covariance, variogramme ou covariance généralisée). Si l'on considère le krigeage comme une fonction d'interpolation, c'est-à-dire si on examine les propriétés analytiques de l'estimateur de krigeage considéré comme un fonction de l'espace euclidien, il est clair que ses propriétés de régularité seront liées à celles du modèle. En fait, comme on a supposé que les éventuelles fonctions de base sont monomiales, les propriétés de continuité et dérivabilité de la fonction d'interpolation seront exactement les mêmes que celles du modèle structural. Cela apparaît immédiatement dans l'expression de l'estimateur de krigeage sous sa forme duale (Matheron, 1981a).

Toutefois, du point de vue cartographique, cette régularité au sens mathématique ne représente pas grand chose concrètement. Car, ne serait-ce que pour des raisons informatiques, on a toujours recours à une discrétisation pour tracer les cartes. Le comportement infinitésimal de l'estimateur est alors évidemment oublié.

Remarque: en parlant ici de discrétisation, je ne pense pas nécessairement à l'utilisation intermédiaire d'une grille d'estimation, en vue du tracé effectif des courbes. Il est vrai qu'une telle pratique est courante, pour des raisons de simplicité. Mais rien n'interdit de tracer directement les courbes de niveau, en suivant les points de valeurs krigées égales (Galli, Murillo et Thomann, 1984, pp. 621-634); rien non plus n'interdit de tracer directement le graphe de la fonction d'interpolation. Mais dans tous les cas, quel que soit le degré de résolution de tels tracés, on finit toujours par travailler sur un nombre fini de points.

Finalement, au point de vue graphique, la seule propriété mathématique du modèle qui puisse avoir un effet immédiat sur le tracé d'une carte, est la discontinuité. On sait en effet que, en présence d'un effet de pépite, l'estimateur présente des discontinuités (au sens mathématique) aux points de mesure. Si donc une donnée correspond à l'une des isovaleurs que l'on veut tracer, on peut avoir l'impression que la courbe de niveau ne respecte pas cette donnée. On voit se dessiner ici le fossé qui sépare les points de vue théorique et appliqué:

- au niveau théorique, le krigeage reste essentiellement un interpolateur exact, qui restitue donc parfaitement les données - fût-ce au prix d'une discontinuité des courbes de niveau.
- au niveau pratique, au contraire, le krigeage peut paraître manquer à la plus importante propriété des interpolateurs: l'exactitude.

Ce hiatus n'a rien de rédhibitoire: le passage à la pratique s'accompagne nécessairement d'éléments d'appréciation personnelle, de jugements de valeur. Tel attachera plus d'importance à l'exactitude, tel autre à la régularité (au sens intuitif) de sa carte; et aucun n'aura "plus raison" que l'autre. La seule erreur serait d'exiger simultanément toutes ces propriétés.

## 2 - L'ECHELLE DE TRAVAIL

Compte-tenu des remarques précédentes, il est donc clair qu'il ne peut exister de définition "passe-partout" de la régularité au plan pratique. En particulier, se limiter à la définition purement mathématique est à la fois très restrictif, et peu réaliste:

- peu réaliste: à cause du nécessaire recours, mentionné plus haut, à une discrétisation qui élimine de toute façon les comportements infinitésimaux.
- très restrictif, surtout: parce que ces comportements infinitésimaux, par nature même, méconnaissent la notion d'échelle.

Ce dernier point est très important, mais n'a rien de nouveau. Pour le praticien, la régularité est associée à la notion de "comportement local", donc de "tendance". Encore faut-il donner un sens à cette tendance, et c'est cette définition qui impose le choix d'une échelle (Matheron, 1971, pp. 142-143).

Une première échelle que l'on peut mentionner est celle du champ étudié tout entier. De ce point de vue, la régularité de la carte est une propriété du phénomène lui-même, non de l'interpolateur: par exemple, on considèrera comme régulière la carte d'une plaine, et comme irrégulière celle d'une chaîne de montagnes.

On peut définir une seconde échelle, associée cette fois au mode d'interpolation. Ce sera ici le voisinage de krigeage, que l'on peut aussi désigner sous le terme plus général d'échelle de travail. (Naturellement, cette seconde échelle peut se trouver confondue avec la première, dans le cas d'un travail en voisinage unique). C'est alors très exactement à cette échelle de travail que sont supposés valides les résultats de l'analyse structurale, c'est-à-dire explicitement: l'ordre  $k$  de stationnarité d'une part, le modèle structural (covariance, variogramme ou covariance généralisée) d'autre part.

## 3 - REGULARITE DE LA CARTE

Plaçons nous donc à l'échelle d'un voisinage de krigeage. Le problème concret est de décrire l'allure effective de l'estimateur de krigeage à cette échelle. Intuitivement, il est aisé de distinguer les tracés qui ont un comportement "lisse", et ceux qui présentent une allure générale erratique, des courbures trop prononcées, des isthmes, etc... Cette notion n'a rien ni de nouveau, ni de choquant: au niveau d'un ajustement polynomial, on sait bien intuitivement qu'un polynôme de degré trop élevé aura un comportement trop "erratique", trop "fluctuant" - bref: trop "irrégulier" au sens proposé ici.

La seule difficulté est le caractère subjectif de cette définition de la régularité. C'est là bien sûr que le point de vue appliqué se démarque de la théorie. Selon les utilisateurs, un même résultat ne sera pas nécessairement jugé de la même façon - et ceci est parfaitement légitime dans les sciences naturelles. Le rôle d'une étude géostatistique (et des présentes réflexions) ne saurait être donc de révéler une Vérité (qui n'existe pas), mais de s'appuyer sur

la théorie pour établir le lien entre les résultats observés et les hypothèses de départ - et pour réajuster celles-ci au besoin.

On a choisi dans ce paragraphe le point de vue de l'interpolation. La raison en est que, dans ce cas, on peut dire pour simplifier que le produit final d'une étude est la carte, dont les propriétés peuvent être comparées de façon quasi immédiate à celles de la variable étudiée. Naturellement, et compte-tenu des précédentes remarques, il n'est pas question de définir ici ce qu'est une "bonne" carte: ce serait là attitude purement dogmatique; mais je me propose d'apporter quelques éléments permettant de critiquer certaines caractéristiques d'un tracé à la lumière des paramètres du krigeage. Et, en premier lieu, je voudrais préciser la conclusion déjà proposée à la Section F-3-e.

#### 4 - DUALITE : IRREGULARITE DU MODELE - REGULARITE DE LA CARTE

La notion de régularité d'une carte est donc celle définie au précédent paragraphe. Comme - on l'a vu - il ne s'agit pas d'une notion absolue, on doit garder l'énoncé de la Section F-3-e sous une forme comparative. On pourrait dire par exemple:

*"plus le modèle est irrégulier, et plus la carte est régulière".*

Tel quel, cet énoncé est évidemment provoquant. Rappelons d'abord qu'il ne s'agit nullement d'un théorème: un contre-exemple sera proposé au prochain paragraphe; il s'agit plutôt d'un avertissement à l'intention des praticiens de la géostatistique. Sous une forme moins péremptoire, on pourrait dire:

*"en règle générale, ce n'est pas en utilisant un modèle structural plus régulier que l'on arrivera à régulariser une carte, bien au contraire."*

Une telle propriété n'est paradoxale qu'en apparence. Ce n'est même pas une nouveauté: le cartographe manuel qui ne respecte pas exactement la valeur d'une mesure afin de garder une "belle" allure à une courbe de niveau, ne fait rien d'autre que rajouter un effet de pépite à son modèle structural implicite. La différence n'est que de vocabulaire: il parlera d'une "incertitude de mesure".

Le mécanisme par lequel le rajout d'un effet de pépite lisse la courbe, a été examiné aux Sections E-8, puis F-3. Mais ce n'était là qu'un cas extrême: on peut très bien envisager une évolution plus progressive de l'irrégularité d'un modèle; le mécanisme de lissage sera toujours le même. On peut ainsi reprendre l'exemple élémentaire déjà cité à la Section E-1.

Je me propose maintenant de revenir à l'exemple de la covariance gaussienne, déjà traité (du point de vue du signe des pondérateurs de krigeage) à la Section E-3. La question est maintenant de savoir comment la variation de portée se répercute sur la forme des courbes de niveau.

- lorsque la portée est grande par rapport au voisinage de travail, le phénomène est très structuré. Il apparaît des poids de krigeage négatifs, et les courbes de niveau peuvent être extrêmement erratiques. Plus précisément, la valeur de l'estimateur (et donc la forme des courbes de niveau) est considérablement affectée par la géométrie des données.

Une valeur particulière de donnée peut donc agir notablement sur la forme d'une courbe de niveau, aussi longtemps que cette courbe ne pénètre pas dans la "zone d'influence" d'une autre donnée. Les anomalies de tracé dues aux particularités locales des données, sont en quelque sorte "diluées" le long des courbes de niveau.

- à l'opposé, pour une portée très courte, c'est-à-dire un modèle très irrégulier, les valeurs particulières des données sont très peu différenciées par l'algorithme d'estimation. Les courbes de niveau sont donc fort peu affectées par les conditions locales. A la limite, pour une portée nulle (effet de pépite pur), la carte se borne à sa dérive (soit: 0, dans l'exemple présent), sauf aux points de données. Les anomalies sont ainsi cette fois focalisées aux points de mesure.

Cet exemple, proposé dans une optique KS, pourrait sans inconvénient être généralisé aux krigeages non stationnaires. En KU avec modèle sous-jacent stationnaire, l'interprétation serait absolument identique. En FAIk, ce serait une illustration expérimentale qu'il subsiste, dans le formalisme des covariances généralisées, des notions aussi concrètes que les "zones d'influence".

En fait, la dualité évoquée dans ce paragraphe est déjà parfaitement familière à tout praticien de la géostatistique. Je crois cependant qu'elle mérite quelques observations supplémentaires:

## 5 - DES POINTS DE VUE CONTRADICTOIRES

C'est bien sûr à dessein que je viens de reprendre l'exemple de la covariance gaussienne, pour souligner combien le point de vue "mathématique pur" peut s'écarter des préoccupations pratiques.

Quelle que soit sa portée en effet, la covariance gaussienne reste indéfiniment dérivable: cela n'a donc pas de signification mathématique de parler de sa plus ou moins grande régularité. Ainsi, comme c'était annoncé à la Section G-1, ce n'est pas la régularité au sens algébrique qui a par elle-même une importance décisive sur le tracé des cartes. Et par suite, ce ne sont pas non plus les propriétés algébriques (continuité, dérivabilité, etc...) de la fonction d'interpolation, qui déterminent par elles-mêmes les qualités (au sens subjectif) d'une carte. Et cette conclusion n'est en rien liée à une quelconque étape de discrétisation.

L'approche du krigeage en FAIk en propose d'ailleurs une illustration spectaculaire. Ainsi, pour une FAI-1 à deux dimensions, on sait que le modèle de covariance généralisée de type "spline plaque mince" conduit à de "belles cartes": et, en particulier, les courbes de niveau sont plus régulières que celles obtenues avec les modèles linéaire ou cubique. Or, ces deux derniers modèles sont, respectivement, moins et plus régulier (au sens mathématique) que le modèle "plaque mince": tel est le contre-exemple annoncé à la proposition de la Section G-4. Il apparaît donc clairement que c'est "autre chose" qui joue le rôle prépondérant dans la régularité des cartes; en l'occurrence, cet "autre chose" a une signification énergétique, mise en évidence par l'équivalence théorique entre les splines et le krigeage (Matheron, 1981b).

On peut maintenant faire le bilan des derniers exemples envisagés:

- a) les propriétés de régularité mathématique du modèle ont évidemment leur importance

- (exemple de l'effet de pépite; cas du schéma monomial en fonction de son degré),
- b) mais ne sont pas déterminantes (exemple du schéma "spline plaque mince" comparé aux schémas linéaire et cubique),
- c) tandis que d'autres facteurs peuvent intervenir de façon décisive (comme la portée de la covariance gaussienne).

Ainsi, il ne semble pas possible d'adhérer aux conclusions présentées dans (Sibson, 1981, p. 22):

*"A continuously differentiable function is visually smooth and (implicit function theorem) has smooth "contour lines" (an important application area). Functions which are not continuously differentiable definitely do not look smooth."*

parce que le mot "smooth" reste ici limité à une signification purement mathématique. Au contraire, il ressort du présent paragraphe que le géostatisticien doit adopter un point de vue de physicien, c'est-à-dire qu'il doit prendre explicitement en compte les "conditions d'expérience". Concrètement, cela signifie surtout que la notion d'échelle doit être introduite non seulement lors de la critique de la carte (cf. Section G-3), mais également au niveau du modèle. A ce propos, une limitation inévitable est imposée par la maille d'information: pour une échelle inférieure à cette maille, les caractéristiques du modèle ne sont pas des vérités qui ont été révélées par l'analyse structurale, mais des CHOIX constructifs qui engagent la responsabilité du géostatisticien, et qui doivent être adaptés au problème particulier qu'il a à résoudre.

Cette approche physicienne pourra être complétée par l'examen d'un autre type de dualité dans le tracé des cartes, cette fois entre régularité et précision (Section G-7); d'autre part, on rappellera les liens étroits entre krigeage et splines (Section G-8), ainsi que l'intérêt de joindre une carte des isovariances aux courbes de niveau (Section G-9). Mais auparavant, il semble intéressant de préciser la position du krigeage par rapport à d'autres méthodes d'interpolation, en essayant de déceler la source de certains malentendus (Section G-6).

## 6 - A LA SOURCE D'UN MALENTENDU ?

On peut lire également dans (Sibson, 1981, p. 22):

*"Our interpolant must satisfy a number of conditions which among other things make it relate naturally to the data and make it reasonably smooth, so that there is a good chance that it looks and behaves very much like the original function from which the data were drawn".*

Comme il s'agit d'une déclaration de principe, et non de la définition de critères précis, il est bien naturel que cet énoncé reste subjectif: "naturally", "reasonably", "good chance", traduisent des opinions, non des concepts mathématiques rigoureux. En ce sens, il semblerait donc que cette position s'inspire du même pragmatisme que celui qui guide le géostatisticien pratique.

Et pourtant, en une seule phrase, on voit apparaître trois affirmations avec lesquelles le krigeage, comme mode d'interpolation de données réelles, ne peut pas et ne doit pas être d'accord:

### a - Précision et irrégularité

La première contrainte que le krigeage ne saurait satisfaire, serait de respecter la simultanéité:

"... make it relate naturally to the data AND make it reasonably smooth ..."

Si l'on interprète l'expression "relate naturally" comme un souci d'exactitude et de précision, il y a là un a priori que rien ne justifie. Cette contrainte semble signifier en effet que la variable que l'on veut interpoler est nécessairement régulière. Sans aller jusqu'à parler d'objets fractals, on sait bien au contraire que les phénomènes naturels sont souvent caractérisés par une extraordinaire irrégularité: quel sens mathématique pourrait-on donner, ne serait-ce qu'à la notion de continuité d'une teneur dans un gisement ? Même au niveau des moyennes locales, les comportements de variables physiques (teneurs, cotes topographiques, températures, pressions, etc...) sont souvent - au sens intuitif - très discontinus.

Il est parfaitement légitime d'exiger d'abord d'un interpolateur la régularité - sous réserve d'avoir au préalable défini précisément cette notion: ce sera par exemple le point de vue des splines (cf. Section G-7). Il est tout aussi légitime, avec la même réserve, d'en attendre d'abord la précision: c'est par exemple le choix du krigeage. Mais vouloir imposer simultanément ces deux contraintes revient à forcer la nature, à moins que l'on n'ait choisi pour la précision et la régularité des définitions si peu exigeantes, que tout se passe finalement comme si aucune contrainte n'avait été imposée.

### b - Estimation et simulation

Une seconde contrainte complète la précédente:

"... so that there is a good chance that it LOOKS and BEHAVES very much like ..."

En plus de l'a priori de régularité qui était déjà présent dans la première contrainte, on trouve ici l'idée supplémentaire que l'interpolateur doit reproduire le comportement de la variable étudiée. En termes géostatistiques, cela signifierait en particulier que l'interpolateur doive respecter le variogramme de la variable.

Une telle exigence est en soi légitime. Mais alors, elle oriente à elle seule le choix de l'algorithme, et sans doute n'y a-t-il plus lieu de continuer à nommer "interpolation" le résultat obtenu. En Géostatistique, ce point de vue a conduit à la mise au point des Simulations Conditionnelles (Journel, 1974). Le lien théorique avec le krigeage en est bien sûr fort étroit; cependant, il s'agit là d'une technique autre, visant à répondre à d'autres questions. A l'évidence, Simulation Conditionnelle et Krigeages (quels qu'ils soient) ne peuvent jouer des rôles interchangeables.

Ainsi, une nouvelle fois, il n'est pas légitime d'exiger simultanément les deux propriétés. Le krigeage ne peut pas rendre compte de la structure de la variable (lissage: cf. Section F-3-a); la simulation conditionnelle est un mauvais estimateur. La recherche d'un outil passe-partout est évidemment à proscrire: à problèmes différents, solutions différentes.

Remarquons enfin que cette seconde critique est renforcée par la précédente: vouloir obtenir simultanément, pour un même "interpolateur", la précision ET le respect de la structure ET la régularité est bien sûr utopique. Mais chaque problème pris séparément peut être résolu, par exemple:

- par le krigeage, si on cherche la précision (du moins, si la précision est définie en termes de variance non conditionnelle);
- par les simulations conditionnelles, si on cherche le respect de la structure;
- par les splines, si on cherche la régularité.

### c - Une explication des divergences ?

Le dernier point éclaire peut-être l'origine du malentendu:

"... the original FUNCTION from which the data were drawn."

Naturellement, dans les sciences de la nature, il n'y a pas de fonction cachée dont les valeurs prises en quelques points particuliers seraient nos données expérimentales. De même, dans l'optique de la Géostatistique, il n'y a pas de fonction aléatoire dont la Variable Régionalisée serait une réalisation: dès les débuts de la géostatistique (Matheron et Formery, 1962, p. 262), l'attention a été attirée sur ce point. La quête d'une telle fonction sous-jacente, déterministe ou aléatoire, relèverait donc au sens strict de la métaphysique.

Autrement dit, la démarche de l'interpolation (ou: de l'"estimation", de la "prédiction", le mot par lui-même est ici sans importance) en sciences naturelles, ne peut pas satisfaire dans le même temps à la totalité des desiderata exposés par SIBSON. Le "bon" interpolateur tel qu'il est défini au début de ce paragraphe, est donc à prendre au sens purement algébrique, au même titre que les différents interpolateurs rencontrés par exemple en calcul des différences finies (cf. par exemple Jordan, 1965). Ce serait une exigence utopique, donc une erreur, que d'attendre toutes les "bonnes" propriétés d'un seul et unique algorithme.

Et, ce qui est encore pire, c'est cette fois faire preuve de dogmatisme que prétendre être seul détenteur des "vrais" critères de qualité. Prenons à titre d'exemple le cas de la cartographie sous-marine: le capitaine de sous-marin légitimement soucieux de ne pas s'échouer exigera naturellement un maximum de précision; et l'éditeur de livres de vulgarisation désirant s'en tenir aux grandes lignes demandera au contraire une carte régulière. Aucun n'a "plus raison" que l'autre; et suite à une utilisation "passe-partout" du krigeage, le premier se plaindra d'avoir une carte trop lissée, et le second d'avoir des courbes de niveau trop irrégulières.

Un développement purement mathématique est - simplement - soit juste, soit faux. Au contraire, une application pratique exige une étape critique. (Cette étape peut d'ailleurs révéler que l'algorithme choisi était inadéquat, mais cela ne remet pas en cause la méthode). Et une source essentielle de malentendus est donc, comme on vient de l'entrevoir dans ce paragraphe, la confusion entre les propriétés objectives de la variable étudiée (appréhendée par des mesures

éventuellement entachées d'incertitude, relatives à un certain support, connues de façon discrète, etc...) et les propriétés mathématiques du modèle (dérivabilité, par exemple).  
Proposons un dernier exemple:

#### d - Un exemple typique

On lit dans (Philip et Watson, 1986b, p. 505):

*"Kriging, as an interpolation method, fails in the first and most important of Sibson's requirements, viz. the surface is not continuously differentiable. In general, it forms cones at data points."*

deux phrases qui méritent sans doute quelques mots de commentaires.

- 1° On a vu (Section G-1) que les comportements infinitésimaux de l'interpolateur sont de toute façon masqués par le recours obligé à une discrétisation. A priori donc, ce "most important requirement" est au mieux un raffinement mathématique.
- 2° De surcroît, il a été établi (Section G-4) que la régularité mathématique est en général plutôt associée à une certaine irrégularité de la carte. Il faut donc choisir si on souhaite vérifier une propriété théorique du modèle, ou bien satisfaire à une propriété pratique du produit final (donc: de la carte). Les deux simultanément peuvent être inconciliables.
- 3° Si les "cones at data points" sont à prendre au sens mathématique, c'est-à-dire s'il s'agit très précisément de points de rebroussement, la remarque de PHILIP et WATSON est évidente. Plus généralement, on observera aux points de données des irrégularités exactement de même nature que celles du modèle structural à l'origine, mais ce seront des irrégularités de nature mathématique: on est ramené à l'observation du 1°.
- 4° Si par contre "cones at data points" signifie des infléchissements (à l'échelle de travail) des courbes de niveau au voisinage des données, la remarque de PHILIP et WATSON devient beaucoup plus inquiétante, et rejoint la discussion initiale sur les pondérateurs négatifs:

Le mécanisme qui suscite ces comportements a été examiné aux Sections E-3 à E-6, puis G-4. Il est d'autant plus fort que la variable est plus structurée, et restitue - même en FAIK - la notion de zone d'influence. Quand on se trouve loin de toute donnée, l'interpolateur tend à décrire le comportement de la dérive; quand au contraire on est proche d'une donnée, il cherche à tenir compte des caractéristiques locales. Il est parfaitement exact que le krigeage n'a ainsi pas de signification homogène, ce qui est d'ailleurs confirmé par l'étude de la carte d'isovariances (cf. Section G-9): mais ceci n'a rien d'anormal, bien au contraire, puisque l'algorithme du krigeage est fondé sur un critère local (minimisation de la variance). Il serait donc extrêmement inquiétant que le voisinage des données ne soit pas privilégié par une méthode qui cherche la précision. Le plus paradoxal dans le point de vue de PHILIP et WATSON, et dans leur contestation du krigeage en général, est qu'ils se réclament de SIBSON (Philip et Watson, 1986b, p. 505):

*"Sibson goes on to list eight desirable properties of an interpolant, not one of which is contained in kriging."*

alors que l'on trouve précisément dans (Sibson, 1981, p. 22) comme septième "desirable property":

*"The interpolant should be localized, in that in some suitable sense only data sites which are reasonably near neighbours should influence the interpolated value at a point".*

## e - Conclusion partielle

Il est bien sûr légitime de s'attacher prioritairement aux propriétés esthétiques d'une courbe. Mais si pour ce faire on met en oeuvre un algorithme qui privilégie la recherche de précision, il ne faut naturellement pas s'étonner que le résultat ne réponde pas à l'attente. L'utilisateur doit d'abord s'en prendre à lui-même; et c'est précisément le fondement d'une bonne étude géostatistique, que de rechercher l'adéquation de la méthode employée au problème posé. Vouloir résoudre tous les problèmes à la fois procède au mieux d'une grande naïveté; prétendre l'avoir fait devient résolument suspect.

Les réflexions du présent paragraphe sont évidemment beaucoup moins éloignées qu'il n'y pourrait paraître de la question initiale des pondérateurs négatifs. Outre en effet que les comportements erratiques sont souvent associés à de tels pondérateurs, la démarche que doit adopter l'utilisateur face à une situation qui ne lui convient pas est exactement la même. En résumé, le savoir-faire du géostatisticien doit trouver un équilibre entre le respect des données d'une part, et la satisfaction de ses desiderata d'autre part; il faut donc se défier simultanément des modèles imposés a priori (les données n'interviennent plus dans leur spécification), comme des procédures trop automatiques (c'est cette fois l'utilisateur qui n'a plus l'initiative).

Cet équilibre n'est pas toujours facile à obtenir; il peut même parfois se révéler impossible. Au moins le krigeage ne prétend-il pas quant à lui être une panacée, et laisse-t-il à l'utilisateur le libre usage de son simple bon sens et de son esprit critique. Le géostatisticien doit être responsable. Sans doute faut-il voir d'ailleurs dans ce fait l'origine de certains jugements plus polémiques que scientifiques sur la géostatistique: il est certes plus facile d'incriminer une méthode que de mettre sa propre responsabilité en cause, surtout si c'est pour reconnaître finalement que les exigences que l'on formule s'apparentent de fort près à la quadrature du cercle...

## 7 - DUALITE : REGULARITE-PRECISION

Le mécanisme de cette dualité apparaît en filigrane dans les paragraphes précédents. Rappelons simplement que, le krigeage se préoccupant en priorité de la précision, la régularité de la carte peut en être affectée.

Rappel important: on garde comme définition de la précision la valeur de la variance d'estimation. On a vu que d'autres définitions seraient envisageables (quoique de manipulation plus difficile), mais on reste ici dans le cadre de la géostatistique linéaire. Rappelons que cette variance d'estimation n'est pas une variance conditionnelle.

Naturellement, cela ne signifie pas qu'une carte régulière soit nécessairement imprécise. Cependant, dans les domaines des sciences de la terre où les variables peuvent être extraordinairement irrégulières, il est vrai qu'il est prudent d'avoir des doutes sur la précision d'une trop belle carte.

Cela signifie encore moins qu'il suffit qu'une carte soit irrégulière pour qu'elle soit précise. Une telle proposition serait cette fois totalement absurde. Du reste, on a vu précédemment que les irrégularités des courbes de niveau tiennent justement à l'hétérogénéité de la précision sur l'ensemble de la carte (Section G-6-d).

Par contre, si une carte a été obtenue par krigeage, elle est optimale en ce sens que la variance d'estimation est minimale en chaque point. On peut naturellement contester l'opportunité de ce critère, surtout dans le cas de variable très éloignée d'une distribution normale. Mais ce qui est sûr, c'est que par rapport à ce critère, la carte deviendra immanquablement moins précise si elle est lissée. Donc, le modèle étant par ailleurs supposé satisfaisant, il est clair qu'on ne rajoute pas de la précision en lissant la carte, bien au contraire. C'est en ce sens qu'il est possible de parler de dualité: précision-régularité. On peut également citer (Dubrule, 1986b):

*"I never wrote that complicated maps are more "accurate", which would be a ridiculous statement. I just tried to say that it is not BECAUSE a map looks nice that it is "accurate". By making this point, I wanted to warn geologists against evaluating a map on its sole cosmetic qualities."*

Naturellement, ces remarques ne doivent pas être comprises comme condamnant les lissages de cartes. Mais elles signifient ceci:

En lissant a posteriori une carte, on ne contribue pas à retrouver une propriété essentielle de la variable, que l'algorithme d'estimation aurait laissé échapper. Au contraire, on apporte une information supplémentaire, étrangère au modèle mathématique, ET ON DOIT ETRE PRET A EN ACCEPTER LES CONSEQUENCES.

Dans le cas du krigeage, ces conséquences sont au moins de deux ordres:

- 1° les variances de krigeage et la carte d'iso-variances deviennent ipso facto dépourvues de signification.
- 2° le processus de lissage tend à donner à la variable cartographiée le sens d'une "régionale": car l'opération de lissage s'apparente à une transformation par moyenne glissante. Il y a à la fin risque de confusion entre la variable et sa dérive.

## 8 - LE POINT DE VUE DES SPLINES

Naturellement, la recherche de la précision (je devrais dire: d'une précision) n'est pas un dogme. On peut au contraire privilégier des contraintes de régularité sur la forme de l'interpolateur. Tel est l'optique de la théorie des splines. Cette idée d'opposition entre les points de vue du krigeage d'une part, des splines d'autre part, est exprimée dans (Dubrule, 1986, p. 330):

*"Finally, in the situation of kriging interpolation, the objective is not to obtain an aesthetic but an accurate map; the shape of (the interpolation function) is a consequence of the minimization criterion. It is the opposite with spline interpolation, where the shape of the interpolation function determines the minimization criterion".*

Et de fait, les approches sont diamétralement opposées. En krigeage, on cherche à tenir compte au maximum de la connaissance des données; en théorie des splines, on cherche à garantir une propriété de l'interpolateur. Dans le premier cas, on se met "au service" de l'information initiale; l'initiative du géostatisticien est limitée à l'analyse structurale d'une part, au choix du plan de krigeage d'autre part. Dans le second cas au contraire, l'utilisateur décide librement des propriétés futures de son outil.

Et pourtant, ce qui est extrêmement instructif, ces deux démarches si étrangères l'une à l'autre au plan pratique sont en fait rigoureusement équivalentes au niveau théorique (Matheron, 1981b, pp. 77-95; Dubrule, 1981). La mise en oeuvre d'une interpolation par spline revient à faire un krigeage (avec en général un modèle structural non classique, que l'on pourrait caractériser au besoin): mais ce modèle sous-jacent n'est pas précisé, et surtout on ne se préoccupe pas de sa compatibilité avec les données expérimentales. Inversement, un krigeage équivaut à la minimisation d'un opérateur, généralement très complexe, dont on ne connaît pas a priori la signification graphique. Les cas où l'équivalence spline-krigeage peut être explicitée, dépendent de plus de la dimension de l'espace dans lequel on travaille; on a évoqué à la Section G-5 le cas du modèle "plaque mince" à 2 dimensions; voir aussi (Dubrule, 1986, p. 330).

Cette équivalence théorique est importante au point de vue méthodologique. Elle donne un sens plus précis aux opérations consistant à "infléchir" une analyse structurale en vue d'obtenir de "bonnes" propriétés des cartes (cf. les discussions des Sections E-8 à E-10). A la limite, on pourrait décider a priori du modèle à utiliser, afin de s'assurer dès le début de l'étude qu'il n'y aura pas d'effets indésirables (pondérateurs négatifs, irrégularités). Mais une telle tactique est évidemment dangereuse: car alors, on ne dispose plus de moyen de juger quantitativement de la précision de l'interpolateur. A la limite, on risque de se retrouver dans le cas déjà envisagé à propos de la covariance gaussienne (cf. Section E-6): le modèle, ajusté en fonction de ce qu'on souhaite et non pas de ce qu'on a observé, conduit à des variances d'estimation dépourvues de signification. On court donc ce risque redoutable d'obtenir finalement des résultats, qui tendraient à persuader l'utilisateur que ses manipulations (régularisation, lissage, etc...) étaient de toute façon bénignes, et que la carte lissée est aussi (voire plus) précise que si l'on avait scrupuleusement respecté les données initiales.

Bien sûr, le simple bon sens devrait garantir contre de telles erreurs d'interprétation - et le géostatisticien est heureusement prévenu contre ce genre de pièges méthodologiques (déjà rencontrés par exemple en pratique du KU, cf. Chauvet, 1986, p. 7). Mais le danger est trop important pour ne pas le rappeler à chaque occasion - comme ici.

## 9 - UTILISATION DES CARTES D'ISOVARIANCES

La confrontation d'une carte et de la carte des variances de krigeage associées, est une opération malheureusement trop souvent négligée. Bien sûr, dans le cas de données à maille régulière, la carte des iso-variances apporte peu d'information; et dans le cas de données irrégulières, on peut prévoir facilement avec un peu d'expérience la structure générale des courbes iso-variances. Au moins cette carte a-t-elle le premier mérite de fixer l'ordre de grandeur des variances de krigeage.

Une deuxième étape plus enrichissante consiste en la superposition des deux cartes: les courbes de niveau, et leur variance de krigeage. C'est par cette confrontation que l'on pourra par exemple donner une interprétation à des comportements erratiques de l'interpolateur. Ainsi, des irrégularités de la carte en zone de forte variance traduiront une "indécision" de l'interpolateur, probablement associée à des pondérateurs négatifs suivant le mécanisme étudié aux Sections E-3 à E-5. Au contraire, si la zone d'irrégularités est associée à de faibles variances, on pourra conclure que ces irrégularités sont une propriété structurale de la variable. On pourra raisonner de la même façon en ce qui concerne les plages très régulières. En particulier, si celles-ci sont associées à de très fortes variances - c'est-à-dire si on est éloigné de toute donnée -, on saura que l'interpolateur du krigeage n'a plus guère que le sens d'une dérive, et que sa structure n'a plus rien à voir avec la structure réelle de la variable.

Ces remarques viennent en complément de la "dualité régularité-précision" de la Section G-7: encore une fois, ce n'est pas la régularité en soi qui permet de discuter de la précision. D'autre part, on met clairement en évidence qu'une carte n'a pas une qualité homogène: une courbe de niveau est plus révélatrice des propriétés de la variable dans les zones de faible variance, et représente plutôt des caractéristiques du mode d'interpolation (artefacts) dans les zones de forte variance.

Dans un autre ordre d'idées, on pourrait utiliser l'information apportée par la carte des variances, pour aider au besoin au lissage de courbes de niveau. Ainsi, il semblerait raisonnable de ne pas modifier la carte - quelque irrégulière qu'elle puisse être - dans les parties bien reconnues, et au contraire de s'autoriser des régularisations importantes dans les zones de forte variance. Cela dit, l'automatisation d'une telle technique ne se ferait sans doute pas sans d'importantes difficultés, liées entre autres à des problèmes de cohérence des courbes de niveau. Et bien sûr, il ne serait plus possible d'associer une variance d'estimation au produit final.

Comme la variance d'estimation ne dépend que de la géométrie de l'information et du modèle structural, remarquons qu'un utilisateur averti peut critiquer aisément sa carte en la superposant au plan d'implantation des données qu'il a utilisées. A défaut de valeurs numériques, il dispose ainsi d'un outil comparatif, permettant de mettre en parallèle les propriétés des courbes de niveau dans les "bonnes" et les "mauvaises" zones de la carte: à l'égard d'éventuelles erreurs de modélisation, cet examen de simple bon sens est même sensiblement plus robuste que l'étude de la carte des variances.

## 10 - ADEQUATION DU MODELE STRUCTURAL

La remarque ci-dessus sur l'importance du "simple bon sens" dans la critique d'un résultat, n'est pas une clause de style. Car le formalisme qui permet de juger de la qualité d'une carte, sollicite le modèle encore plus que le tracé de la carte lui-même.

C'est d'un douteux privilège que bénéficient les méthodes plus frustes d'interpolation: IDW ("inverse distance weighting"), polygones d'influence, etc., ou même des théories plus élaborées comme les splines. Car, n'ayant pas de "compte à rendre" aux données, ces algorithmes ne peuvent être jugés numériquement. A l'opposé, le krigeage autorise le calcul des variances d'estimation; mais si le modèle utilisé n'est pas réaliste, ces résultats n'auront aucune signification pratique. Reprenons rapidement l'exemple déjà envisagé aux Sections E-3 à E-6: le cas de l'extrapolation en KS, à 1 dimension, avec une covariance gaussienne (on suppose pour simplifier qu'il n'y a que deux points de données):

1° On sait que les poids de krigeage sont de signes opposés, et de valeur absolue pouvant devenir très importante si la portée est grande par rapport à la configuration de krigeage.

2° Lorsque le point à estimer s'éloigne de la paire de données,

- le poids de la donnée la plus proche croît à partir de la valeur 1, passe par un maximum d'autant plus grand que la portée est grande, puis décroît asymptotiquement vers 0,
- le poids de la donnée la plus éloignée décroît à partir de 0, passe par un minimum puis croît asymptotiquement vers 0,
- et la somme de ces poids décroît de 1 vers 0 (décroissance monotone).

3° Dans le même temps, la variance de krigeage croît de façon monotone, partant de 0 et tendant asymptotiquement vers la valeur du palier de la covariance gaussienne.

Cette situation peut paraître choquante. En particulier, il semble irréaliste que la variance d'extrapolation soit bornée par la variance a priori de la variable, d'autant plus que, par le jeu des poids de krigeage alternés, les valeurs estimées risquent fort de prendre des valeurs considérablement extérieures à l'intervalle des deux données. Ce paradoxe tient à ce que le modèle envisagé est trop exigeant; et bien que les équations soient correctes, il serait totalement illusoire de réaliser un tel krigeage avec, simultanément, 1) un modèle aussi régulier, 2) aussi peu de données, et 3) l'hypothèse de stationnarité. Sans peut-être s'en rendre compte, le géostatisticien a apporté beaucoup trop d'information dans son modèle, et cette information dépasse de beaucoup celle contenue dans ses données. Appliqué dans de telles conditions, un tel modèle structural est finalement aussi arbitraire que les algorithmes "non critiques" dénoncés ci-dessus. Malheureusement, il permet - lui - de calculer une variance d'estimation - qui sera ici fortement sous-estimée.

Remarque: On pourrait me faire grief d'avoir orienté ces commentaires sur l'exemple d'une extrapolation. Ce choix n'a bien sûr rien de restrictif et, mutatis mutandis, les conclusions pourraient s'en appliquer tout aussi bien à l'interpolation au sens strict dès que la configuration de krigeage est grande par rapport à la portée.

Une incidente: ce serait d'ailleurs un exercice instructif que d'examiner, dans le cadre de la distribution multigaussienne, à quelle condition les valeurs des données peuvent être considérées comme probables (à un seuil de tolérance fixé) dans le cadre du modèle. Intuitivement, plus la portée de la covariance est grande, plus les deux valeurs doivent être proches; ou encore, pour deux valeurs données, il existe une portée maximum au delà de laquelle ce jeu de données devient définitivement improbable dans le modèle.

Cet exemple apporte un autre enseignement. Comme on l'avait déjà noté à la Section E-3, il est évident qu'un jeu de données, nécessairement fini, ne permet pas de "diagnostiquer" une propriété aussi exigeante que l'indéfinie dérivabilité. Autrement dit, l'indéfinie dérivabilité n'est pas une propriété objective (au sens défini dans: Matheron, 1978, p. 77) - c'est une propriété du modèle, non de la réalité. Cette propriété, qui infléchira naturellement les estimations ultérieures, est donc à la fois féconde et hasardeuse:

- féconde donc, puisqu'elle contient plus d'information qu'il n'y en a objectivement dans les données;
- et hasardeuse par le fait même, puisqu'il y a toujours le risque de constater après coup que le modèle est finalement totalement inadapté aux données.

Nous soulevons ici le problème de la "dialectique de la Variable Régionalisée et du modèle" (Matheron, 1978, p. 91):

"Le modèle, une fois choisi, va tendre à vivre sa vie propre, qui nous échappe en partie, et ne coïncide plus qu'en partie avec notre intuition physique de la réalité et les caractéristiques des Variables Régionalisées que nous avons choisies. Le modèle, en effet, tend à déborder la réalité, en ce sens que, si nous nous abandonnons sans esprit critique au formalisme mathématique, si nous cédon à la tentation du passage à la limite, nous nous laisserons entraîner au-delà du domaine où le modèle est opératoire."

Et remarquons de plus qu'il n'est pas possible d'esquiver cette dialectique, sous peine de déboucher sur des manipulations au mieux purement tautologiques, au pire dépourvues de toute signification objective. Pour revenir par exemple au point de départ de ces réflexions, c'est-à-dire au problème posé par BARNES et JOHNSON, il est évident d'une part qu'il faut impérativement disposer d'un modèle en continu (ne serait-ce que pour faire du krigeage étendu), et que d'autre part les échantillons disponibles ne permettront jamais d'accéder à une telle connaissance continue du phénomène.

## 11 - CONCLUSION ; RETOUR AU PROBLEME DES PONDERATEURS NEGATIFS

Ce qui rend la Géostatistique moins attrayante que d'autres méthodes, c'est qu'elle ne prétend pas que la réalité "doit" être simple, et qu'elle laisse l'utilisateur face à ses responsabilités. Phase essentielle de toute étude, l'Analyse Structurale comporte ces deux aspects complémentaires:

- elle limite la liberté du géostatisticien, en lui révélant les contraintes imposées par les données.

- mais, dans le même temps, elle ne prétend pas résoudre complètement le problème de la caractérisation du modèle: le géostatisticien doit prendre des initiatives, au risque de se tromper du tout au tout.

Le géostatisticien doit être à la fois humble (les données sont à prendre comme elles sont) et constructif (ces mêmes données sont insuffisantes pour caractériser un modèle). On se trouve à l'opposé du point de vue - fort dogmatique en vérité - qui consisterait

- 1° à spécifier le modèle a priori sans tenir compte des données réelles,
- 2° mais dans le même temps à présenter ce modèle arbitraire comme une "Vérité" absolue.

Confusion des propriétés de la variable et du modèle, confusion aussi de ce que l'on observe (ou néglige d'observer...) et de ce que l'on souhaite: deux écueils qu'une bonne étude géostatistique doit prioritairement s'efforcer d'éviter. Et c'est en réalité une des forces de la Géostatistique que de donner à l'utilisateur les moyens de surmonter ces difficultés en connaissance de cause.

En particulier, une fois l'étude effectuée, il est indispensable d'entreprendre une critique a posteriori des résultats. Il est normal, lorsqu'un problème est assez complexe, d'avoir à revenir sur certaines hypothèses et modifier certains paramètres. Aussi, le point de vue "boîte noire" doit être proscrit, sauf si l'utilisateur accepte en connaissance de cause de courir les importants risques qui lui sont inhérents.

Revenant maintenant au problème de BARNES et JOHNSON, on adoptera à la lumière de cette longue digression un double point de vue:

- un point de vue de physicien qui garde toujours à l'esprit les conditions d'une expérience, et sait en "apprécier correctement dans chaque cas particulier les seuils de robustesse et de réalisme" (Matheron, 1978, p. 93),
- un point de vue de technicien qui sait adapter ou changer ses outils en fonction du problème concret qui lui est posé.

En géostatistique appliquée, les mathématiques doivent être considérées comme un outil (d'ailleurs indispensable), jamais comme un but en soi.

## H - QUELQUES REPONSES AU PROBLEME DES PONDERATEURS NEGATIFS

### 1 - REMARQUE PREALABLE

J'examine pour terminer quelques attitudes possibles du géostatisticien confronté à des estimateurs négatifs (la variable étant supposée positive). Je ne me propose pas ici de faire une liste exhaustive des solutions qui sont développées à ce jour dans la littérature géostatistique: les différences peuvent d'ailleurs parfois se situer simplement au niveau du détail technique. Par contre, je privilégierai l'aspect "commentaires", à la lumière des remarques présentées dans l'ensemble des paragraphes précédents.

J'exprime le problème en termes d'estimateurs négatifs, parce que c'est bien sûr sous cette forme qu'il apparaît de la façon la plus immédiate au praticien. Mais on a vu que la critique de fond doit porter sur le sens même du phénomène - donc: sur les pondérateurs - plutôt que sur des manifestations spectaculaires mais superficielles (les valeurs estimées). De même, une action corrective bien comprise devra porter sur le système de krigeage lui-même (cf. Section C), non sur son résultat.

Je supposerai toujours dans les paragraphes qui suivent, que l'analyse structurale et le plan de krigeage, qui ont causé l'apparition des pondérateurs négatifs contestés, ont été réalisés dans des conditions raisonnables. Par hypothèse donc, les résultats "unsatisfactory" ne peuvent être imputés à des données fausses ou à des manipulations erronées: ils sont significatifs.

## 2 - LE STATU-QUO

Cette première attitude est bien sûr de loin la plus simple: on décide purement et simplement de passer outre, c'est-à-dire de conserver des valeurs estimées négatives.

Ce n'est pas là une proposition académique. On est dans l'hypothèse où analyse structurale et estimation ont été effectuées dans les règles de l'art. Le résultat est donc théoriquement cohérent, même si non satisfaisant au plan pratique. On se trouve en quelque sorte dans le cas d'école où tous les commentaires des paragraphes précédents s'appliquent (en particulier concernant la stabilité des résultats). La seule question qui demeure est donc: indépendamment de leur aspect "désagréable", est-ce que ces estimateurs négatifs sont rédhibitoires ?

Naturellement, il ne peut y avoir de réponse toute faite. Du reste, la seule considération esthétique peut être décisive: ainsi, une carte d'isopaques présentant des valeurs négatives fera sûrement une impression désastreuse au naturaliste. Encore faut-il distinguer si ces valeurs négatives sont seulement des "zéros mal estimés" ou traduisent une anomalie de plus grande ampleur.

De toute façon, il convient de savoir à quelle échelle on veut utiliser ces résultats du krigeage. Supposons que le domaine estimé n'excède pas l'enveloppe des données, et que celles-ci soient réparties de façon homogène; la configuration de krigeage est donc homogène sur l'ensemble du domaine estimé: on est par hypothèse dans le cas où l'estimation est globalement non biaisée. La moyenne des erreurs d'estimation est voisine de 0. Autrement dit, si localement on observe un estimateur négatif (donc: une sous-estimation), cela signifie qu'en d'autres zones du domaine, on pourra avoir des effets de surestimation - et un rééquilibrage global. Si le but du krigeage est simplement d'avoir une estimation globale assortie d'une zonéographie grossière, on pourra vraisemblablement se contenter de ce premier résultat. Par contre, il n'est évidemment pas envisageable d'utiliser ces valeurs numériques pour une estimation locale (sur des panneaux, par exemple), encore moins pour une sélection.

Il est clair que ces remarques sont essentiellement fondées sur l'homogénéité de la configuration de krigeage. Ainsi, elles ne s'appliqueront plus si on travaille en extrapolation, s'il existe un effet de bord important, ou si les données sont réparties de façon très hétérogène (clusters).

### 3 - LA MISE A ZERC

Considérer comme nulles les valeurs estimées négatives, peut éventuellement se justifier si les estimateurs ne sont que faiblement négatifs. De toute façon, il est clair que cette manipulation introduira en toute rigueur un biais sur l'estimation globale.

En fait, une telle opération a souvent le sens d'une délimitation du champ. Par exemple, en cartographie pétrolière, on considérera que l'isopaque "0" détermine la frontière d'une structure (d'un dôme, par exemple): on gardera les valeurs krigées à l'intérieur de cette frontière, et on rejettera les autres. Ce point de vue correspond très exactement à une sélection, et traduit donc une opération non linéaire.

D'après l'ensemble des observations précédentes, on peut prévoir que la structure de cette frontière sera très lissée par rapport à la limite réelle du champ (cette frontière estimée aura les caractères de dérivabilité du modèle). Aussi, une telle méthode sera très dangereuse pour estimer des volumes ou des surfaces. De plus, comme bien souvent on ne dispose pas de données nulles, la ligne de niveau "0" est en grande partie obtenue en extrapolation. On cumule ainsi toutes les difficultés: effet de bord qui rend le modèle localement sujet à caution, comportement de toute façon non stationnaire, extrapolation. Le résultat est alors inmanquablement entaché d'une erreur d'estimation considérable, et qui malgré cela est probablement encore sous-estimée parce que l'on a dépassé les limites de réalisme d'utilisation du modèle.

C'est peut-être là l'occasion de rappeler qu'une donnée nulle est une donnée comme une autre (même si non intéressante économiquement) qui contient autant d'information qu'une donnée "riche", et qu'il est toujours important de fermer le domaine étudié pour éviter les inconvénients rencontrés ici.

Enfin, surtout si elle est envisagée comme procédure automatique, cette mise à zéro systématique a le redoutable inconvénient d'escamoter le problème: la source même des difficultés demeure présente, et l'utilisateur ne fait que fuir une discussion critique qui pourrait être pleine d'enseignement. Le résultat proposé par un programme qui appliquerait systématiquement une telle troncature, ne serait donc qu'un simple trompe-l'oeil. Aussi, cette idée de mise à zéro des valeurs négatives est à déconseiller fortement, et le simple statu quo - quoique plus choquant en apparence - semble bien préférable si l'utilisateur ne souhaite pas mettre en oeuvre des procédures de correction plus élaborées.

### 4 - REAJUSTEMENT DU MODELE STRUCTURAL

Les commentaires des Sections E-7 à E-9 développent les mécanismes par lesquels une modification du modèle structural peut éliminer des pondérateurs négatifs. Au plan pratique toutefois, une différence importante existe entre les deux méthodes proposées:

- l'adjonction (ou l'augmentation) d'un effet de pépîte,
- le rajout de conditions d'universalité.

La modification de l'effet de pépité demeure une opération assez arbitraire. Car, l'analyse structurale ayant été réalisée dans les règles de l'art, l'effet de pépité initialement ajusté n'est pas plus contestable que les autres traits structuraux du modèle. Il est vrai que la caractérisation de ce comportement à l'origine fait fatalement intervenir une part de choix, puisque les comportements d'échelle inférieure à la maille des données sont en toute rigueur inaccessibles expérimentalement. Il est vrai également qu'il n'y a pas une "solution" unique au problème de l'analyse structurale, et l'expérience montre bien que deux géostatisticiens différents peuvent proposer, sur un même jeu de données, des réponses en apparence différentes (très différentes au plan de l'expression algébrique). Cet aspect "science expérimentale" de la géostatistique réserve une place importante au tour de main de l'utilisateur, et donc laisse une latitude certaine dans l'ajustement d'un modèle: ce point est illustré en particulier par l'exercice proposé dans (Matheron, 1978, pp. 93-96).

Cependant, les modifications à apporter à l'effet de pépité pour éliminer les pondérateurs négatifs, sont la plupart du temps d'un autre ordre de grandeur que cette latitude. La valeur que l'on va donner à l'effet de pépité sera donc en grande partie arbitraire. Je ne donne pas ici à ce mot une intention péjorative; mais il est clair que, par l'adjonction d'un effet de pépité, on rajoute une information qui n'est pas contenue dans les données de départ. Cette fois encore, le géostatisticien doit prendre ses responsabilités. Il peut y être aidé, en donnant à ce surplus d'effet de pépité le sens d'une erreur de mesure (souvent arbitraire). La seule difficulté méthodologique est que le modèle final cherche à la fois à tenir compte de la structure observée des données, et à anticiper les propriétés futures de l'estimateur: c'est une solution de compromis.

Le rajout de conditions d'universalité est une opération beaucoup moins radicale. Cette fois, l'utilisateur ne modifie pas fondamentalement l'image qu'il se fait du phénomène, et en tous cas ses manipulations sont moins arbitraires que précédemment. En fait, deux attitudes sont possibles:

- a) On peut se contenter de rajouter des conditions d'universalité, sans modifier le modèle structural. Cela revient simplement à restreindre la classe des estimateurs que l'on s'autorise. Dans la modélisation globale que l'on fait de la Variable Régionalisée, on décide d'accorder une importance plus grande à la partie "régulière" du phénomène (donc aux fonctions de base): on espère que les incidents dus au choix du modèle structural seront ainsi minimisés. Naturellement, si on rajoute trop de conditions d'universalité, on ira à l'encontre du but poursuivi, puisque ce sera cette fois la composante "régulière" qui risquera d'être à l'origine de comportements erratiques, et donc de provoquer des estimateurs négatifs.
- b) La première attitude était en fait une demi-mesure. Si l'on décide de rajouter des conditions d'universalité - donc de modifier l'idée que l'on se fait de la composante régulière du phénomène - il est raisonnable de changer également de modèle structural: car le phénomène global demeure quant à lui inchangé. Il est donc logique de recommencer l'analyse structurale. On passe d'un modèle en FAIk à un modèle en FAI(k+1), ce qui en théorie est parfaitement licite. Cette fois, on espère que le modèle structural sera purgé de ses caractéristiques trop régulières, qui étaient vraisemblablement à l'origine des pondérateurs négatifs. L'intérêt de ce point de vue est que l'on peut pratiquer cette nouvelle analyse structurale dans les règles de l'art: on garde un contact étroit avec

les données, la critique du modèle est toujours possible, et la part d'arbitraire notablement réduite.

Mais quel que soit le choix qui a été fait, le résultat ne peut être garanti. L'addition de conditions d'universalité est une opération beaucoup moins brutale que le rajout d'un effet de pépète, et par suite ses effets risquent d'être bien plus limités. Mais précisément, l'intérêt en est que l'on a l'impression de faire beaucoup moins violence aux données. De plus, si l'on adopte la seconde attitude, le modèle - bien qu'"influencé" par l'utilisateur - est raisonnablement adapté aux données, de sorte que les cartes d'iso-variances restent significatives.

## 5 - MODIFICATION DU PLAN DE KRIGEAGE

On a vu (Section C) que le second facteur pouvant générer des pondérateurs négatifs, est la géométrie de l'information. Naturellement, le jeu de données est supposé fixé une fois pour toutes. Mais, sauf dans le cas assez rare de krigeage en voisinage unique, on est maître du choix de son voisinage de krigeage.

La détermination du voisinage de krigeage (en estimation minière, on parle traditionnellement du choix d'un "plan de krigeage") est une manipulation typiquement géostatistique. Ainsi, (Journel et Huijbregts, 1978, pp. 344-356) lui consacrent tout un paragraphe, inventoriant les différents problèmes pratiques que l'on peut rencontrer: anisotropies, regroupement d'information, krigeage de lacunes, etc... Bien que banal, ce choix est évidemment lourd de conséquences, tant au niveau du temps calcul (mais, en général, on fait toujours attention à cet aspect de la question) qu'au niveau des propriétés de l'estimateur. Ce second point, qui risque d'être négligé par un utilisateur inexpérimenté, justifie sans doute quelques rappels de géostatistique pratique:

### a - Nombre de données

Le premier paramètre qui vient à l'esprit pour la constitution du système de krigeage, est évidemment le nombre de points de données qui va être utilisé. Ce nombre décide de la taille du système, et donc détermine directement le temps calcul nécessaire à l'estimation. On peut imaginer les choix extrêmes suivants:

- Un seul point est pris dans chaque voisinage. Cette option n'est possible que dans une optique KO ou KS. Si, comme cela est raisonnable, la donnée choisie est celle qui est la plus proche du point à estimer, le krigeage revient à remplir les polygones d'influence par une valeur estimée constante (et égale à la valeur de la donnée dans le cas du Krigeage Ordinaire).

(Remarquons que rien n'interdirait théoriquement de choisir une donnée autre que la plus proche du point à estimer. Bien sûr, un choix aussi étrange relèverait de la seule responsabilité de l'utilisateur, et les résultats observés ne sauraient servir de base à une critique raisonnable du krigeage...)

- On prend tous les points de mesure disponibles: c'est l'optique du voisinage unique. Son inconvénient majeur est de nature informatique: son coût en temps calcul d'abord, et le

risque de propagation d'erreurs d'arrondis au niveau de la résolution du système de krigeage. Mais le calcul en voisinage unique présente des avantages certains au point de vue de l'interpolation, et pour le tracé direct des courbes de niveau (Matheron, 1981a; Galli, Murillo et Thomann, 1984).

Le nombre minimum de données théoriquement requis pour assurer la régularité du système de krigeage, ne dépend que du degré de la dérive (ou, pour être plus correct, de l'ordre de la FAIk). En général bien sûr, on ne se satisfera pas de cette valeur minimum et, dans les limites d'un coût informatique donné, on cherchera à élargir le nombre de données effectivement utilisées dans chaque voisinage de krigeage, ceci pour plusieurs raisons:

- raisons de calcul: on cherche à stabiliser numériquement la solution du système. On espère, par l'effet de lissage évoqué à la Section F-3-a, obtenir un estimateur régulier - au sens où cela a été défini précédemment. Naturellement, comme on l'a déjà vu, cette régularisation pourra être mise en échec dans le cas où apparaîtraient des pondérateurs négatifs: il est bien clair que la quantité d'information n'est pas toujours déterminante pour garantir de "bonnes" propriétés à l'estimateur (voir aussi Section H-5-b).
- raisons de tracé: on veut éviter des discontinuités trop importantes sur les cartes lors d'un travail en voisinages glissants. En effet, si deux voisinages contigus n'ont que peu (ou pas) de données en commun, on observe évidemment des discontinuités sur les valeurs estimées: le krigeage est essentiellement un estimateur local, et il n'y a pas de contrainte de raccord entre voisinages. Aussi, lorsqu'un krigeage est réalisé dans de mauvaises conditions, il peut arriver que la carte finale soit bien plus représentative du pavage du domaine en voisinages de krigeage, que de la structure réelle de la variable. Une solution pratique à ce problème consiste à utiliser des voisinages glissants qui se recouvrent largement, et tels que pour deux voisinages consécutifs, les données n'appartenant pas à leur intersection soient affectées d'un poids de krigeage faible.
- raisons théoriques: dans le modèle, l'ensemble des combinaisons linéaires autorisées (donc: des estimateurs linéaires non biaisés) sera d'autant plus riche qu'il y aura davantage de données. On espère donc, en augmentant le nombre de données, diminuer la variance de krigeage.

## b - Qualité de l'information

Il apparaît clairement, dans les remarques précédentes, que le nombre de données n'est pas un critère suffisant pour juger de la qualité d'un plan de krigeage. La géométrie de l'information joue également un rôle très important.

Comme exemple simpliste, on peut évoquer les doublets de données. Au plan théorique, si les fonctions de base sont correctement choisies, le système de krigeage est régulier dès qu'il n'y a pas de données confondues - et donc la résolution des équations de krigeage ne devrait pas poser de problèmes. Dans la pratique par contre, si des données sont très voisines l'une de l'autre, le système devient très instable et à la limite, l'estimation n'a plus aucune signification. (Rappelons que cette instabilité peut souvent se manifester par la présence de pondérateurs négatifs). Il ne suffit donc pas d'avoir beaucoup de données, encore faut-il qu'elles soient de bonne qualité. On retrouve la dualité de la Section F-3-c: le krigeage, fondé sur un

modèle global pour réaliser une estimation locale, exige d'abord une bonne homogénéité des données.

Plus précisément, il faut que les données mises à contribution à l'étape d'estimation, soient utilisées dans des conditions qui soient en bon accord avec les conditions de l'analyse structurale. La première idée est ainsi de respecter l'échelle de validité du modèle structural: ainsi, un voisinage de krigeage ne doit pas excéder les dimensions pour lesquelles le variogramme a été modélisé. De même, il est illusoire d'utiliser individuellement des données trop voisines les unes des autres ("clusters") si cet effet de cluster ne s'est traduit au niveau du variogramme que par un effet de pépite - c'est-à-dire si la connaissance du phénomène à très petite échelle n'en a pas été améliorée: on est donc amené à effectuer des regroupements d'information. Et naturellement, la géométrie du voisinage de krigeage doit tenir compte des éventuelles anisotropies: anisotropie de la variable étudiée bien sûr, mais aussi anisotropie de l'information (maille rectangulaire, données le long de sondages, etc...).

Si un plan de krigeage est mal adapté, on dépasse les limites de réalisme du modèle. Les données sont trop sollicitées par l'algorithme d'estimation, et le résultat final peut être dépourvu de sens: c'est par exemple ce qui se passe, au cours d'une étude par ailleurs correcte, lorsqu'on atteint les limites du domaine (effets de bord). Ce risque est d'autant plus grand que le modèle structural est plus "riche", ou, si l'on préfère, plus exigeant avec les données: cette remarque doit également être prise en compte dans le choix du plan de krigeage.

### c - Choix d'un plan de krigeage

A la lumière des observations précédentes, on peut énoncer une première règle de caractère général:

*"Lors de l'élaboration du plan de krigeage, on doit d'abord chercher à privilégier*

- la QUANTITE D'INFORMATION si la variable à estimer est FAIBLEMENT STRUCTUREE
- la QUALITE DES DONNEES (concrètement: leur implantation) si la variable à estimer est FORTEMENT STRUCTUREE"

Proposons deux exemples extrêmes:

- 1' Le modèle pépitique pur correspond à une variable totalement non structurée (bruit blanc). Lors du krigeage, toutes les données sont affectées du même poids (positif, bien sûr), et la variance de krigeage dépend seulement du nombre de données, et non de leur implantation. On a vu à la Section G-4 les conséquences de ce modèle sur l'allure des courbes de niveau.
- 2' Le modèle à covariance gaussienne présente au contraire un cas de phénomène très structuré. La disposition des données se révèle alors déterminante pour le krigeage, du moins si les distances au sein de la configuration de krigeage sont inférieures à la portée pratique. Face à des dispositions défavorables des données (doublets, clusters), le système de krigeage "réagira", par exemple en générant des poids négatifs. Autrement dit, lors de l'élaboration du plan de krigeage, il faudra veiller avec le plus grand soin à écarter les configurations "à risque", c'est-à-dire plus précisément:

- éliminer les doublets ou multiplats de données

- et s'assurer que les données entourent bien le domaine à estimer, pour éviter de travailler en extrapolation.

Entre ces deux cas extrêmes, complétons l'exemple déjà cité à la Section E-1 (Matheron, 1971, pp. 130-131): l'estimation d'un segment par KO, avec un modèle de variogramme monomial. Lorsque le variogramme est de degré inférieur à 1, il est plus avantageux (i.e: la variance de krigeage est plus faible) d'utiliser comme données les deux extrémités du segment, plutôt que le seul milieu du segment: il faut préférer le nombre à une bonne implantation. Cette priorité s'inverse lorsque le degré du variogramme est supérieur à 1 (plus grande régularité).

Notons au passage le rôle particulier des modèles markoviens (ici, le variogramme linéaire): ils constituent en quelque sorte une frontière, entre les modèles pour lesquels l'apparition de pondérateurs négatifs est plutôt de nature "accidentelle", et ceux au contraire pour lesquels cette apparition sera plutôt la règle générale. Mais il ne s'agit là que d'un résultat empirique, non d'un théorème rigoureux. On démontre en effet (Matheron, 1986) que ce n'est qu'à 1 dimension qu'il est possible de proposer des modèles structuraux qui ne généreront jamais que des pondérateurs positifs (et seulement en KS ou KO); par contre, en FAIk ou à plusieurs dimensions, il est toujours possible de trouver des configurations de krigeage conduisant à des poids négatifs. On a d'ailleurs déjà mentionné à la Section E-9 qu'il était intuitivement prévisible que le rajout de conditions d'universalité augmente les chances de rencontrer de tels poids négatifs.

Toutes les remarques du présent paragraphe sont faites en supposant que le modèle est obtenu - globalement - dans de bonnes conditions. Mais on a vu que certains modèles sont plus exigeants avec les données que d'autres. Les deux cas extrêmes proposés ici en sont une illustration. Ainsi, le modèle à covariance gaussienne est indéfiniment dérivable: toute réalisation d'une Fonction Aléatoire Stationnaire à covariance gaussienne est donc une fonction analytique. En toute rigueur, la connaissance de cette fonction sur quelque ouvert que ce soit devrait caractériser la fonction dans tout l'espace. Au plan pratique, on mesure le caractère exorbitant d'une telle propriété, même s'il faut observer que dans la réalité, une Variable Régionalisée n'est jamais reconnue que sur un nombre fini de points... On met là en évidence l'origine profonde des phénomènes "pathologiques" observés lorsqu'on travaille avec une covariance gaussienne.

Il ne faut d'ailleurs pas croire que ce modèle soit le seul à impliquer des propriétés irréalistes. Ainsi, le modèle pépitique pur se révèle également très contraignant: la variance d'extension de  $v$  à  $v'$  (où  $v$  et  $v'$  sont deux ouverts quelconques) est nulle. On peut bien sûr observer une nouvelle fois que les données sont toujours en nombre fini et que le domaine à estimer, même s'il est censé être continu (pour une estimation de bloc par exemple), sera toujours discrétisé pour des raisons informatiques. Cela dit, dans le modèle, la variance de krigeage peut toujours être rendue inférieure à une valeur fixée arbitrairement: il suffit pour cela d'augmenter le nombre de données, prélevées n'importe où dans le champ. Cette fois encore, on met en évidence une propriété irréaliste du modèle. Une des tâches prioritaires de la Géostatistique est précisément de mettre l'utilisateur trop pressé en garde contre une mise en oeuvre non critique de modèles par ailleurs théoriquement corrects: nous avons déjà été confrontés à ce problème au sujet de l'interpolation (Section G); la détermination du plan de krigeage est sans doute l'étape où cette phase de critique est la plus décisive. Pour compléter cette discussion sur le modèle pépitique pur, on pourra se reporter à l'étude du krigeage des grandes mailles

(Matheron, 1971, pp. 105-110, pp. 137-138 et p. 200), déjà citée à la Section E-4. Mais, pour clore par une affirmation péremptoire, on peut dire que le modèle péritique pur - comme précédemment le modèle à covariance gaussienne - est un modèle "à risque"; en termes anthropocentriques, il "attend trop des données".

#### d - Modification du plan de krigeage

Lorsque l'on a choisi de modifier le plan de krigeage pour éliminer des phénomènes indésirables, il convient bien entendu de s'inspirer des règles générales évoquées ci-dessus. Mais une difficulté supplémentaire doit être prise en compte.

En effet, sauf dans le cas où les anomalies sont assez nombreuses pour justifier une reprise complète de l'estimation, on aura en général tendance à proposer des modifications locales. Le plan de krigeage ne sera remanié que dans les zones où des difficultés ont surgi. Aussi, ce faisant, on court le risque de proposer une modification ad hoc, et donc de s'éloigner notablement d'une utilisation correcte du modèle. En poussant jusqu'à l'absurde, on pourrait choisir tout simplement les données "qui arrangent", qui assurent un estimateur positif ou mieux encore, qui garantissent la positivité de tous les pondérateurs: par exemple, la donnée la plus proche... Ce faisant, on réduit le krigeage aux estimateurs les plus frustes que l'on puisse imaginer - à ceci près que ces manipulations sont beaucoup plus coûteuses... De toute façon, l'honnêteté intellectuelle de l'utilisateur est dangereusement sollicitée.

Il est vrai que la plupart du temps, la présence de pondérateurs négatifs n'améliore guère la précision du krigeage. Aussi, semble-t-il, on pourrait sans grand inconvénient écarter les données qui ont été affectées de tels poids. Encore faut-il noter que cette affirmation est établie à partir d'un modèle (la précision est donnée par la variance théorique de krigeage); or, précisément, la situation anormale que l'on cherche à corriger est due en général à une inadéquation du modèle (global) aux conditions locales. On se trouve donc dans la situation dangereuse où, dans le même temps, on se heurte aux limites du modèle, et on demande à celui-ci de se juger lui-même.

Une autre difficulté tient à l'hétérogénéité que l'on fait surgir à l'échelle du champ. Le plan de krigeage est orienté (on pourrait dire: biaisé) en fonction des particularités locales, ou même, en fonction de ce que l'on souhaite obtenir. De plus, si c'est l'ESTIMATEUR que l'on astreint à être de signe positif - et non les pondérateurs -, cela revient à conditionner les poids de krigeage par les données: on se trouve exactement aux antipodes d'une utilisation correcte de la Géostatistique Linéaire. C'est pourquoi J. Rivoirard souligne qu'il serait souhaitable de mettre alors en oeuvre les techniques de Géostatistique Non Linéaire: mais bien sûr, pour ce faire, les hypothèses requises sont beaucoup plus fortes, et cette opération n'est pas toujours possible.

Même si, comme cela est correct, on se limite à prendre en compte la géométrie de l'information pour modifier le plan de krigeage, des risques d'erreur subsistent. Ainsi, le Krigeage Aléatoire (Journel et Huijbregts, 1978, pp. 352-355), bien que séduisant, n'est pas une réponse satisfaisante: on se trouve au contraire en général dans le cas où les données ont une configuration exceptionnelle (bordure du champ, lacune, par exemple), et n'ont donc pas une implantation "aléatoire" à l'intérieur du voisinage de krigeage. Le Krigeage Aléatoire, qui est

une technique en quelque sorte encore plus globale que le krigeage usuel, sera donc encore moins adapté aux circonstances anormales; même si les pondérateurs négatifs ont disparu, la valeur estimée sera probablement de peu de signification et, surtout, le modèle sera impuissant à diagnostiquer cette anomalie. En particulier, la variance de krigeage sera accrue par rapport à la variance du krigeage usuel, mais elle sera complètement inadaptée aux caractéristiques locales.

Une autre méthode consiste à opérer des regroupements d'information, et à affecter un pondérateur de krigeage unique à chacun de ces regroupements. L'idée est de réduire la taille du système de krigeage, et donc de pouvoir utiliser davantage de données à coût informatique fixé. Là encore, il peut s'agir d'une arme à double tranchant. D'abord, selon les observations de la Section H-5-c, cette méthode est plus adaptée aux variables faiblement structurées: plus la variable est structurée, et plus il est souhaitable de limiter ces regroupements à des informations suffisamment éloignées du domaine à estimer. D'autre part, les regroupements sont en général guidés par des considérations de symétrie de l'information (compte-tenu bien sûr des éventuelles anisotropies du modèle structural); aussi, il ne faut pas trop espérer jouer sur ces regroupements pour résoudre les problèmes de type "effet de bord". Par contre, on peut éventuellement aider ainsi à résoudre les problèmes liés à l'existence de lacunes dans l'information.

La mise en oeuvre de regroupements revient à rechercher un estimateur dans une classe plus pauvre que celles des Combinaisons Linéaires Autorisées. A ceci près, la variance de krigeage a exactement la signification usuelle, ce qui n'était pas le cas pour le Krigeage Aléatoire. Elle est bien sûr supérieure à ce qu'elle aurait été par un krigeage usuel (KS, KO, ou FAIk) réalisé avec les mêmes données prises individuellement. En contrepartie de cette perte de précision, on peut favoriser l'influence de la tendance dans le voisinage de krigeage, et ainsi à la fois limiter les risques d'estimateurs négatifs, et atténuer les discontinuités entre voisinages de krigeage.

En résumé, la méthode des regroupements peut parfois être de pratique délicate, mais elle permet une grande souplesse dans la caractérisation de l'estimateur. Elle est d'ailleurs employée assez systématiquement dans les programmathèques de krigeage utilisées en Géostatistique Minière. De toute façon, la discussion du plan de krigeage, même dans le cas élémentaire où les données sont toutes considérées individuellement, doit être une étape essentielle de toute étude géostatistique. Une approche méthodique de cette question fait l'objet de (Rivoirard, 1984). Dans la pratique, l'élaboration et à plus forte raison le réajustement du plan de krigeage sont sans doute les opérations qui sollicitent le plus le savoir-faire du géostatisticien.

## 6 - PASSAGE A LA GEOSTATISTIQUE NON-LINEAIRE

Le mécanisme qui conduit de pondérateurs négatifs à des estimateurs négatifs, peut être associé de façon flagrante à de simples problèmes de géométrie de l'information: c'est le cas par exemple en bordure de champ, lorsqu'on travaille en extrapolation.

Mais il peut aussi être révélateur d'une difficulté beaucoup plus fondamentale. Comme cela a été mentionné à la Section D, l'étude de la probabilité de coïncidence entre un pondé-

rateur négatif et une donnée forte se révèle très complexe. Cependant, l'intuition et l'expérience permettent de pallier cette lacune du formalisme théorique: on sait ainsi que les risques d'incidents sont accrus lorsque la variable étudiée a une distribution très éloignée de la normale. Par exemple, si la variable présente à la fois un fort effet zéro et une très longue queue d'histogramme, il pourra plus facilement arriver qu'une très forte valeur affectée d'un pondérateur négatif ne puisse pas être compensée par les autres données, de poids positifs, mais trop faibles. Ce cas n'est qu'une illustration de ce que les estimateurs linéaires (dont en particulier le krigeage) s'"éloignent" de l'estimateur idéal (l'espérance conditionnelle) lorsque la loi s'"éloigne" de la distribution gaussienne.

Remarque: c'est par la force des choses que j'utilise ici le verbe "éloigner" au sens intuitif. Ce ne serait pas chose facile que de donner une définition rigoureuse de la dégradation de qualité d'une classe d'estimateurs, parce que cette définition devrait être indépendante du modèle. Et ce serait peut-être encore plus difficile de définir un concept de distance entre lois de distributions, qui puisse être directement relié à cette dégradation.

La prise en compte de la distribution des données relève de la Géostatistique Non Linéaire. La technique du Krigeage Disjonctif revient à affecter aux données des poids dépendant de leur valeur. On peut ainsi espérer réduire considérablement les risques d'apparition d'estimateurs négatifs. En contrepartie, les conditions requises sur les données (stationnarité en particulier) sont beaucoup plus exigeantes que pour la Géostatistique Linéaire. On ne sait pas actuellement construire des estimateurs non linéaires en hypothèse non stationnaire.

Une des étapes essentielles de la Géostatistique Non Linéaire est l'anamorphose gaussienne. (On espère que cette transformation suffit à éliminer les difficultés liées à des distributions trop éloignées de la normale: il y a d'ailleurs là un pari, puisqu'il s'agit d'une anamorphose ponctuelle, alors que les formalismes développés par la suite suppose la binormalité). On pourrait alors essayer, à la lumière de cette technique, de recourir à une transformation de la variable pour éliminer les risques d'estimateur négatif, sans pour autant mettre en oeuvre tout l'arsenal des méthodes non linéaires. Une telle optique n'est pas rare: pour des variables que l'on sait de distribution très dissymétrique (teneurs en Uranium, par exemple), il est fréquent de passer en logarithmes pour réaliser l'estimation. Bien sûr, ce faisant, on est assuré de ne pas générer d'estimateur négatif, puisque la dernière étape est une exponentiation. Mais il ne faut pas oublier que ce mode de calcul introduit un biais, à cause précisément du caractère non linéaire de la transformation; et la correction rigoureuse de ce biais nécessite l'hypothèse de lognormalité. Si cette hypothèse n'est pas vérifiée, l'utilisation du formalisme de correction de biais n'a pas de sens (cf. par exemple: Carlier, 1964, p. 163). Aussi, sauf circonstance exceptionnellement favorable, la rigueur impose d'appliquer intégralement les algorithmes de Géostatistique Non Linéaire.

## 7 - CHOIX D'ALGORITHMES NOUVEAUX

Pour répondre aux problèmes d'estimateurs négatifs, cette option est évidemment celle qui ouvre la voie aux développements les plus nombreux et les plus divers. Il n'est donc pas question

ici d'en dresser un inventaire complet, même en se limitant aux seuls aménagements du krigeage.

On trouve dans (Dubrule, 1986, pp. 38-39) un commentaire sur quelques méthodes existant pour résoudre le krigeage sous contraintes de positivité. Ainsi, (Barnes et Johnson, 1984, pp. 231-244) utilisent une méthode du simplexe pour imposer la positivité des pondérateurs du krigeage. De même, (Limic et Mikelic, 1984, pp. 423-429) étudient l'estimation de variables positives, et définissent un cadre d'hypothèses pour lequel la contrainte de positivité de l'estimateur implique la positivité de tous les pondérateurs; ils proposent une solution fondée sur l'algorithme de Dantzig.

Plus généralement, (Cazes, 1978) ou (Mallet, 1980) envisagent la question de la régression sous contrainte. Ce point de vue peut en particulier s'appliquer au cas du krigeage. La contrainte de positivité porte, cette fois, sur l'estimateur lui-même. Antérieurement, cette idée d'introduire des relations d'inégalité dans les problèmes d'interpolation a été développée en théorie des splines (Atteia, 1967; Laurent, 1972, chapitre IX).

Les méthodes de krigeage sous contrainte se classent donc en deux familles, suivant que ces contraintes portent sur les pondérateurs ou les estimateurs. Mais de toute façon, le krigeage n'est plus une projection sur un espace vectoriel, mais seulement sur un espace convexe. Le prix minimum à payer pour ceci est la perte de la propriété de linéarité du krigeage: le krigeage d'une somme n'est plus la somme des krigeages, et la technique du krigeage des lacunes (Journal et Huijbregts, 1978, pp. 351-352) n'est plus licite.

## a - Contraintes sur les pondérateurs

Imposer une condition de positivité à l'ensemble des poids de krigeage, est évidemment une contrainte exigeante. Car cette condition est suffisante, mais nullement nécessaire dans le cas général pour garantir ce que recherche le praticien, à savoir la positivité des estimateurs.

Un des intérêts d'un tel choix, est qu'il n'implique aucune hypothèse sur la loi de la variable. En ce sens, on adopte un point de vue qui relève strictement de la Géostatistique Linéaire, même si les techniques de résolution qu'il faut mettre en oeuvre ne sont plus linéaires. On se trouve dans le cadre exact des commentaires de la Section E: les difficultés trouvent leur origine à la fois dans la structure du phénomène et dans la géométrie de l'information, et on cherche à les résoudre en restreignant la classe des estimateurs autorisés, sans remettre en cause ni cette structure, ni cette géométrie. Le point de vue des projections, évoqué ci-dessus, montre que la variance de krigeage fournie par un tel krigeage sous contrainte sera toujours supérieure (au moins égale) à ce qu'elle aurait été par un krigeage usuel.

Les seules différences entre les différents algorithmes proposés ne devraient donc être liées qu'au choix d'un programme d'optimisation, et donc ne devraient se répercuter que sur les performances des programmes.

Observation: la condition de positivité sur tous les pondérateurs est une contrainte assez radicale. On a vu (Section E-9) que dans un formalisme FAIk ou KU, l'apparition de pondérateurs négatifs peut être un phénomène assez fréquent, et surtout parfaitement naturel (au sens qu'il ne traduit pas d'"anomalie" locale) puisqu'il cherche à rendre compte des dérives locales.

Aussi, la contrainte de positivité sur les pondérateurs risque de réduire de façon excessive la classe des estimateurs autorisés, et de conduire à des résultats beaucoup trop éloignés de l'optimalité. Remarquons que la technique du Krigeage Positif (Barnes et Johnson, 1984) échappe à cette remarque, car elle se place explicitement dans une optique KO; mais il est probable que, bien que théoriquement correcte, une généralisation aux cas non stationnaires en serait peu satisfaisante.

Toujours dans (Barnes et Johnson, 1984), on relève les deux propositions suivantes, page 237 d'abord:

*"The optimal values of the positive weights obtained by Positive Kriging are equal to the weights generated by the Ordinary Kriging procedure using only those samples associated with the optimal basic weights."*

page 238 ensuite:

*"The set of negative weights generated by an initial execution of the OK procedure is not necessarily equal to the optimal set of non-basic weights".*

Rappelons que, dans la terminologie du Krigeage Positif, les "basic weights" sont ceux qui sont positifs à l'optimalité; et les "non-basic weights" sont ceux qui sont mis à zéro par l'algorithme d'optimisation. La seconde proposition en particulier est importante, car elle explique pourquoi il n'est pas possible d'accéder à l'optimum du KO sous contrainte par un processus itératif: on aurait pu imaginer la démarche suivante:

- 1° on réalise un KO sans contrainte, avec le jeu de données initiales;
- 2° si tous les pondérateurs du KO sont positifs, la première proposition permet d'affirmer que l'optimum du Krigeage Positif est ainsi atteint;
- 3° si certains pondérateurs sont négatifs, on écarte les données correspondantes, et on recommence un KO sur les données restantes;
- 4° si tous les pondérateurs sont positifs, on s'arrête; sinon, on reprend au 3°.

Mais la seconde proposition de BARNES et JOHNSON prouve que cette procédure est erronée: rien ne garantit que le résultat ainsi obtenu soit effectivement l'optimum. En réalité, cette méthode itérative n'est autre qu'une modification ad hoc du plan de krigeage. Mais pour obtenir réellement l'optimum théorique sous contrainte, il faut effectivement avoir recours à des techniques de programmation quadratique.

Une difficulté apparaît si le Krigeage Positif est utilisé en cartographie. Car, à l'intérieur d'un même voisinage de krigeage, rien ne garantit que la classe de "basic weights" reste la même pour tous les points à estimer. Aussi, d'après la première proposition ci-dessus, tout se passe comme si les points du voisinage de krigeage n'étaient pas tous estimés à l'aide des mêmes données. Il est donc à craindre des discontinuités qui, contrairement à celles causées par les voisinages glissants, sont presque imprévisibles et difficiles à contrôler.

## b - Contraintes sur les valeurs estimées

Les méthodes qui définissent des contraintes sur les futures valeurs de l'estimateur sont plus variées, et on n'en proposera pas ici un inventaire. On peut toutefois distinguer celles qui requièrent des hypothèses sur la loi de la variable étudiée, et celles qui s'affranchissent de telles hypothèses.

Les premières s'apparentent en fait à la Géostatistique Non Linéaire. Naturellement, elles sont de ce fait plus exigeantes avec les données, et ne peuvent - en l'état actuel de la théorie - être mises en oeuvre dans un cadre non stationnaire. Les secondes au contraire s'adaptent bien au travail en FAIk, en particulier dans une optique cartographique.

En matière de tracé de courbes de niveau, une importante difficulté est soulignée dans (Dubrule et Kostov, 1986, p. 39): le krigeage est un outil d'estimation locale (on pourrait dire qu'il travaille point par point), et il fera également un usage purement local des relations d'inégalité. Comme on a vu à la Section G-4 en particulier, la régularité (au sens intuitif) des courbes de niveau n'est pas directement maîtrisée par l'utilisateur, contrairement à ce qui se passe quand on travaille avec des splines. Aussi, la présence de contraintes purement ponctuelles peut se traduire au niveau du tracé par des comportements totalement inacceptables: DUBRULE et KOSTOV en proposent un exemple spectaculaire (p. 40).

Pour lever cette difficulté, DUBRULE et KOSTOV partent du formalisme du Krigeage Dual, et prouvent qu'on peut lui appliquer l'approche des splines sous contrainte (Dubrule et Kostov, 1986, p. 43). Notons que cette fois, le point de vue adopté est celui de l'interpolation, et que tout contexte probabiliste est oublié. Il n'y a donc pas lieu de s'interroger sur les lois des variables: on se trouve ici à l'opposé des méthodes de Géostatistique Non Linéaire, et le formalisme proposé est parfaitement adapté au krigeage non stationnaire. Le prix qu'il faut payer pour ces améliorations, est la perte de la variance d'estimation.

### c - Brève conclusion

Ce rapide survol de quelques méthodes actuellement existantes montre qu'il ne peut exister de réponse dogmatique au "problème" des poids ou des estimateurs négatifs. Outil simple - on pourrait dire: passe-partout - requérant peu d'hypothèses, le krigeage peut naturellement apparaître peu satisfaisant dans des situations particulières. Mais l'arsenal d'outils mathématiques disponibles est suffisant pour proposer des perfectionnements, pour la satisfaction aussi bien de ceux qui souhaitent légitimement ces pondérateurs négatifs, que de ceux qui les redoutent tout aussi légitimement. Le vrai problème n'est pas là.

Le vrai problème d'une Géostatistique bien comprise est d'abord de satisfaire aux exigences de réalisme. Rechercher un outil qui fournirait une "bonne" solution à tous les problèmes simultanément est - au mieux - totalement utopique, surtout si les exigences a priori sont contradictoires... Affirmer avoir construit un tel outil est plus simplement malhonnête. La justesse d'un algorithme est évidemment indispensable. Mais l'exécution du programme n'est pas la phase essentielle d'une étude. La critique et le choix des données en amont, la critique des résultats en aval, puis éventuellement la décision d'une nouvelle exécution ou le choix d'un nouvel algorithme, telles sont les étapes cruciales d'un travail, où le Géostatisticien doit se comporter bien plus en Physicien qu'en Mathématicien. On espère que cette note aide à adopter cette attitude pour examiner le cas des pondérateurs et des estimateurs négatifs.

## BIBLIOGRAPHIE

- Barnes R.J. et Johnson T.B., 1984.- Positive kriging, Geostatistics for Natural Resources Characterization - NATO ASI series, C122, Reidel ed., Dordrecht
- Carlier A., 1964.- Contribution aux méthodes d'estimation des gisements d'uranium - Thèse présentée à la Faculté des Sciences de l'Université de Paris
- Cazes P., 1978.- Méthodes de régression, la régression sous contrainte - Les cahiers de l'analyse des données, vol. 2, Dunod
- Chauvet P., 1986.- La géostatistique non stationnaire: KU ou FALK ? - Ecole Nationale Supérieure des Mines, Paris
- Chauvet P., 1987.- The comparison between the gamma logs and the grades in the estimation of a uranium deposit - Geostatistical Case Studies, Reidel ed., Dordrecht
- CHILES J.P., 1977.- Modeles de schemas de transition adaptes aux simulations dans le plan - Ecole Nationale Supérieure des Mines, Paris
- Chiles J.P., 1979.- La dérive à la dérive - Ecole Nationale Supérieure des Mines, Paris
- Delfiner P., Delhomme J.P. et Pelissier-Combescure J., 1983.- Applications of geostatistical analysis to the evaluation of petroleum reservoirs with well logs - presented at the 24th annual Logging Symposium of the SPWLA, Calgary
- Dubrulle O., 1981.- Krigeage et Splines en cartographie automatique. Application à des exemples pétroliers - Ecole Nationale Supérieure des Mines, Paris
- Dubrulle O., 1984.- Comparing Splines and Kriging - Computers & Geosciences, vol. 10, n° 2-3
- Dubrulle O. et Kostov C., 1986.- An Interpolation Method Taking into Account Inequality Constraints - Math. Geol., vol. 18, n° 1
- Galli A., Murillo E. et Thomann J., 1984.- Dual kriging: its properties and its uses in direct contouring - Reidel ed., Dordrecht
- Journel A., 1974.- Simulations conditionnelles: théorie et pratique - Ecole Nationale Supérieure des Mines, Paris
- Journel A., 1986.- Constrained Interpolation and Qualitative Information - The Soft Kriging Approach - Math. Geol., vol. 18, n° 3
- Journel A. et Huijbregts Ch., 1978.- Mining Geostatistics - Academic Press, London
- Jordan Ch., 1965.- Calculus of finite differences - Chelsea Publishing Company, New York
- Laurent P.J., 1972.- Approximation et optimisation - Hermann ed., Paris
- Limic M. et Mikelic A., 1984.- Constrained Kriging Using Quadratic Programming - Math. Geol., vol. 16, n° 4
- Mallet J.L., 1980.- Régression sous contraintes linéaires: application au codage des variables aléatoires - Revue de Statistique Appliquée, vol. 38, n°1
- Matheron G. et Formery Ph., 1963.- Recherche d'optimum dans la reconnaissance et la mise en exploitation des gisements miniers - Annales des Mines, vol. VI, Juin 1963
- Matheron G., 1969.- Le Krigeage Universel - Ecole Nationale Supérieure des Mines, Paris

- Matheron G., 1970.- Krigeage Universel pour une dérive aléatoire - Ecole Nationale Supérieure des Mines, Paris
- Matheron G., 1971.- The theory of regionalized variables and its applications - Ecole Nationale Supérieure des Mines, Paris
- Matheron G., 1973.- The intrinsic random functions and their applications - Advances in Applied Probability, vol. 5
- Matheron G., 1978.- Estimer et choisir - Ecole Nationale Supérieure des Mines, Paris
- Matheron G., 1979.- Comment traduire les catastrophes, ou la structure des Fonctions Aléatoires Intrinsèques Généralisées - Ecole Nationale Supérieure des Mines, Paris
- Matheron G., 1981a.- Remarques sur le krigeage et son dual - Ecole Nationale Supérieure des Mines, Paris
- Matheron G., 1981b.- Splines and kriging: their formal equivalence, Down-to-earth statistics: solutions looking for geological problems - D.F. Merriam ed., Syracuse University
- Matheron G., 1986.- Sur la positivité du poids de krigeage - Ecole Nationale Supérieure des Mines, Paris
- Philip G.M. et Watson D.F., 1986a.- Matheronian geostatistics - Quo vadis? - Math. Geol., vol. 18, n° 1
- Philip G.M. et Watson D.F., 1986b.- Geostatistics and spatial data analysis - Math. Geol., vol. 18, n° 5
- Rivoirard J., 1984.- Le comportement des poids de krigeage - Ecole Nationale Supérieure des Mines, Paris
- Sibson R., 1981.- A brief description of Natural Neighbour Interpolation in: Interpreting Multivariate Data - Barnett V. ed., J. Wiley & sons, New York

