ESTIMATION SOUS CONTRAINTES D'INÉGALITÉS

V. LANGLAIS *

TABLE DES MATIÈRES

RÉSUMÉ 212
ABSTRACT 212
A – INTRODUCTION 213
B – LES MÉTHODES 213
1. Minimisation de la forme quadratique Z'UZ sous contraintes 213
2. L'espérance conditionnelle 214
3. Méthode incrémentale 214
C – COMPARAISON DES MÉTHODES SUR UN EXEMPLE 216
1. Présentation des données 216
2. Analyse structurale 218
3. Minimisation de la forme quadratique . 219
4. Espérance conditionnelle 220
5. Méthode incrémentale 220
D – CONCLUSION 222

ILLUSTRATIONS

Figure 1 – Krigeage avec les seules données exactes 217
Figure 2 – Coupe sur le krigeage avec les altitudes réelles 217
Figure 3 – Coupe sur le krigeage avec les données exactes 217
Figure 4 – Différents variogrammes 218
Figure 5 – Coupe sur l'estimation par minimisation d'une forme quadratique 219
Figure 6 – Coupe sur l'estimateur espérance conditionnelle 220
Figure 7 – Coupe sur l'estimation par la méthode incrémentale 220
Figure 8 – Coupe sur une simulation 1 221
Figure 9 – Coupe sur une simulation 2 221

* Ecole Nationale Supérieure des Mines de Paris, Centre de Géostatistique – 35, rue Saint-Honoré, 77305 FONTAINEBLEAU (FRANCE)

Etudes Géostatistiques VI - Séminaire C.F.S.G. sur la Géostatistique, Juin 1989, Fontainebleau. Sci. de la Terre, Sér. Inf., Nancy, 1991, 30, pp 211 à 224

RÉSUMÉ

Pour traiter les problèmes d'estimation en présence de données exactes et de données où seul un encadrement de la variable est connu, on dispose de plusieurs méthodes. La minimisation d'une forme quadratique sous contraintes d'inégalités répond au problème aussi bien dans le cas stationnaire que dans le cas non-stationnaire. Cependant, elle a l'inconvénient de caler certaines contraintes (les contraintes actives) à leurs bornes, ce qui n'est pas souhaitable pour certains problèmes (limite de couche). Les autres méthodes évitent cet écueil. Si l'espérance conditionnelle, sous réserves d'hypothèses multigaussiennes, est le meilleur estimateur que l'on puisse construire, et si elle fournit une variance d'estimation, elle se heurte rapidement à des problèmes pratiques (précision, nombre limité de contraintes, rapidité) outre les hypothèses fortes requises. Enfin, la méthode incrémentale évite les problèmes pratiques de l'espérance conditionnelle mais pose quelques problèmes de mise en oeuvre (ordre de traitement des contraintes, anamorphose). Pour toutes ces méthodes se pose le problème de l'analyse structurale avec l'ensemble de l'information où quelques outils simples permettent un contrôle de la modélisation. Enfin, ces méthodes sont comparées sur un exemple dans le cas stationnaire.

ABSTRACT

Several methods are available for estimation problems involving exact data and inequality data (the value of the variable is known to be in an interval). Minimizing a quadratic form under inequality constraints can work in both the stationary and non-stationary cases. But in some cases, the constraints reach one of their limit values (binding constraints), which is not satisfactory for some problems (layer delimitation). The other methods do not have this disadvantage. Although the conditional expectation, under the multigaussian hypothesis, is the best estimator that can be built (it also provides the estimation variance) some practical problems arise: accuracy, limitation of the number of inequalities, computer time. The method of truncated normal scores avoids these practical problems of the conditional expectation, but leads to problems when putting it into practice. Firstly, the order in which inequality constraints are processed is important, and an anamorphosis is required. These methods have one common problem, that of the structural analysis with inequality data. Some basic tools are presented to check the modeling. Lastly, these methods are tested and compared in a stationary case.

A – INTRODUCTION

Il arrive que l'information disponible se présente sous deux formes : des données exactes (la valeur de la variable est connue au point x) et des contraintes (seul un encadrement de cette valeur est connu). On voudrait construire un estimateur qui prenne èn compte ces deux types de données, ce qui est d'autant plus indispensable lorsque l'estimateur, construit avec les seules données exactes, ne respecte pas les données du type intervalle. Pour construire un tel estimateur, on dispose de plusieurs méthodes : minimisation d'une forme quadratique sous contraintes, espérance conditionnelle (conditionnement par les données et les contraintes), et enfin une méthode incrémentale. Il est alors intéressant de comparer sur un exemple les résultats obtenus par ces trois méthodes et aussi d'aborder un de leurs problèmes communs qui est l'analyse structurale avec toute l'information (données exactes et contraintes).

B – LES MÉTHODES

1 – MINIMISATION DE LA FORME QUADRATIQUE Z^tUZ SOUS CONTRAINTES

On ne rappelle ici que briévement cette méthode proposée par Dubrule et Kostov (1986).

On notera $\begin{pmatrix} U & V \\ V^t & W \end{pmatrix}$ l'inverse de la matrice de krigeage $\begin{pmatrix} K & F \\ F^t & 0 \end{pmatrix}$.

On minimise la forme quadratique Z^tUZ sous les contraintes suivantes :

$$Z_\alpha = z_\alpha \quad \text{aux n points de données}$$

$$Z_\alpha \in I_\alpha \quad \text{aux p points de contraintes.}$$

Ces contraintes sont en fait des contraintes d'exactitude de l'interpolateur. Sous réserve d'avoir un nombre suffisant de données exactes (nombre qui dépend du degré de la FAI–k considérée) correctement situées (Langlais (1989)), la solution existe et est unique. Une fois les contraintes affectées d'une valeur, l'interpolateur n'est autre que le krigeage calculé à partir des (n+p) points.

Cette méthode permet d'obtenir rapidement un estimateur aussi bien dans le cas stationnaire que non–stationnaire. Mais elle présente cependant un inconvénient : le calage aux bornes. En effet, lorsqu'une contrainte Z_i est active la solution de la minimisation conduit à affecter à la contrainte sa borne supérieure ou inférieure. Cette solution n'est pas toujours satisfaisante car en général on sait qu'en ce point de contrainte la valeur de la variable est strictement comprise entre les bornes. Ce comportement quelque peu brutal des contraintes actives conduit à envisager d'autres méthodes évitant cet écueil.

2 – L'ESPERANCE CONDITIONNELLE

Sous réserve d'hypothèse multigaussienne, on pense évidemment à prendre comme estimateur l'espérance conditionnelle, soit :

$$Y^*(x) = E(Y(x) | Y_\alpha , Y_i \in I_i)$$

où α décrit les données exactes et i les contraintes. De plus cette méthode permet de calculer une variance d'estimation. Cette méthode requiert des hypothèses fortes, et pratiquement le nombre de contraintes est limité ; de plus le temps de calcul est très long (exemple: 61 données, 10 contraintes, précision médiocre, 2500 points à kriger en voisinage unique nécessitant 2h30 cpu). Bien que l'espérance conditionnelle soit le meilleur estimateur, on est amené à mettre en oeuvre une méthode plus rapide.

3 – MÉTHODE INCRÉMENTALE

Cette méthode proposée par Matheron affecte à chaque contrainte une valeur dans le cadre d'une fonction aléatoire multigaussienne.

On notera Y_α la gaussienne au point de donnée α et Y_i la gaussienne au point de contrainte i. On décrit d'abord le procédé de simulation :

Contrainte n°1 : $Y_1 = \lambda^\alpha Y_\alpha + R_1 = Y_1^{KS} + R_1$

Y_1^{KS} = krigeage simple au point de contrainte n°1 avec les n données.

On cherche une valeur de Y_1 telle que $Y_1 \in I_1$, c'est-à-dire $R_1 \in J_1$ où J_1 est l'intervalle I_1 translaté du krigeage. Par hypothèse Y_1^{KS} et R_1 sont indépendants ; on peut donc simuler R_1 comme une gaussienne seuillée appartenant à l'intervalle J_1.

On affecte alors au point 1 la valeur Y_1^s qui est la somme du krigeage simple Y_1^{KS} et de la simulation R_1^s. Puis on passe la contrainte suivante :

Contrainte n°2 : $Y_2 = Y_2^{KS} + R_2$

Cette fois Y_2^{KS} est le krigeage simple au point de contrainte n°2 à partir des n données et de la première contrainte :

$$Y_2^{KS} = \lambda^\alpha Y_\alpha + \lambda^1 Y_1^s$$

R_2 est alors simulé comme précédemment.

Contrainte n°p : $Y_p = Y_p^{KS} + R_p$

Y_p^{KS} est le krigeage simple au point de contrainte n°p avec les n données et les (p–1) contraintes déjà traitées ; R_p est simulé à son tour.

Cette méthode permet bien d'obtenir des valeurs aux points de contraintes strictement comprises dans l'intervalle.

Ces simulations aux points de contraintes peuvent être utilisées comme données conditionnantes dans des simulations conditionnelles classiques.

– Estimation

Pour passer à une estimation, on peut envisager de prendre l'espérance conditionnelle du résidu $E(R_i | R_i \in J_i)$. Mais il faudrait alors avoir recours à un cokrigeage et non à un krigeage comme c'était le cas précédemment. En effet, pour la première contrainte, on a maintenant:

$$Y_1^* = Y_1^{KS} + E(R_1 | R_1 \in J_1)$$

Pour la seconde : $Y_2 = Y_2^{CKS} + R_2$ avec : $Y_2^{CKS} = \lambda^\alpha Y_\alpha + \beta^1 Y_1^*$ (cokrigeage simple).

Il faut cette fois introduire la covariance entre Y(x) et Y_1^*, c'est–à–dire entre Y(x) et $E(R_1 | R_1 \in J_1)$, et ainsi de suite au fur et à mesure des itérations.

Dans la pratique, on a pris $Cov(Y(x), Y_1^*) = Cov(Y(x), Y_1)$, covariance a priori.

– Ordre

L'ordre d'apparition des contraintes reste un choix arbitraire de l'utilisateur ; il paraît illusoire de faire la moyenne de toute la combinatoire, ce qui serait peut–être un raffinement superflu. Il faut cependant choisir un ordre "sensé" ; il ne paraît pas raisonnable de commencer par une contrainte seule loin de toute donnée exacte, car seule la moyenne de ces données interviendrait dans le krigeage simple. Il est préférable de choisir une autre contrainte comme point de départ afin que les données exactes interviennent autrement que par leur moyenne.

Un ordre simple à mettre en oeuvre est de commencer par la contrainte telle que la variance du résidu R_i sachant que R_i appartient à J_i est la plus faible. Ce critére est réactualisé à chaque étape. Ce critère prend en compte :

- l'amplitude de l'intervalle, on débute par les contraintes dont l'intervalle est de faible amplitude :

- la valeur du krigeage par rapport à l'intervalle, on débute par les contraintes les plus fortes :

la variance d'estimation au point i avec les n données et les (i–1) contraintes déjà estimées, on débute par les contraintes les mieux conditionnées :

Cette méthode donne des résultats rapides et satisfaisants (valeurs strictement comprises entre les bornes), mais elle dépend de l'ordre de traitement des contraintes. De plus, l'hypothèse multigaussienne est nécessaire en toute rigueur, mais il faut tester si d'un point de vue pratique le résultat est très sensible à cette hypothèse.

C – COMPARAISON DES MÉTHODES SUR UN EXEMPLE

1 – PRÉSENTATION DES DONNÉES

La variable étudiée est l'altitude d'une région dont on connait en fait la topographie sur une maille assez fine, ce qui a permis de choisir des contraintes réalistes. On a choisi 61 données exactes et 10 contraintes en accord avec la topographie connue. Pour fixer les idées sur l'influence de l'information disponible sur l'interpolateur, on a calculé le krigeage à partir des 61 données exactes (figure 1). Le tableau ci-dessous présente les contraintes choisies (bornes inférieure et supérieure) :

n°	x	y	min	max	vrai
62	110.	130.	2000.	2300.	2250.
63	95.	130.	2500.	3000.	2600.
64	85.	230.	2500.	3000.	2850.
65	140.	120.	2500.	3000.	2800.
66	125.	180.	2200.	3000.	2700.
67	215.	75.	3000.	3500.	3330.
68	50.	210.	1400.	1600.	1490.
69	185.	25.	1500.	2500.	1820.
70	15.	195.	1500.	1800.	1740.
71	45.	245.	2100.	2500.	2250.

Comme on peut le constater sur les figures 1 et 3, les 61 données exactes ne suffisent pas au respect des contraintes sauf pour la contrainte n°69 (185., 25.) où le krigeage est compris entre 1500. et 2500.. Il est donc nécessaire d'introduire les contraintes pour obtenir un interpolateur qui respecte mieux l'information. On a comparé les différentes méthodes par une coupe plus lisible qu'une carte.

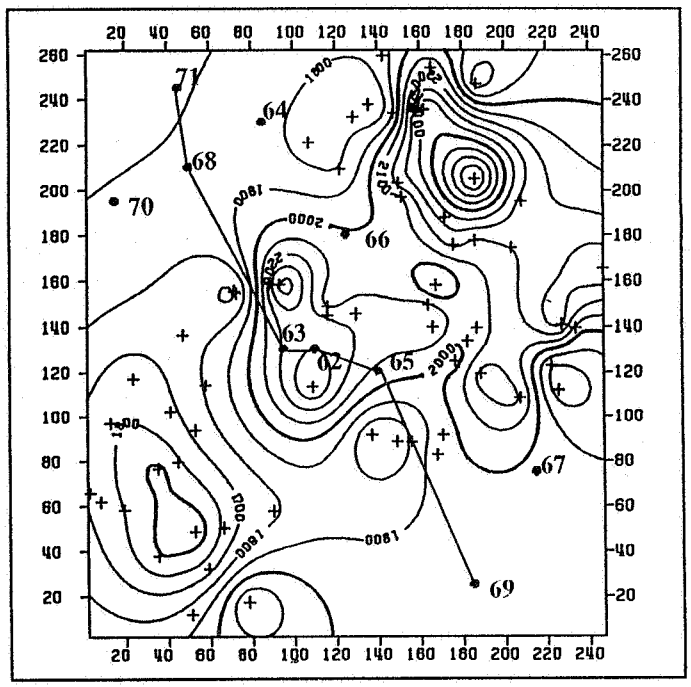


Fig. 1 : Krigeage avec les seules données exactes
+ : données exactes ; ● : contraintes
1111 : valeur réelle : — : coupe

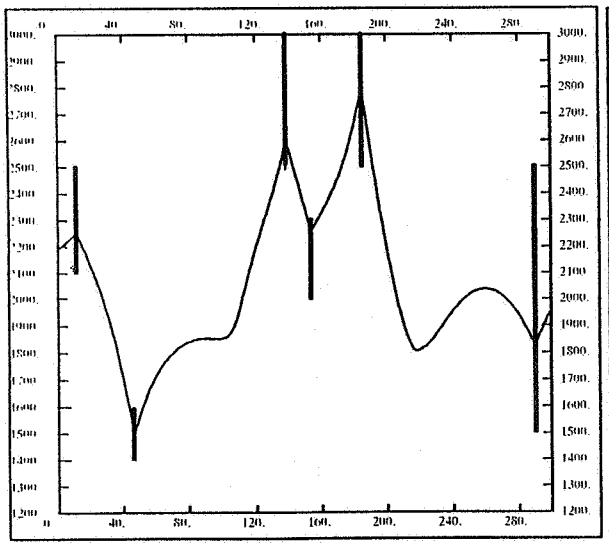


Fig. 2 : Coupe sur le krigeage avec les altitudes réelles

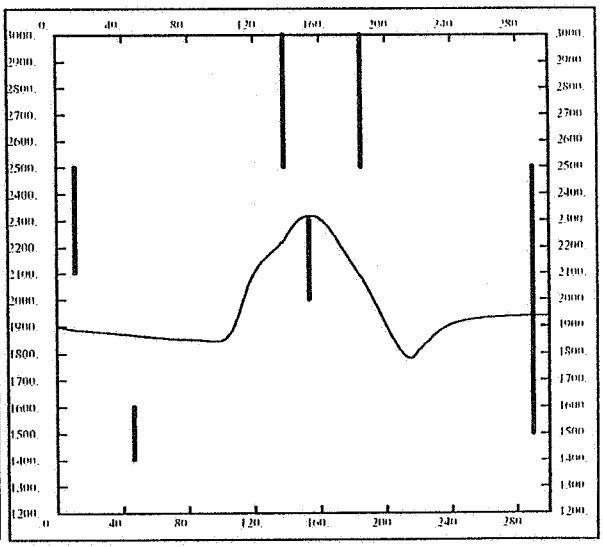
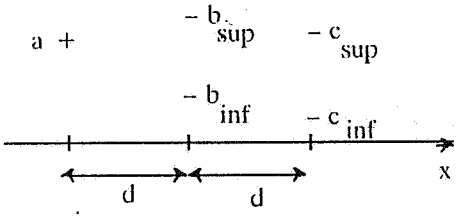


Fig. 3 : Coupe sur le krigeage avec les données exactes

S'il n'y a aucun problème à calculer un variogramme sur les données exactes (figure 4), le problème se pose sur la manière d'introduire l'information supplémentaire que constituent les contraintes. Dans le cas présent, on peut calculer un variogramme avec les données exactes et les vraies valeurs aux points de données (valeurs en général inconnues).

D'autre part, on peut calculer des variogrammes dits inférieur et supérieur qui donnent un encadrement lâche d'un variogramme calculé avec les données exactes et les contraintes affectées de valeurs respectant les bornes. Pour cela, on décrit l'ensemble des couples possibles et lorsqu'on est en présence d'un couple donnée exacte–contrainte ou contrainte–contrainte on prend l'écart le plus grand (respectivement le plus petit) pour calculer le variogramme supérieur (respectivement inférieur). Il faut remarquer qu'en général ce ne sont pas des variogrammes possibles car, selon le couple considéré, ce n'est pas la même borne de la contrainte qui intervient comme le montre l'exemple suivant :



$$\gamma_{inf}(d) = \frac{1}{2}\,(0 + 0)$$

$$\gamma'_{sup}(d) = \frac{1}{2}\left((a - b_{inf})^2 + (b_{sup} - c_{inf})^2\right)$$

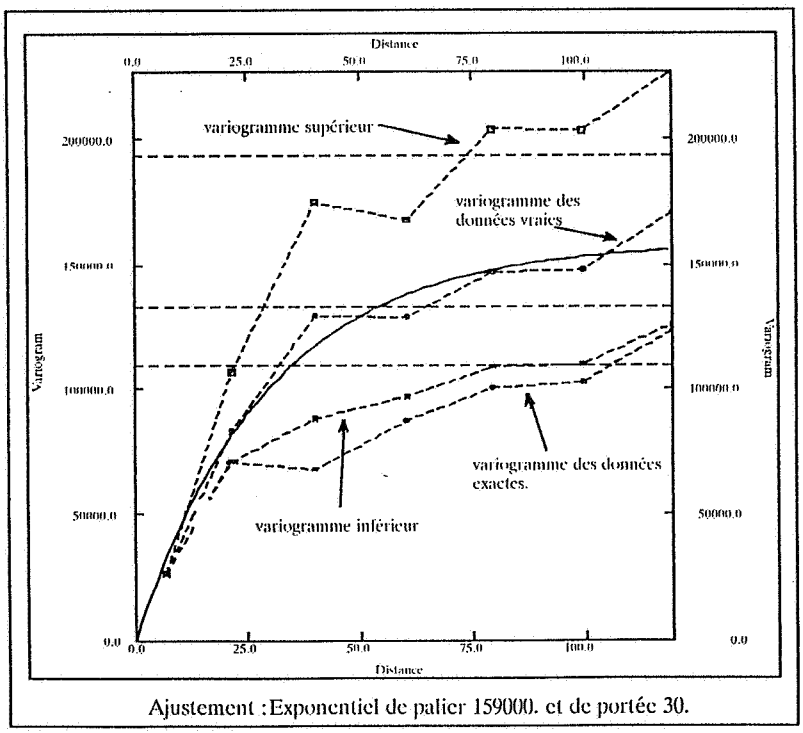


Fig. 4 : Différents variogrammes

Sur la figure 4 on voit immédiatement que le variogramme sur les données exactes n'est pas représentatif des contraintes. Cependant, d'après les variogrammes inférieur et supérieur, on peut supposer que les contraintes n'apporteront pas de grand changement dans la portée du variogramme. Cette hypothèse se confirme lorsqu'on calcule le variogramme à partir des valeurs réelles connues. En effet, dans les deux modèles, les portées sont de 30. et 35., cette différence n'aura que peu d'influence sur l'estimation.

3 – MINIMISATION DE LA FORME QUADRATIQUE

Sur l'exemple choisi, la minimisation de la forme quadratique conduit à neuf contraintes actives (calées à l'une des bornes), seule la contrainte n°69 (185., 25.) est inactive. On remarquera que c'est aussi la seule pour laquelle le krigeage ordinaire avec les données exactes satisfait les bornes. Il est intéressant de remarquer que la minimisation avec le variogramme des données vraies n'amène pratiquement aucun changement. Seule la valeur de la contrainte inactive est légèrement rapprochée du krigeage.

Selon la nature du problème traité, un comportement aussi brutal (9 inégalités calées à l'une de leurs bornes) est peu satisfaisant et conduit à l'emploi d'une autre méthode.

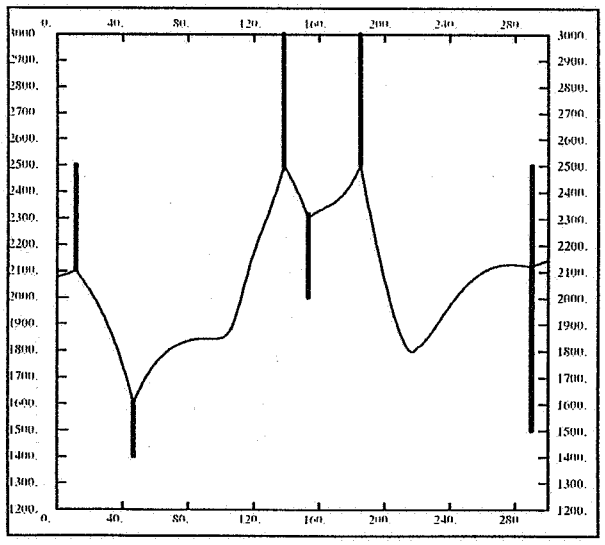


Fig. 5 : Coupe sur l'estimation par minimisation d'une forme quadratique

4 – ESPERANCE CONDITIONNELLE

Cette méthode évite l'ecueil précédent : les valeurs estimées aux points de contraintes sont strictement comprises dans les intervalles correspondants, mais il ne faut pas perdre de vue qu'elle nécessite une forte hypothèse sur les données (hypothèse multigaussienne), et que des problèmes de précision numérique apparaissent rapidement lorsque le nombre de contraintes augmente (12 en pratique) . On remarque que cette méthode fournit un estimation moins lisse que précédemment.

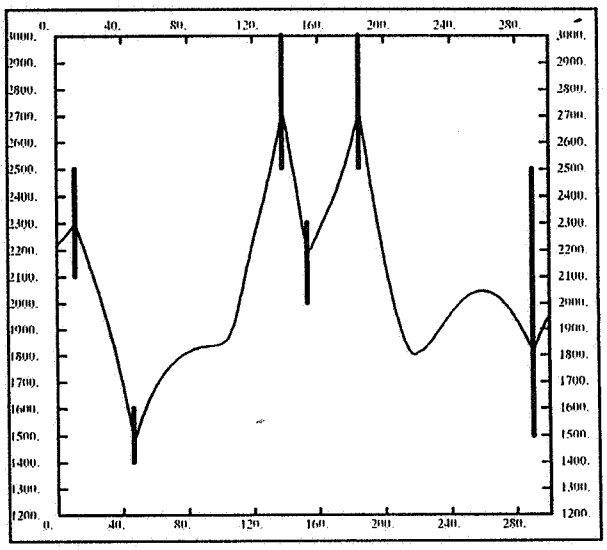


Fig. 6 : Coupe sur l'estimateur espérance conditionnelle

5 – MÉTHODE INCRÉMENTALE

– Estimation

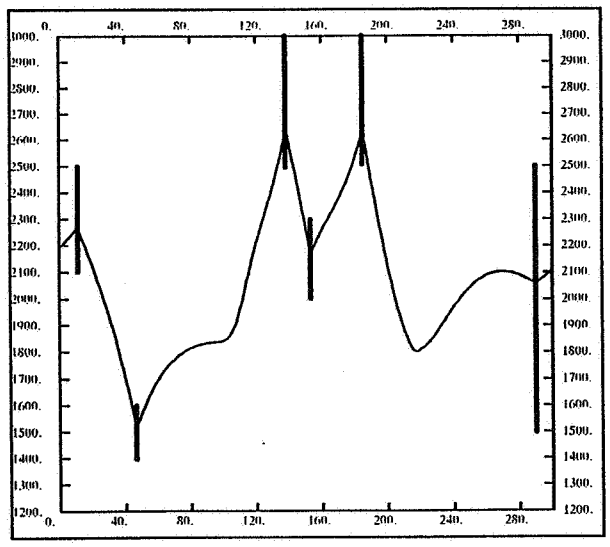


Fig. 7 : Coupe sur l'estimation par la méthode incrémentale

La méthode incrémentale, avec le critère d'ordre préconisé, conduit à traiter les contraintes dans l'ordre suivant : 68, 62, 70, 67, 64, 71, 63, 65, 66 et 69. On débute ainsi par la contrainte n°68 qui est celle dont l'intervalle est de plus faible amplitude, tandis que les n°66 et 69 sont rejetées en fin de traitement du fait de la grande amplitude de leurs intervalles et de leur faible conditionnement par des données environnantes.

– Simulations

Cette méthode permet également d'obtenir des simulations aux points de contraintes, simulations qui peuvent être utilisées comme données conditionnantes pour des simulations conditionnelles classiques. On présente ici deux coupes obtenues en utilisant deux simulations différentes aux points de contraintes et la même simulation non conditionnelle.

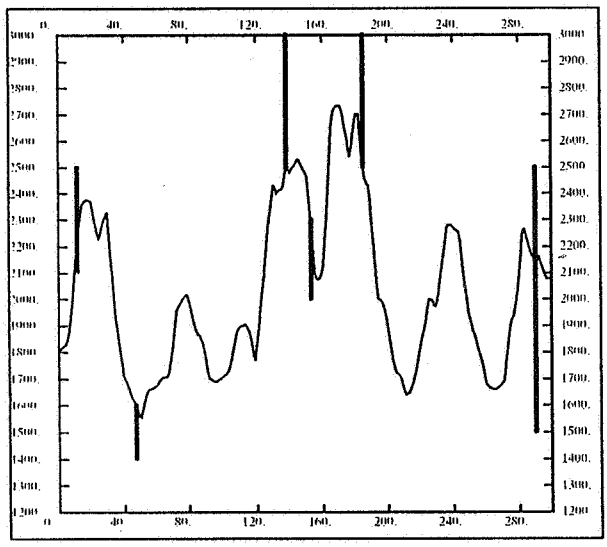


Fig. 8 : Coupe sur une simulation (1)

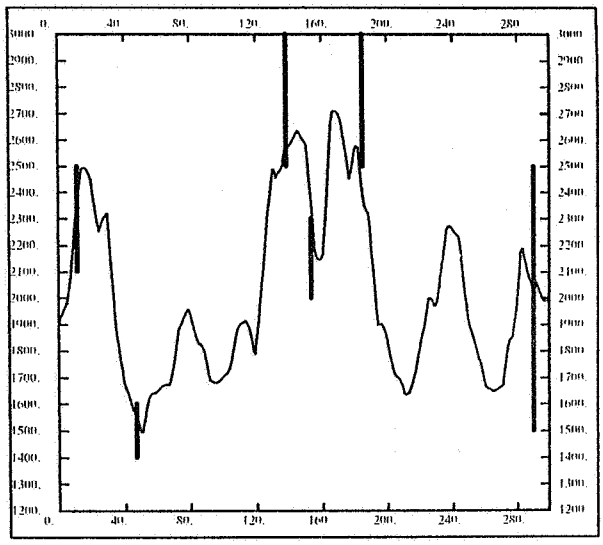


Fig. 9 : Coupe sur une simulation (2)

Le tableau ci-dessous résume les différentes valeurs estimées ou simulées aux points de contraintes.

N°	borne inf.	minimisation sous contraintes	espérance conditionnelle	méthode incrémentale	borne sup.	krigeage avec les données	Simulation 1	Simulation 2
62	2000.	2300.	2172.	2164.	2300.	2303.	2 082.	2 191.
63	2500.	2500.	2724.	2637.	3000.	2212.	2 635.	2 735.
64	2500.	2500.	2724.	2620.	3000.	1839.	2 588.	2 570.
65	2500.	2500.	2724.	2636.	3000.	2074.	2 671.	2 503.
66	2200.	2200.	2480.	2404.	3000.	2010.	2 216.	2 544.
67	3000.	3000.	3470.	3107.	3500.	1971.	3 022.	3 082.
68	1400.	1600.	1472.	1508.	1600.	1887.	1 511.	1 454.
69	1500.	2119.	1807.	2052.	2500.	1941.	2 112.	1 987.
70	1500.	1800.	1655.	1657.	1800.	1890.	1 797.	1 593.
71	2100.	2100.	2298.	2268.	2500.	1906.	2 237.	2 370.

– Anamorphose

Pour appliquer la méthode incrémentale et l'espérance conditionnelle, on a anamorphosé les données brutes : soit calée sur les données exactes, soit en introduisant les contraintes avec une faible pondération, ou encore celle calée sur les valeurs réelles. Il est également intéressant de tester la méthode incrémentale sur les données brutes afin d'observer l'importance de l'hypothèse de distribution gaussienne.

Pour cela on a simplement centré les données brutes. Lorsque l'on compare les résultats obtenus pour deux simulations différentes et l'espérance conditionnelle du résidu sur les données anamorphosées (anamorphose sur les données réelles) et sur les données centrées, on remarque que la simulation, ou l'espérance conditionnelle du résidu, joue un rôle prépondérant et non pas le type d'anamorphose.

		données anamorphosées			données centrées		
		simulation 1	simulation 2	espérance conditionnelle du résidu	simulation 1	simulation 2	espérance conditionnelle du résidu
Données anamorphosées	simulation 1	1	0,948	0,950	0,994	0,963	0,970
	simulation 2	0,948	1	0,972	0,943	0,995	0,978
	espérance conditionnelle du résidu	0,950	0,972	1	0,939	0,970	0,994
Données centrées	simulation 1	0,994	0,943	0,939	1	0,965	9,67
	simulation 2	0,963	0,995	0,970	0,965	1	0,984
	espérance conditionnelle du résidu	0,970	0,978	0,994	0,967	0,984	1

Coefficients de corrélation entrs les différentes estimation ou simulation

On remarque que de façon systématique la corrélation est meilleure à simulation donnée ou espérance conditionnelle et non pas à anamorphose fixée.

D – CONCLUSION

La minimisation sous contraintes fournit une manière simple de résoudre ce problème d'estimation. Elle ne se limite pas au cas stationnaire mais elle présente l'inconvénient de mettre certaines contraintes à leurs bornes, ce qui n'est pas une réponse satisfaisante pour certains problèmes. L'espérance conditionnelle est bien sûr le meilleur estimateur que l'on puisse construire mais elle soulève des problèmes pratiques notamment la précision des calculs. La méthode incrémentale évite le calage aux bornes, mais elle est sensible à l'ordre dans lequel sont traitées les contraintes, elle est peu sensible à l'anamorphose.

Toutes ces méthodes nécessitent une analyse structurale rendue plus délicate par la présence de contraintes, mais quelques outils simples fournissent un moyen de contrôler la modélisation.

Enfin, l'exemple montre que dans toutes les méthodes, les contraintes jouent pleinement leur rôle, les résultats sont peu différents car les contraintes sont en faible nombre.

REFERENCES BIBLIOGRAPHIQUES

Dubrule, O. and Kostov, C., (1986). An interpolation method taking into account inequality constraints: I – Methodology. *Math. Geol.*, 18, No 1, 1986, pp. 33–51.

Journel, A.G., (1974). *Simulations Conditionnelles – Théorie et Pratique*. Thèse de Docteur– Ingénieur, ENSMP.

Kostov, C., (1985). Constrained interpolation in geostatistical applications. M.Sc. Thesis, Dept. of Applied Earth Sciences, Stanford University, Cal., 1985.

Kostov, C. and Dubrule, O., (1986). An interpolation method taking into account inequality constraints: II – Practical approach. *Math. Geol.*, 18, No 1, 1986, pp. 53–73.

Langlais, V., (1989). Krigeage sous contraintes d'inégalités en voisinage glissant. Note interne du Centre de Géostatistique, ENSMP, Fontainebleau.

Maréchal, A., (1978). Gaussian anamorphosis models. Cours du Centre de Géostatistique, ENSMP, Fontainebleau.

Matheron, G., (1970). *La Théorie des Variables Régionalisées et ses Applications*. Cahiers du Centre de Morphologie Mathématique, Fasc. 5, ENSMP, 1970.

Matheron, G., (1972). Quelques aspects de la montée. Note interne du Centre de Géostatistique, ENSMP, Fontainebleau.