

Interpolation optimale et cartographie

par J.P. CHILÈS et G. MATHERON

Centre de Morphologie mathématique
Ecole nationale supérieure des Mines de Paris

Les phénomènes que l'on a à cartographier ont des comportements très divers, mais on y distingue à peu près toujours un aspect aléatoire et un aspect structuré. Le recours à des théories fondées sur les fonctions aléatoires est alors naturel. Mais la détermination de la fonction de covariance, qui synthétise les caractéristiques structurales de la variable étudiée, n'est pas automatisable dans le cadre de la théorie classique, surtout lorsque le phénomène présente une dérive (ou trend). La théorie récente des fonctions intrinsèques généralisées permet, elle, de prendre en compte les dérives, et de déterminer automatiquement la covariance. Cette théorie généralise celle de N. Wiener et s'applique aisément dans l'espace à deux dimensions. Après un exposé rapide des grandes lignes de la théorie, on donne un exemple d'application à la cartographie de la topographie.

INTRODUCTION

Dans des domaines d'activité très divers, on a l'habitude de synthétiser sous forme de cartes l'information dont on dispose au sujet de tel ou tel phénomène. Par exemple, il pourra s'agir de présenter sous forme de courbes de niveau $z = C^{ste}$ la valeur $z(x)$ en chaque point x du plan de la grandeur à laquelle on s'intéresse (cote topographique, pression, température, densité de population, puissance ou teneur d'une formation géologique minéralisée ou pétrolifère, etc.) Dans certaines situations, la valeur $z(x)$ est connue ou peut l'être facilement en chaque point x : c'est le cas, par exemple, d'une carte topographique lorsque l'on dispose des photographies aériennes. Souvent, au contraire, on ne connaît les valeurs $z(x_i) = z_i$ de z qu'en un nombre limité de points $x_i, i = 1, 2, \dots, N$ (stations météorologiques, sondages pétroliers, etc.) qu'il n'est pas possible de multiplier indéfiniment en raison d'un prix de revient élevé.

Il s'agit alors de tirer le meilleur parti possible d'une information rare et coûteuse, donc d'utiliser une méthode d'interpolation optimale entre les points expérimentaux. Cela suppose évidemment le choix d'un critère d'optimalité. Plus précisément, pour estimer la valeur inconnue $z(x)$ en un point donné x , nous devons préciser :

- la classe C des estimateurs $z^*(x) = f(z_i)$ considérés comme utilisables,
- le critère d'optimalité permettant de choisir le meilleur estimateur de cette classe.

Un grand nombre de critères possibles ont été proposés sinon effectivement utilisés. Nous nous limiterons ici au plus simple d'entre eux, qui consiste à **minimiser la variance d'estimation**. Parler de variance suppose, évidemment, le recours à une interprétation probabiliste : la fonction $z(x)$ qu'il s'agit de cartographier est considérée comme une réalisation d'une fonction aléatoire (FA) $Z(x)$. Les estimateurs possibles sont alors de la forme $f(Z_1, \dots, Z_N), f \in C$ (avec $Z_i = Z(x_i)$),

et le problème consiste à trouver la fonction f de N variables appartenant à la classe C et réalisant le minimum de :

$$\sigma_{est}^2 = E [(Z(x) - f[Z_1, \dots, Z_N])^2] \quad (f \in C)$$

Pour pouvoir résoudre ce problème, nous devons être capables de calculer, en principe pour chaque $f \in C$ l'espérance mathématique écrite ci-dessus : cela suppose, à titre de prerequisite, la connaissance de certaines caractéristiques de la FA, en nombre d'autant plus élevé que la classe C est plus vaste. Mais alors se pose le problème de l'inférence statistique : il est nécessaire que les données disponibles (les Z_i) permettent une estimation au moins approximative des paramètres dont nous avons besoin pour calculer les variances d'estimations.

Cette contrainte nous oblige, en pratique, à restreindre considérablement la classe C à l'intérieur de laquelle nous pourrions travailler. Si C est trop grande, en effet, l'inférence statistique n'est pas possible. Par exemple, si C est la classe la plus générale possible, c'est-à-dire celle de toutes les fonctions mesurables $f(Z_1, \dots, Z_N)$ de N variables, l'estimateur optimal est l'espérance conditionnelle de $Z(x)$ à Z_1, \dots, Z_N fixés, mais pour calculer effectivement cette espérance conditionnelle, il faudrait connaître la loi conditionnelle correspondante, et cela les données disponibles ne le permettent pas en général.

En vue de diminuer le nombre des prerequisites indispensables, il convient donc de réduire la classe C : en pratique, on se limitera à la classe des combinaisons linéaires des données expérimentales, soit même à une sous-classe de celle-ci pour des raisons que nous allons voir maintenant.

Le meilleur estimateur linéaire affine

Si nous prenons pour C la classe des combinaisons linéaires affines

$$f(Z_1, \dots, Z_N) = a + \sum \lambda_i Z_i$$

il est facile de voir que l'estimateur optimal de $Z(x)$ est de la forme

$$Z^*(x) = m(x) + \sum \lambda_i(x) [Z_i - m_i]$$

avec $m(x) = E(Z(x))$, $m_i = E(Z_i)$ et des coefficients λ_i déterminés par :

$$(1) \quad \sum_i \lambda_i \sigma_{ij} = \sigma_{jx}$$

(σ_{ij} = Covariance (Z_i, Z_j), σ_{jx} = Cov ($Z_j, Z(x)$). Cet estimateur est le « meilleur predicteur linéaire » classique depuis les travaux de N. Wiener. Du point de vue de l'inférence statistique, sa mise en

œuvre nécessite la connaissance des moments d'ordre 1 et 2 (les espérances et les covariances) au nombre de $N+1 + (N+1)(N+2)/2$. S'il n'y a pas d'autre source d'information que les N valeurs numériques prises par les Z_i , l'estimation de ces moments n'est donc pas possible en général, à moins d'introduire des hypothèses supplémentaires permettant de réduire le nombre des paramètres à estimer.

Telle est la raison pour laquelle on introduit souvent l'hypothèse suivante (dite **stationnarité d'ordre 2**) :

- Les espérances sont égales à une même constante $E(Z_i) = E(Z)$.
- La covariance de $Z(x)$ et $Z(y)$ est une fonction $\sigma(y-x) = \sigma(h)$ dépendant seulement du vecteur $h = y-x$ et non pas séparément des deux points d'appui x et y .

Si cette hypothèse est vérifiée, le nombre des paramètres du problème se trouve substantiellement réduit, et l'on montre que leur estimation est possible. Malheureusement, beaucoup de phénomènes réels sont incompatibles avec cette hypothèse de stationnarité : en cartographie marine, par exemple, on sait d'avance que la profondeur augmente, en moyenne, lorsqu'on s'éloigne de la côte ; ou encore, en recherche pétrolière, on s'intéresse à une structure particulière (un dome, etc.) et de ce fait même — la stationnarité n'est pas recevable.

Le Krigeage universel

En face d'un phénomène manifestement non stationnaire, la première idée qui vient à l'esprit consiste à essayer de l'interpréter, au moins localement, comme la somme d'une « dérive » (fonction non constante, mais régulière et lentement variable dans l'espace) et d'une fonction aléatoire stationnaire. Le modèle correspondant est :

$$Z(x) = Y(x) + m(x)$$

où $Y(x)$ est une FA stationnaire (d'espérance nulle et de covariance $\sigma(h) = \sigma(y-x)$) et $m(x)$ un polynome de degré k donné à coefficients inconnus, soit :

$$m(x) = \sum a_i f^i(x)$$

(les f^i sont les monomes de degré $\leq k$ et les a_i des coefficients inconnus). En l'absence de toute information sur les coefficients a_i , on ne peut minimiser la variance d'estimation que si celle-ci ne dépend pas des a_i : il est facile de voir que cette condition équivaut à $E(Z^*(x) - Z(x)) = 0$,

quels que soient les a_i . La classe C_k des estimateurs correspondants est celle des combinaisons linéaires

$$Z^*(x) = \sum \lambda_i Z_i$$

vérifiant la condition dite **d'universalité** :

$$(2) \quad \sum \lambda_i f^i(x_i) = f^i(x)$$

En exprimant le minimum de la variance d'estimation compte tenu de la condition (2), on obtient le système :

$$\sum_i \lambda_i \sigma_{ij} = \sigma(x_j - x) + \sum_l \mu_l f^l(x_j)$$

qui diffère de (1) par l'intervention au second membre des multiplicateurs de Lagrange μ_l . L'estimateur correspondant porte le nom de **Krigeage Universel**.

Dans le cadre de ce modèle, l'inférence statistique de la covariance $\sigma(h)$ est plus difficile que dans le cas stationnaire, à cause des biais qu'introduit l'estimation préalable des coefficients a_i : la situation est la même qu'en statistique ordinaire lorsque l'on procède à l'estimation de la variance en présence d'une moyenne inconnue, mais les biais sont beaucoup plus importants. De plus, ce modèle présente deux défauts assez graves :

- La séparation de $Z(x)$ entre une « dérive » fonctionnelle $m(x)$ et une partie aléatoire $Y(x)$ est passablement artificielle (elle correspond, cependant, assez bien à la distinction usuelle en géophysique, entre anomalie régionale et anomalie locale).
- L'hypothèse de stationnarité stricte des résidus $Y(x)$ est assez forte, et n'est pas toujours compatible avec la réalité expérimentale.

Mais en fait, ces deux hypothèses ne sont pas réellement utiles et il est possible de s'en affranchir.

Les fonctions aléatoires intrinsèques d'ordre k

De fait, ce qui caractérise vraiment le Krigeage Universel, ce sont les deux circonstances suivantes :

- on recherche l'estimateur optimal dans la classe C_k des combinaisons linéaires vérifiant les conditions d'universalité (2)
- la variance d'estimation des estimateurs de cette classe s'exprime à l'aide d'une fonction $\sigma(h)$ d'un seul argument vectoriel h .

Or ceci n'implique ni la dichotomie $Z(x) = m(x) + Y(x)$, ni la stationnarité des résidus $Y(x)$. Considérons, en effet, l'erreur d'estimation $Z(x)$

— $\sum \lambda_i Z(x_i)$. En changeant les notations, nous voyons que cette erreur est une combinaison linéaire de la forme $\sum \lambda'_\alpha Z(x_\alpha)$ où les x_α , au nombre de $N+1$, sont les points expérimentaux x_i et le point x à estimer) et que la condition d'universalité s'écrit :

$$(3) \quad \sum \lambda'_\alpha f^\alpha(x) = 0$$

Convenons de dire qu'une combinaison linéaire attribuant le poids λ'_α au points x_α est **autorisée à l'ordre k** si elle vérifie la condition (3), c'est-à-dire annule les polynômes de degré inférieur ou égal à k .

La classe C_k des estimateurs du krigeage universel est ainsi définie par la condition que l'erreur d'estimation doit être une combinaison linéaire autorisée à l'ordre k . Dès lors, pour rendre possible l'inférence statistique, il suffit d'imposer la condition de stationnarité aux seules combinaisons linéaires autorisées (et non pas à la FA elle-même). Cette condition s'exprime en disant que pour toute combinaison linéaire autorisée (attribuant les poids λ'_α à des points donnés x_α), la FA :

$$x \rightarrow \sum \lambda'_\alpha Z(x + x_\alpha)$$

est stationnaire (en x). On montre qu'il existe alors une fonction $K(h)$ appelé **covariance généralisée** telle que l'on ait :

$$(4) \quad E[(\sum \lambda'_\alpha Z(x_\alpha))^2] = \sum_{\alpha, \beta} \lambda'_\alpha \lambda'_\beta K(x_\alpha - x_\beta)$$

pour toute combinaison linéaire autorisée.

On peut d'ailleurs remplacer $Z(x)$ par $Z(x) + \sum A_i f^i(x)$ (avec des coefficients A_i constants ou aléatoires) sans modifier la valeur des combinaisons linéaires autorisées. Le véritable instrument de travail est donc ici, non pas $Z(x)$ elle-même, mais la classe de toutes les FA égales à $Z(x)$ à un polynôme près de degré $\leq k$ et à coefficients aléatoires. C'est cette classe d'équivalence que nous appelons FA Intrinsèque d'ordre k , ou FAI- k .

Si $Y(x)$ est un élément quelconque de cette classe, les autres sont de la forme :

$$Z(x) = Y(x) + \sum A_i f^i(x)$$

La notion de dérive est donc implicitement contenue dans celle de FAI- k , mais il n'est nullement nécessaire de supposer $Y(x)$ stationnaire, et, en général, une FAI- k n'admet pas de représentation stationnaire. Par exemple, le mouvement brownien est une FAI-0 (c'est-à-dire une FA à accroissements stationnaires) admettant une covariance généralisée de la forme $K(h) = -a|h|$.

Mais ce n'est en aucune façon une FA stationnaire.

Du fait que l'erreur est, par construction, toujours une combinaison linéaire autorisée pour les estimateurs de la classe C_k , la variance d'estimation ne dépend que de $K(h)$: l'estimateur optimal s'obtient donc en résolvant le même système que dans le cas de krigeage universel. Il suffit simplement de remplacer $\sigma(h)$ par $K(h)$.

Les covariances polynomiales

Il faut bien voir que les FAI-k ont un degré de généralité beaucoup plus élevé que les FA stationnaires ordinaires. Par exemple, si l'on prend $k + 1$ coefficients positifs a_0, \dots, a_k , il existe toujours une FAI-k admettant la covariance généralisée

$$(5) \quad K(h) = \sum_{i=0}^k (-1)^{i+1} a_i |h|^{2i+1}$$

alors que cette fonction n'est manifestement pas une covariance (stationnaire) au sens usuel.

Ce modèle (5) à covariance généralisée polynomiale se prête bien à une procédure entièrement automatisée d'inférence statistique, car les coefficients a_i interviennent linéairement et sont donc particulièrement faciles à estimer. On peut ajouter un terme supplémentaire représentant un bruit blanc (associé soit à des erreurs de mesures, soit à une microrégionalisation de très faible portée) et on obtient ainsi le modèle à $k + 2$ paramètres utilisé dans le programme de cartographie automatique élaboré par le Centre de Morphologie Mathématique.

A l'usage, il est apparu que l'on obtenait très fréquemment un accord raisonnable avec les données expérimentales sans dépasser l'ordre $k = 1$ ou 2 (pas toujours, cependant, et il y a des types de données qui se prêtent mal à une représentation par le modèle à covariance polynomiale). Les anisotropies éventuelles du phénomène sont, en effet, bien prises en charge par la « dérive » (toujours implicitement présente dans le modèle FAI-k, comme on l'a vu ci-dessus), et la modulation des coefficients a_i permet l'ajustement à des gammes très variées de phénomènes.

Sur les figures 1, 2 et 3, on peut voir des réalisations de FAI-gaussiennes d'ordre 0,1 et 2 respectivement (les représentations choisies sont celles pour lesquelles $Z(x)$ et ses dérivées jusqu'à l'ordre $k-1$ s'annulent au centre de la figure). Pour $k = 1$ et plus encore pour $k = 2$, on est manifestement très loin de la stationnarité au

sens usuel. On a plutôt l'impression de structures (en col ou en selle) bien individualisées. Il est remarquable que des phénomènes de ce genre puissent être représentés par un formalisme pratiquement identique à celui de la stationnarité (la seule restriction que l'on doit s'imposer consiste à ne jamais manipuler des des combinaisons linéaires autorisées) et cela suggère que le domaine d'application pratique des FAI-k doit être pratiquement illimité.

Un exemple : mémorisation d'une carte I.G.N.

Les cartes topographiques obtenues par restitution photogrammétrique sont assez précises, mais demandent un travail assez long. Or beaucoup d'applications n'exigent pas une très grande précision, si bien qu'on peut se demander si des méthodes plus rapides ne pourraient pas donner des résultats satisfaisants. Dans ce but nous avons fait une étude expérimentale d'une partie de la carte de Noirétable au 1/25 000. L'idée consiste à ne retenir qu'un échantillonnage photogrammétrique de quelques centaines de points et à reconstituer la carte à partir de ces points par une méthode d'interpolation. Nous avons comparé trois schémas d'échantillonnage différents.

Il va de soi que la méthode d'interpolation a une grande influence sur la qualité des résultats. Si on veut limiter au strict nécessaire le nombre de points à retenir, il est indispensable d'utiliser une méthode qui tire le meilleur parti possible de l'information qui lui est fournie. La théorie des FAI-k telle qu'elle a été développée au Centre de Morphologie mathématique de Fontainebleau, répond à cette condition. Elle a déjà été appliquée dans de nombreux domaines comme l'estimation minière ou forestière, la bathymétrie, la météorologie... En topographie, des méthodes analogues, mais moins générales, ont été développées, notamment par K. Kraus et E.M. Mikhail (voir réf. [10] et [5]).

Nous avons cherché à visualiser sur l'exemple de Noirétable l'influence du choix du schéma d'échantillonnage sur la qualité de la carte. Pour cela l'I.G.N. nous a fourni les données relatives à trois schémas d'échantillonnage photogrammétrique différents :

- Le premier comprend 219 points caractéristiques situés sur des lignes de niveau ou correspondant à des extrema.
- Le second comprend 420 points situés sur des lignes de crête ou des talwegs.
- Le troisième comprend les 420 points situés sur ces lignes caractéristiques, ainsi qu'un

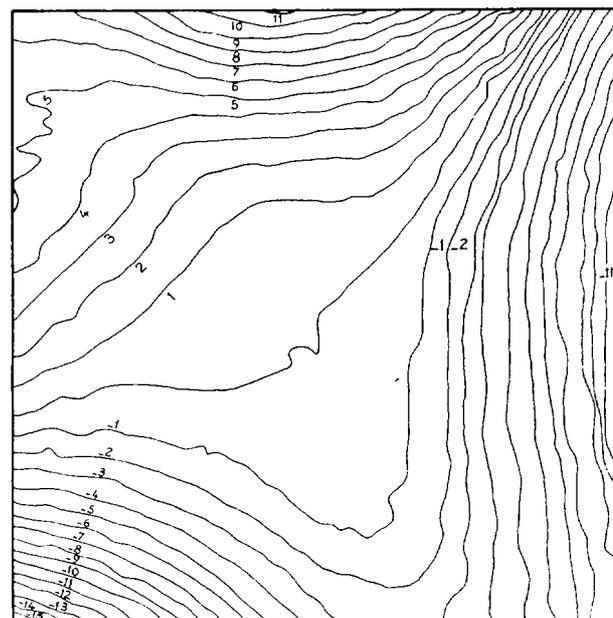
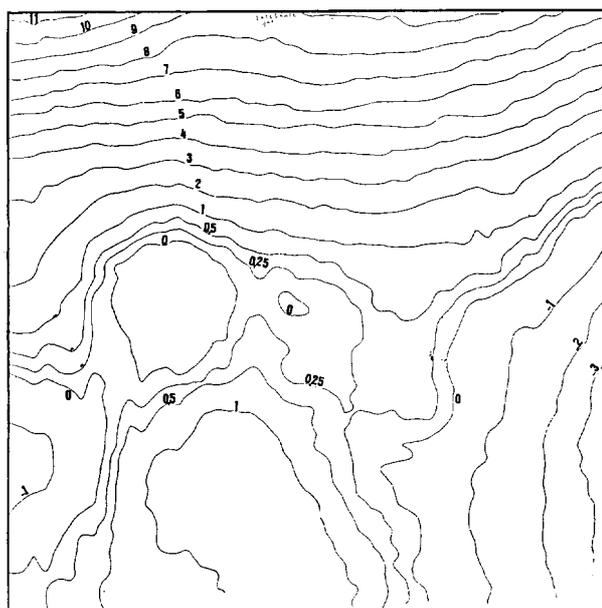
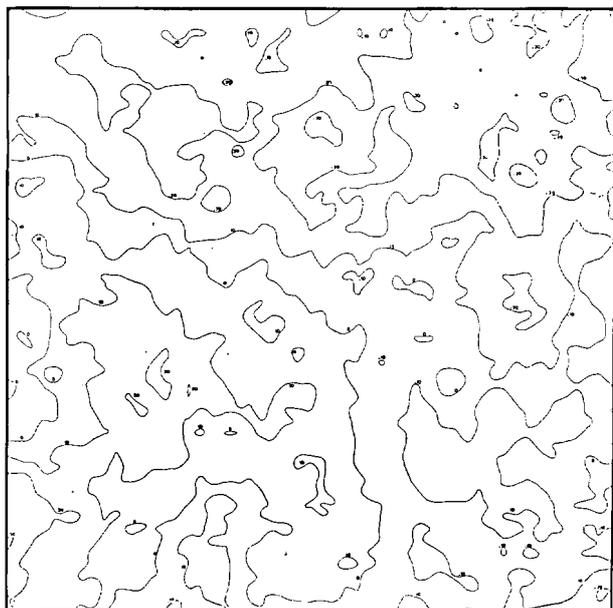


Fig. 1, 2 et 3. — Réalisations de FAI — k de covariance généralisée $K(h) = (-1)^{k+1} |h|^{2k+1}$ avec $k = 0$ (en haut à gauche) $k = 1$ (ci-dessus) $k = 2$ (ci-contre) — d'après J.P. Orfeuil [9].

semis complémentaire de 153 points, soit 573 points ou total ; c'est à peu près la somme des deux schémas précédents.

On aurait bien sûr pu tester d'autres schémas d'échantillonnage, notamment un échantillonnage à maille régulière. Mais cette étude ne prétend pas être exhaustive et ces trois schémas donnent déjà de bonnes indications.

Nous avons donc tracé trois cartes, correspondant à chacun des jeux de données. Le tracé proprement dit est fait par des interpolations simples entre les nœuds d'une grille régulière. Ce sont les valeurs aux nœuds de cette grille qui sont estimées par krigeage. Nous avons pris une maille de 1,5 mm au 1/25 000, correspondant par conséquent à une maille de 37,50 m sur le terrain. Les cartes reproduites ici sont agrandies ; elles sont environ au 1/17 000 et non au 1/25 000.

L'écart-type d'estimation, qui caractérise la précision de l'estimation, est en moyenne de 5 m environ.

L'étude des cartes topographiques présente l'avantage qu'on connaît la carte réelle, ou du moins une carte qui en est très approchée, obtenue par restitution photogrammétrique. C'est en prenant cette carte comme référence que nous allons examiner les cartes obtenues à partir des différents schémas d'échantillonnage. Il va de soi que les trois cartes tracées ici ne pourront pas permettre de retrouver tous les détails, puisqu'elles sont élaborées à partir de quelques centaines de données seulement, alors qu'il en faudrait plusieurs milliers si on voulait numériser avec précision les courbes de niveau.

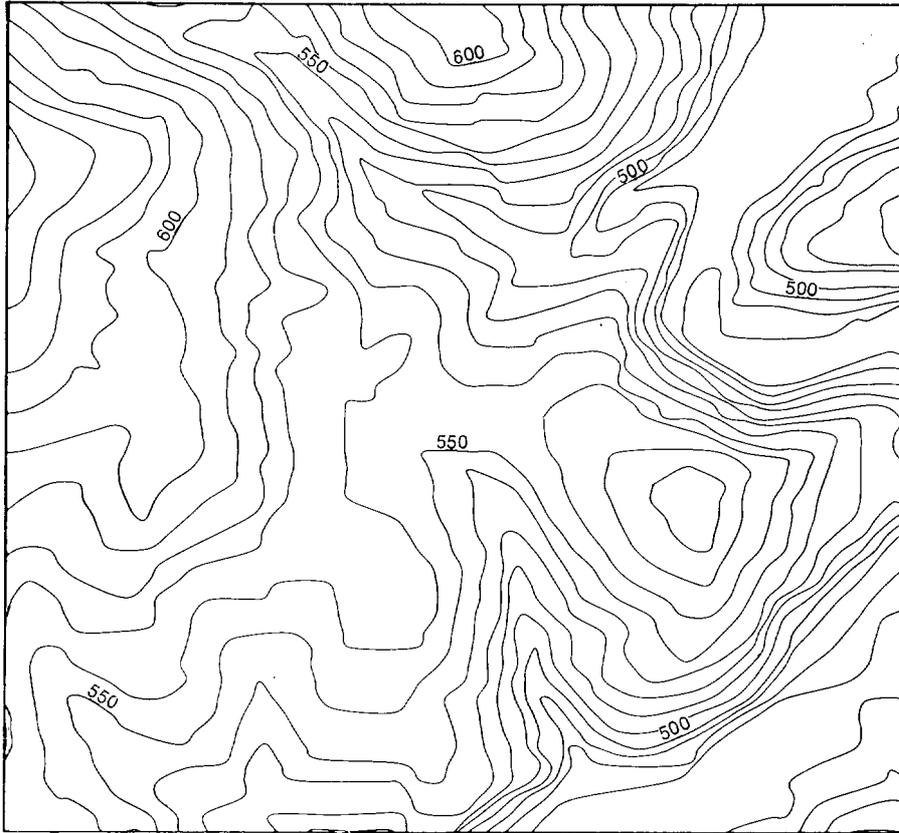


Fig. 4. — Carte obtenue par restitution photogrammétrique.

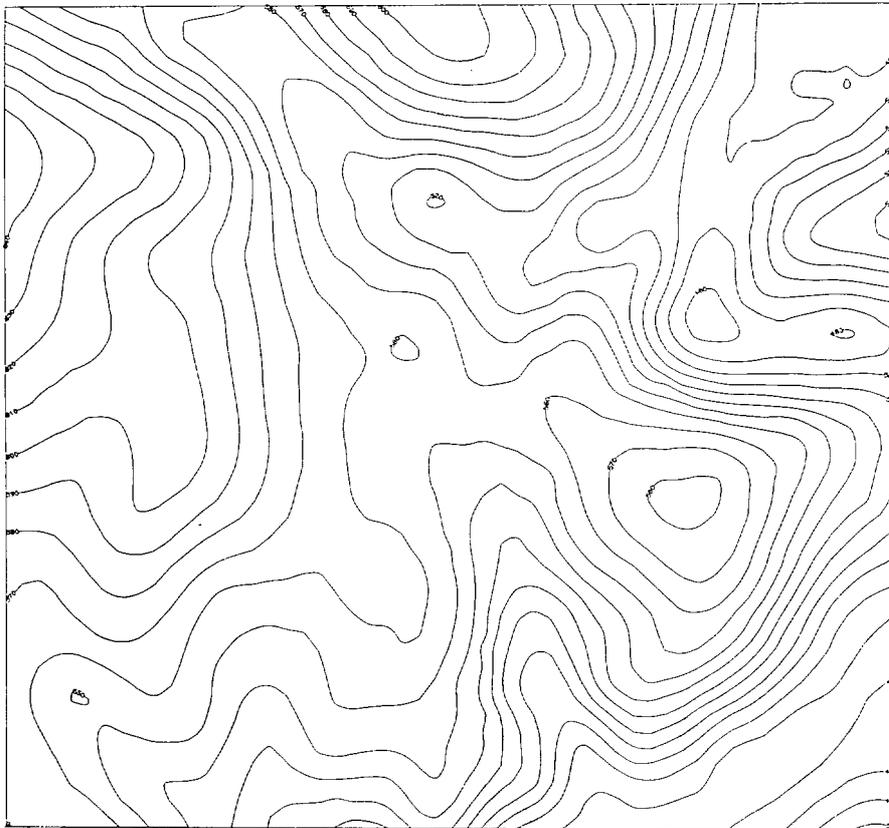


Fig. 5. — Carte de la même région obtenue à partir d'un échantillonnage de 219 points situés sur des lignes de niveau ou correspondant à des extrêmes.

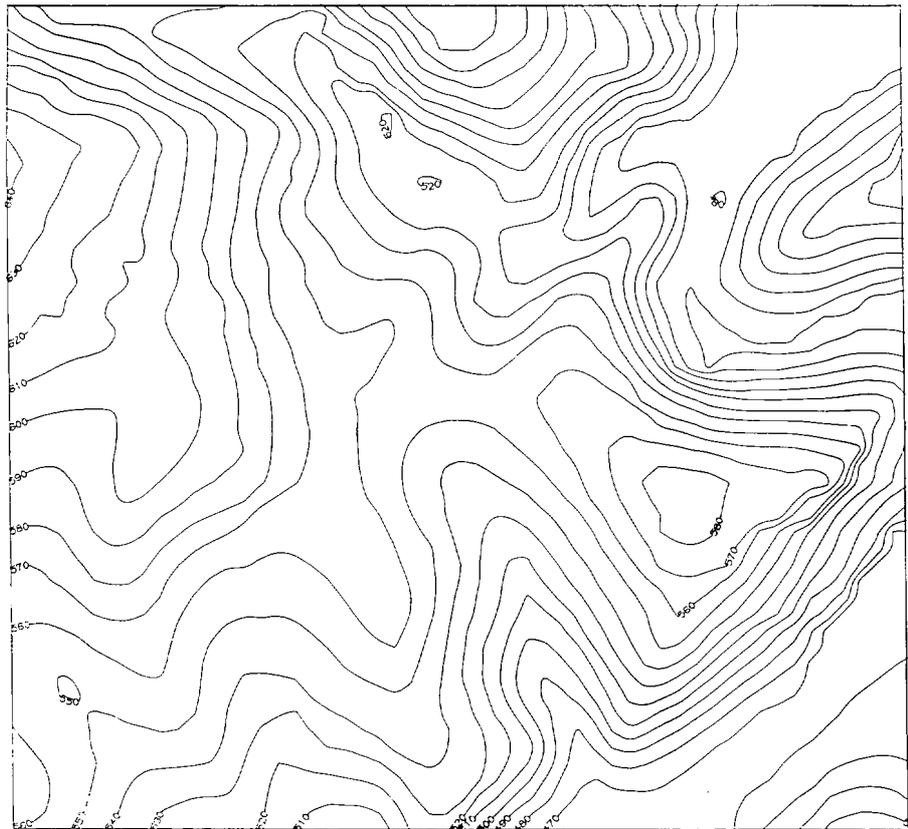


Fig. 6. — Carte de la même région obtenue à partir d'un échantillonnage de 420 points situés sur des lignes de crête ou des talwegs.

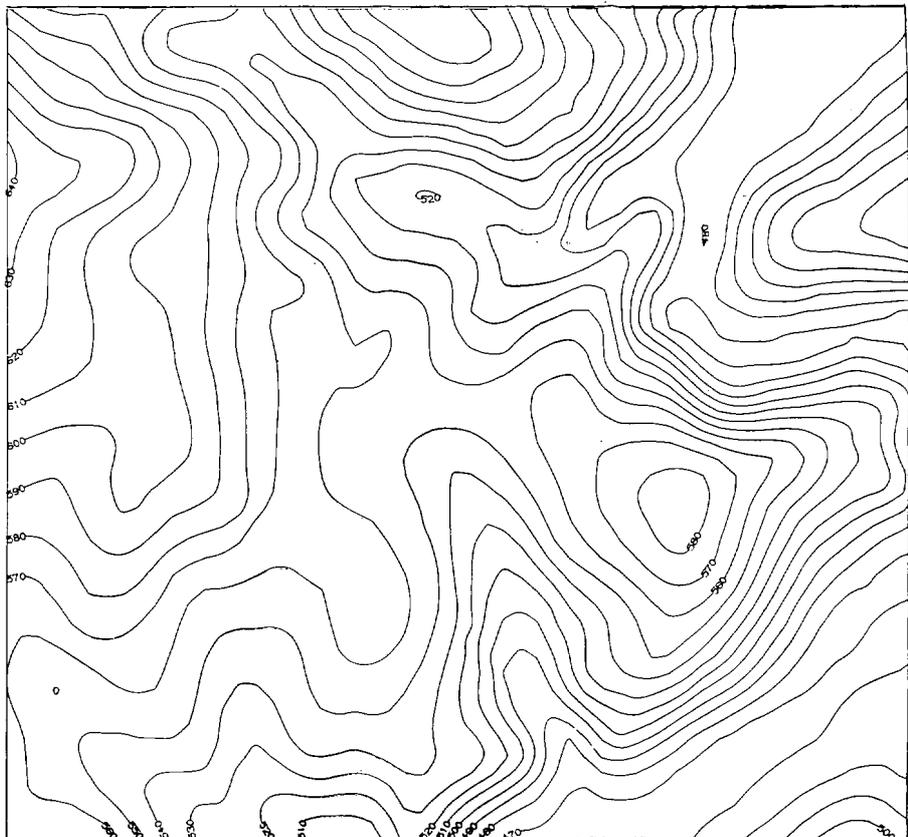


Fig. 7. — Carte de la même région obtenue à partir d'un échantillonnage de 420 points de la figure 6 complétés par un semi complémentaire de 153 points.

Examinons donc les différentes cartes :

- La première, tracée à partir de 219 points caractéristiques donne une bonne reconstitution des lignes de niveau ; ce n'est pas surprenant, puisque la plupart des points caractéristiques se trouvent sur des lignes de niveau. Si on veut donc simplement, en chaque point de la carte, lire l'altitude correspondante avec une assez bonne précision, cet échantillonnage limité donne déjà de bons résultats. Cependant les lignes de niveau sont arrondies, si bien que les traits caractéristiques du relief ressortent mal.
- La seconde, tracée à partir d'un échantillonnage de 420 points des lignes de crête et des talwegs, suit plus mal les lignes de niveau réelles, mais fait ressortir ces lignes caractéristiques ; elle les exagère même un peu trop.
- La troisième, qui intègre à peu près les deux schémas d'échantillonnage précédent, a les qualités des deux premières cartes sans avoir leurs défauts : on obtient un bon tracé des lignes de niveau, et les éléments caractéristiques du relief apparaissent bien.

Il est sans doute possible de réduire l'échantillonnage des lignes caractéristiques ; mais on ne peut guère espérer faire cette carte avec moins de 400 points.

Le krigeage permet donc de reconstituer correctement des cartes topographiques à partir d'un échantillonnage photogrammétrique limité. Si on veut seulement pouvoir lire avec une précision correcte l'altitude de la surface topographique en tout point de la carte, un échantillonnage limité à des points caractéristiques situés surtout sur les lignes de niveau donne de bons résultats. Si par contre on veut également pouvoir lire le relief, c'est-à-dire si la carte doit faire ressortir les traits caractéristiques du relief, il est indispensable d'ajouter un échantillonnage des lignes de crête et des talwegs. Il va de soi que le krigeage ne crée pas d'information, mais se contente d'utiliser au mieux l'information qui lui est fournie. Il ne prétend donc pas remplacer la restitution photogrammétrique lorsque l'on a besoin d'une carte très précise, ou alors il faut faire un échantillonnage très serré. Mais dans les nombreux cas où une grande précision n'est pas indispensable, cette méthode permet d'accélérer la production des cartes. Ceci nécessite bien sûr un calcul sur ordinateur, mais celui-ci n'est pas très long. A titre indicatif, le krigeage d'une grille 61×61 comme celles qui ont été calculées pour cette étude demande sur un ordinateur CDC 6600 environ 60 secondes de travail en

mémoire centrale. L'ensemble des opérations, de l'analyse structurale au krigeage est entièrement automatisé (voir réf. [4]).

Un autre intérêt de cette méthode est de permettre une compression de l'information. En effet, si on veut digitaliser une carte, il suffit de stocker les quelques centaines de points caractéristiques retenus au lieu de stocker une digitalisation des lignes de niveau qui demanderait plusieurs milliers de points. Ceci n'est peut-être pas justifié pour des cartes que l'on ferait tracer souvent, car il faut chaque fois faire le calcul de la carte. Mais cette solution serait valable pour des cartes peu demandées, ou pour des cartes exigeant des mises à jour ; il suffit dans ce dernier cas de changer quelques points caractéristiques au lieu de refaire la discrétisation des lignes de niveau.

BIBLIOGRAPHIE

- [1] J.P. CHILES et P. CHAUVET : Le krigeage — méthode pour la cartographie des fonds marins. *Revue Hydrographique Internationale*, vol. LII, n° 1, janvier 1975.
- [2] P. DELFINER et J.P. DELHOMME : Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride. Colloque sur l'élaboration des projets d'utilisation des ressources en eau sans données suffisantes. UNESCO — AIHS — OMN — Madrid 1973.
- [3] P. DELFINER et J.P. DELHOMME : Optimum Interpolation by Kriging. *Proceedings of NATO Advanced Study Institute for Display and Analysis Spatial Data*. July 1973. Ed. Wiley and Sons, London.
- [4] P. DELFINER et J.P. DELHOMME : Présentation du programme BLUEPACK. Centre de Morphologie Mathématique de Fontainebleau. Note interne, 1973.
- [5] K. KRAUS and E.M. MIKHAIL : Linear least-squares interpolation and filtering in computational photogrammetry. Presented paper, Commission III, I.S.P. Ottawa, 1972.
- [6] G. MATHERON : Les variables régionalisées et leur estimation. Masson et C^{ie} Paris 1965.
- [7] G. MATHERON : La théorie des variables régionalisées et ses applications. *Les Cahiers du Centre de Morphologie Mathématique*, Fontainebleau 1970.
- [8] G. MATHERON : The Intrinsic Random functions and their applications — *Adv. In App. Prob.* — 5-439-468 — 1973.
- [9] J.P. ORFEUIL : Simulation du Wiener-Lévy et de ses intégrales. Centre de Morphologie Mathématique de Fontainebleau. Note interne, 1972.
- [10] P.B. STEWARDSON, K. KRAUS and D.C. GSELL : DACS — Digital Automatic Contouring System. XII International Congress for Photogrammetry, Ottawa 1972.