

INTRODUCTION AU KRIGEAGE A L'USAGE DES HYDROLOGUES

Par Ch. OBLED

Professeur à l'ENS d'Hydraulique de Grenoble
Institut de Mécanique
B. P. 68 - 38402 St Martin d'Heres Cedex

RESUME :

Cet article tente d'abord de rattacher l'approche géostatistique et la technique de krigeage à des méthodes plus familières à l'hydrologue, comme la corrélation multiple et l'analyse des séries chronologiques. Ensuite, on développe le krigeage dans les cas les plus couramment utilisés en hydrologie en analysant le cas de l'effet de pépite et des données entachées d'erreur. Enfin, on fournit quelques idées d'interprétation de la méthode en la formulant différemment (krigeage duale) et en la rattachant à des techniques d'interpolation courantes comme l'interpolation polynomiale et les fonctions splines.

ABSTRACT :

In a first part, an attempt is made to connect the geostatistical approach and the kriging technique with methods more commonly known by hydrologists, like multiple regression and time series analysis. Then, the kriging technique is developed for the most usual cases, including nugget effects and the case of error in observed data. In the last part, some alternate interpretations are suggested, either by reformulating the method as dual kriging, or by coming back to more familiar fitting techniques like polynomial interpolation or spline-functions methods.

L'hydrologie présente, par rapport à d'autres domaines des sciences, certaines particularités qui jaillissent sur ses méthodes. Ce sont, parmi d'autres :

- la nécessité de fournir assez rapidement des réponses opérationnelles (dimensionnement d'ouvrages, prévisions diverses sur l'état futur du système, etc.)
- en ne possédant sur le système (le cycle de l'eau dans un bassin versant) que des données partielles

Ceci est vrai aussi bien dans le temps - les séries sont courtes par rapport aux temps de récurrence de certains phénomènes - que dans l'espace - les coûts limitent la couverture instrumentale en comparaison des surfaces à appréhender -

enfin, le système lui-même est souvent mal connu, et il est aussi illusoire d'envisager de l'étudier expérimentalement en laboratoire que de lui appliquer des lois déterministes, valables certes au niveau microscopique, mais qui ne renseignent guère sur le niveau macroscopique intéressant.

Cela est vrai par exemple pour une entrée privilégiée du système hydrologique : la précipitation pluvieuse. Ce n'est pas par des considérations hydro- et thermodynamiques que nous appréhenderons sa distribution sur la superficie d'un bassin, pas plus que par des expériences en laboratoire. Et les données recueillies au cours d'un épisode seront le seul point de départ objectif, même s'il apporte avec lui quelques sources d'erreur, pour appréhender l'entrée du système.

Ceci étant vrai pour d'autres parties du cycle de l'eau ou pour d'autres objectifs opérationnels, on comprend que l'Hydrologie ait été par le passé un utilisateur exigeant des statistiques et de l'analyse des données : Elle a même contribué parfois de manière significative à leur développement. De même les systèmes de prévision en temps réel ou l'enregistrement de modèles hydrologiques ont fait largement appel à des techniques d'identification issues de l'automatique : celles-ci furent d'abord assimilées, puis adaptées aux problèmes hydrologiques.

Il était donc normal qu'avec la disponibilité de plus en plus fréquente de réseaux suffisamment denses, ou d'informations télédéetectées, l'hydrologue s'intéresse aux techniques de traitement des phénomènes distribués dans l'espace et échantillonnés de manière plus ou moins adéquate. Un tel ensemble de techniques existait déjà dans les domaines voisins : en météorologie, en écologie forestière ou en géologie, plus particulièrement dans la partie orientée vers l'exploitation minière. Cette dernière communauté, dotée de moyens importants, sut reprendre et développer ces techniques jusqu'à les promouvoir sous le nom de Géostatistique.

On y retrouve des méthodes d'usage général, et dont les géologues ne peuvent revendiquer ni la primauté ni l'exclusivité, mélangées avec des solutions originales propres à des problèmes miniers. Malheureusement, cet ensemble assez formalisé a peu à peu développé sa propre terminologie. Il apparaît à certains comme une espèce d'école, à laquelle il faut faire acte d'allégeance, au point que se développent des polémiques retentissantes (cf. articles de Philip et Watson, 1986, et Srivastava, 1986, dans *Mathematical Geology*). Quant aux hydrologues, faute de reconnaître aisément dans la Géostatistique une continuité avec le savoir déjà acquis et les approches statistiques classiques, ils reculent devant un investissement intellectuel qui leur semble excessif, et se privent d'une approche qui, sans être révolutionnaire, est à coup sûr enrichissante.

Dans cet article, nous essaierons de présenter les éléments essentiels de la Géostatistique en partant de techniques couramment connues et pratiquées par les hydrologues de réseau, en particulier les techniques de corrélation linéaire ou de régression multiple. Nous tenterons donc essentiellement d'être didactique, fût-ce au risque d'être long, et le lecteur décidera de ces deux adjectifs, lequel lui semble le plus approprié...

PLAN

	Page
I Rappels sur le modèle linéaire, la régression multiple et les modèles autorégressifs.	178
I .1 La régression multiple	178
I .1.1 La théorie	178
I .1.2 La pratique de la régression	179
I .1.3 Les sous produits du modèle linéaire	180
I .2 Des exemples d'applications proches de la Géostatistique	182
I .2.1 Reconstitution de série à une station	182
I .2.2 Séries temporelles	183
I .2.3 Extensions dans le cas de séries temporelles	184
1.2.4 Premier pas dans l'espace	187
I .3 Conclusions préliminaires	189
II Le krigeage et ses outils	190
II.1 Les principes et les différentes conditions de non biais	190
II.1.1 Les principes	190
II.1.2 Les différentes conditions de non biais	190
II.2 Formulation du krigeage en variogramme et interprétation	194
II.2.1 Ecriture du système en variogramme	194
II.2.2 Le variogramme : son estimation	195
II.2.3 Le variogramme : son interprétation	199
II.2.4 Variogramme et expression des variances	202
II.2.5 Carte krigée et écart type d'estimation	203
II.3 Traitement des erreurs et effet de pépite	205
II.3.1 Effet de pépite	205
II.3.2 Données entachées d'erreur	209
II.3.3 Comparaison entre les deux approches précédentes	212

III Extension de la méthode	213
III.1 Krigeage d'une moyenne spatiale	213
III.2 Liaison avec les autres techniques d'interpolation	215
III.2.1 L'interpolation polynomiale	215
III.2.2 Le krigeage en interpolateur	216
III.2.3 Cas particuliers des fonctions splines	219
Conclusions	220
Remerciements	220
Bibliographie	221

I RAPPELS SUR LE MODELE LINEAIRE, LA REGRESSION MULTIPLE ET LES MODELES AUTO-REGRESSIFS

I.1 La régression multiple

I.1.1. La théorie

On dispose souvent en hydrologie d'une variable à expliquer Y, que l'on souhaite relier à des variables présumées explicatives X_1, \dots, X_n . La relation la plus simple à envisager, qui exprime directement une certaine proportionnalité entre l'effet et la cause présumée, est le modèle linéaire, du type :

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n + a_0 + \varepsilon \quad (1)$$

l'écart ε traduit l'inadéquation du modèle à la réalité.

Ayant choisi la forme du modèle, il faut l'ajuster aux propriétés des variables considérées. Pour cela on choisit un critère, parmi d'autres possibles.

On pourrait décider, par exemple, de déterminer les a_i de sorte que, en espérance, donc sur la population complète des Y, X_1, \dots, X_n , le carré de l'écart soit minimum, i.e. :

$$E[\varepsilon^2] = E[(Y - a_1 X_1 - a_2 X_2 - \dots - a_n X_n - a_0)^2] = \min \quad (2)$$

La minimisation s'effectue en dérivant successivement cette expression par rapport à chacun des paramètres a_1, \dots, a_n , et en écrivant que la dérivée est nulle quand la fonction est minimum. Si l'on considère ce dernier paramètre a_0 , on peut se demander la raison de son introduction. En fait, il vient équilibrer, dans l'équation (1), un éventuel décalage des moyennes entre les deux membres. On veut, en espérance, que :

$$E[Y] = a_1 E[X_1] + \dots + a_n E[X_n] + a_0 + E[\varepsilon] \quad (3)$$

Or, on souhaite qu'en espérance l'erreur soit nulle : $E[\varepsilon] = 0$. D'autre part les Y, X_1, \dots, X_n sont des valeurs quelconques, prises à partir d'une origine arbitraire pour chacune d'elles : il ne faudrait pas que le modèle -et l'optimum obtenu- dépende de ce choix arbitraire. D'où la nécessité d'un degré de liberté supplémentaire dans l'équation, soit a_0 , pour permettre l'égalisation des espérances. A défaut, une contrainte apparaîtrait sur les a_i , et réduirait la capacité de minimisation.

Dans les conditions retenues, l'équation (3) peut être soustraite de l'équation (1) avant de calculer $E[\varepsilon^2]$.

$$Y - E[Y] = a_1 (X_1 - E[X_1]) + \dots + a_n (X_n - E[X_n]) + \varepsilon \quad (4)$$

On se ramène alors à n inconnues a_1, \dots, a_n , la (n+1)^{ème} a_0 étant obtenue par l'équation (3) momentanément séparée des autres.

On montre alors que la minimisation conduit au système, (cf. pour des détails, l'ouvrage de J. Johnston, 1972) avec les notations :

variance de $X_i = \sigma_i^2$, covariance de X_i et $X_j = \text{cov}(X_i, X_j)$

$$\begin{bmatrix} \sigma^2_{X_1} & & & & \\ & \sigma^2_{X_1} & \text{cov}(X_1, X_2) & & \\ & \text{cov}(X_2, X_1) & \sigma^2_{X_2} & & \\ & & & \ddots & \\ & & & & \sigma^2_{X_n} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_j \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \text{cov}(Y, X_1) \\ \text{cov}(Y, X_2) \\ \text{cov}(Y, X_j) \\ \vdots \\ \text{cov}(Y, X_n) \end{bmatrix} \quad (5)$$

ou encore

$$R_{XX} \cdot A = R_{YX}$$

où R_{XX} représente la matrice des covariances des X_i, X_j entre eux, et R_{YX} le vecteur des covariances entre Y et les X_i . Mais qui globalement s'écrit en fait :

$$\begin{bmatrix} R_{XX} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ E[X_1] \dots E[X_n] & 1 \end{bmatrix} \begin{bmatrix} A \\ \vdots \\ a_0 \end{bmatrix} = \begin{bmatrix} R_{YX} \\ \vdots \\ E[y] \end{bmatrix} \quad (6)$$

I 1.2. La pratique de la régression :

En pratique, on va procéder un peu différemment, en ce sens que l'on dispose généralement d'un échantillon ou tableau de données, comportant par exemple P observations conjointes de Y et $X_1 \dots X_n$, soit :

$$\begin{bmatrix} y_1 \\ \vdots \\ y_k \\ \vdots \\ y_p \end{bmatrix} \quad \begin{bmatrix} x_{11} & \dots & x_{11} & x_{n1} \\ \vdots & & \vdots & \vdots \\ x_{1k} & x_{1k} & x_{nk} & \\ \vdots & & \vdots & \vdots \\ x_{1p} & x_{1p} & x_{np} & \end{bmatrix} \quad \text{où } k \text{ est l'observation courante} \quad (7)$$

$Y(p \times 1) \qquad X(p \times n)$

La pratique habituelle consiste à remplacer, dans les équations (1) à (5), l'opérateur espérance mathématique $E[]$ par son approximation : la moyenne arithmétique sur un échantillon. C'est ainsi que (1) \rightarrow (1'), pour la kème observation :

$$Y_k = a_1 X_{1k} + \dots + a_n X_{nk} + e_k \quad (1')$$

mais surtout, les conditions :

$$E[e] = 0 \rightarrow \sum_{k=1}^P e_k = 0$$

$$E[e^2] = \text{Min} \rightarrow \sum_{k=1}^P (Y_k - a_1 X_{1k} - \dots - a_n X_{nk} - a_0)^2 = \text{Min}$$

Comme on le sait, et pour des raisons d'échantillonnage :

$$\frac{1}{P} \sum_{k=1}^P X_{ik} = \bar{X}_i \quad \text{diffère de } E[X_i]$$

$$\frac{1}{P} \sum_{k=1}^P (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j) \quad \text{diffère de } \text{cov}(X_i, X_j) \quad (8)$$

En conséquence, les coefficients obtenus a_0, a_1, \dots, a_n diffèrent des coefficients "théoriques" de la population a_0, a_1, \dots, a_n .

Et des développements théoriques nombreux ont été faits pour tester l'amplitude des fluctuations de ces coefficients en fonction de l'échantillonnage.

On notera aussi, en pratique, que si on centre les données des tableaux (7), Y et X, à l'aide des moyennes empiriques calculées sur l'échantillon alors la matrice des covariances empiriques se résume au produit matriciel.

$$\frac{1}{P} X^T X$$

et si $P > n$, alors cette matrice symétrique est en général définie positive et donc inversible, et le système est soluble. Ce n'est plus vrai si le nombre d'observations $P < n$, bien qu'en espérance, le modèle linéaire soit pertinent...

I 1.3 Les "sous-produits" du modèle linéaire.

Parmi ceux-ci, on retiendra : (cf. par exemple cours de D. Duband, 1982)

- le coefficient de corrélation multiple :

$$R_{Y|X_1, \dots, X_n}^2$$

qui exprime le pourcentage de σ^2_Y expliqué par la connaissance de X_1, \dots, X_n

- l'écart type résiduel, c'est-à-dire l'écart type, optimisé par la méthode, de v :

$$\begin{aligned} \sigma_v^2 &= \sigma^2(Y - Y^*) = \sigma_Y^2(1 - R_{Y|X_1 \dots X_n}^2) \\ &= \sigma_Y^2 - \sum_i a_i \cdot \text{cov}(Y, X_i) \end{aligned} \quad (9)$$

en notant $Y^* = a_1 X_1 + \dots + a_n X_n + a_0$ la valeur modélisée.

On les donne ici en valeurs théoriques, sur la population, alors qu'en pratique ils sont calculés sur l'échantillon et affectés d'une incertitude d'échantillonnage correspondante.

D'autres résultats viennent étayer ceux-ci, en particulier la notion de corrélation partielle. Celle-ci exprime l'apport d'une variable X_1 quand par exemple X_1, X_2, X_3, \dots ont déjà été utilisées et surtout quant les variables X_1, X_2, X_3, \dots sont intercorrélées.

Dans le cas simple de trois variables, on cherche à déterminer le modèle $Y = a_1 X_1 + a_2 X_2 + a_0 + v$ où Y et X_1 sont corrélées, mais aussi Y et X_2 , ce qui pose la question d'insérer ou non X_2 car il existe une corrélation "parasite" entre X_1 et X_2 , qui peut être forte.

Partant du modèle :

$$Y = \beta_1 X_1 + \beta_0 + \varepsilon_{Y|X_1}$$

on voit que la part de Y non expliquée par X_1 est $\varepsilon_{Y|X_1}$:

Par contre, X_2 est en partie aussi expliquée par X_1 :

$$X_2 = \gamma_1 X_1 + \gamma_0 + \varepsilon_{X_2|X_1}$$

et cette information, due à X_1 , déjà prise en compte dans Y , doit être retirée (pour mesurer le véritable apport de X_2), ce qui est le cas dans $\varepsilon_{X_2|X_1}$

En dépit de la corrélation totale $\text{cov}(Y, X_2)$, il n'y aura donc apport de la variable X_2 à un modèle contenant déjà X_1 que si la corrélation totale des résidus :

$$r(\varepsilon_{Y|X_1}, \varepsilon_{X_2|X_1}) \quad \text{notée plus simplement} \quad r_{Y, X_2|X_1}$$

est différente de 0, c'est à dire si $\varepsilon_{X_2|X_1}$ explique une partie du résidu $\varepsilon_{Y|X_1}$.

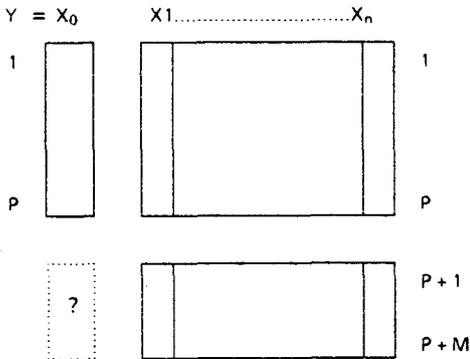
Cette corrélation totale entre résidus s'appelle corrélation partielle entre Y et X_2 quand l'effet de X_1 a déjà été pris en compte. Et ce n'est que si cette corrélation partielle est importante que l'introduction de X_2 améliore significativement le modèle. Ceci est entre autres utilisé dans les techniques de sélection de variables.

Nous retrouverons cet effet en Géostatistique, sous le nom d'"effet d'écran".

1.2 Des exemples d'applications ou de problèmes proches de la Géostatistique :

1.2.1: Reconstitution de série à une station.

Dans un réseau de $n + 1$ stations X_0, X_1, \dots, X_n , la plupart des stations ont fonctionné $P + M$ années, à l'exception de la station X_0 arrêtée au bout de P années. Il est alors intéressant de chercher à compléter la série en X_0 sur les M années manquantes.



On pense donc généralement à caler un modèle où X_0 joue le rôle d' Y , et tel que la valeur reconstituée soit :

$$X_0^* = a_1 X_1 + \dots + a_n X_n + a_0$$

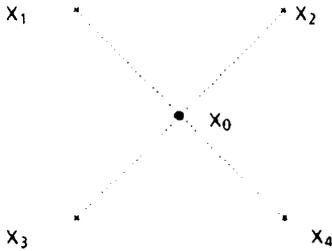
L'échantillon P étant supposé fourni (en particulier $P \gg M$), on peut estimer les moyennes, variances et covariances :

$$\bar{X}_i, \sigma_{X_i}, \text{cov}(X_i, X_j) \quad \forall i, j = 0, \dots, n$$

On applique alors le modèle pour chacune des observations $k = P + 1, \dots, P + M$. On notera au passage que les valeurs reconstituées, ou estimées X_0^* ont même moyenne, mais une variance un peu inférieure à $\sigma_{X_0}^2$ puisque la vraie variance s'écrit :

$$\sigma_{X_0}^2 = \sigma_{X_0}^2 + \sigma_{\epsilon_{X_0|X_1 \dots X_n}}^2 \quad (13)$$

Néanmoins, pour chaque observation, i.e pour chaque ensemble de valeurs mesurées dans l'espace aux points $1, \dots, n$, on a reconstitué une valeur au point 0, la meilleure valeur possible connaissant les valeurs en $1, \dots, n$.



On aura ainsi noté au passage qu'il n'est peut-être pas utile, pour expliquer X_0 , d'utiliser les n stations : Pour des raisons de corrélation partielle, un sous ensemble de stations proches suffit généralement.

1.2.2. Les séries temporelles

Cette fois la variable Y à expliquer est la valeur d'une variable ou processus au temps t à partir des valeurs aux temps $t-1, t-2, \dots, t-n$. Le modèle est alors :

$$Y = X_t = a_1 X_{t-1} + \dots + a_n X_{t-n} + a_0 + \varepsilon \quad (14)$$

Ici une particularité est qu'il s'agit partout de la même variable X mais décalée dans le temps.

On peut imaginer que, par exemple, on veut prévoir les débits d'étiage de septembre à partir des mois de juin à août. Pour cela, on dispose de plusieurs "observations" : les années antérieures, supposées indépendantes.

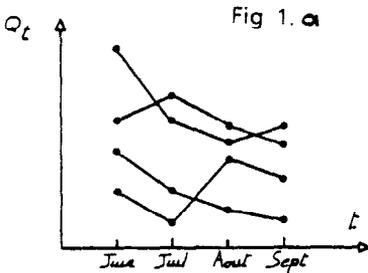


Fig 1. a

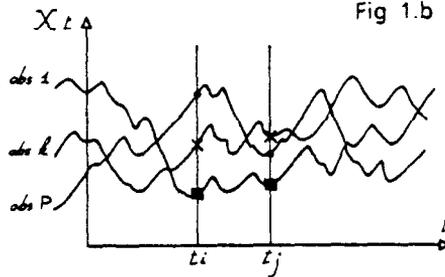


Fig 1. b

Dans ce cas on peut aisément calculer sur les données $cov(X_{t,i}, X_{t,m}) \forall i, m = 0, \dots, 3$ et résoudre le système. Plus généralement, ce modèle peut s'appliquer à un processus temporel, dont on connaît plusieurs observations, encore appelées "réalisations" $k = 1, \dots, P$. La corrélation entre deux intervalles de temps t , et t_j peut se calculer à l'aide des $k = 1, \dots, P$ couples disponibles pour remplir le système.

I 2.3. Extensions dans le cas des séries temporelles.

Si on peut imaginer d'observer plusieurs fois la crue nivale de printemps de la Durance, ou la décrue estivale de la Loire à Blois, il est des cas où l'on ne dispose que d'une observation, éventuellement longue. Prenons par exemple le débit annuel du Niger à Koulikouro (cf. Salas *et al*, 1978) disponible entre 1906 et 1980.

On cherche un modèle "autorégressif" du type :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + a_0 + \varepsilon \quad (15)$$

On remarquera qu'ici, si on suppose la série "homogène" dans le temps, ou encore "stationnaire", alors les moyennes et variances sont identiques pour $t, t-1, \dots$

$$E[X_t] = E[X_{t-1}] = \dots = \mu_x$$

$$\sigma_x = \sigma_{x_{t-1}} = \dots = \sigma_x$$

Cette stationnarité n'étant qu'une hypothèse, mais qui va permettre d'écrire que :

$$\text{cov}(X_t, X_{t-1}) = C_t \quad \forall t$$

D'autre part, si $E[X_t] = \mu_x$, on a, en prenant l'espérance de (15) :

$$\mu_x = a_1 \mu_x + a_2 \mu_x + a_0$$

d'où : $a_0 = (1 - a_1 - a_2) \mu_x$, ce qui définit a_0

Si au contraire on impose un modèle moins libre,

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon \quad (16)$$

alors on a une contrainte sur a_1 et a_2 car la condition sur les espérances devient : $1 = a_1 + a_2$ et le modèle est moins optimisé. En général, soit on introduit a_0 , soit on travaille en valeurs centrées, c'est à dire $\mu_x = 0$ d'où $a_0 = 0$. Cela suppose que l'on connaisse bien la moyenne, ce qui est relativement courant en temporel, quand la série est "longue" par rapport à la durée sur laquelle la corrélation entre deux valeurs est non nulle (temps de décorrélation).

Un autre problème, dans le cas d'une seule réalisation, provient du calcul ou plutôt de l'estimation par exemple de $C_t = \text{cov}(X_t, X_{t-1})$. On fait alors l'hypothèse ergodique. Celle-ci exprime que :

- si l'on dispose d'une longue réalisation,
- et si la corrélation s'éteint au-delà d'un certain décalage de L pas,
- alors un tronçon donné, décalé du tronçon précédent de plus de L pas, n'a pratiquement plus rien à voir avec lui. Ils sont "indépendants".

On remplacera alors l'Espérance, ou moyenne d'ensemble (sur l'ensemble des réalisations) par une moyenne dans la réalisation observée (fig. 2).

M tronçons d'une réalisation :

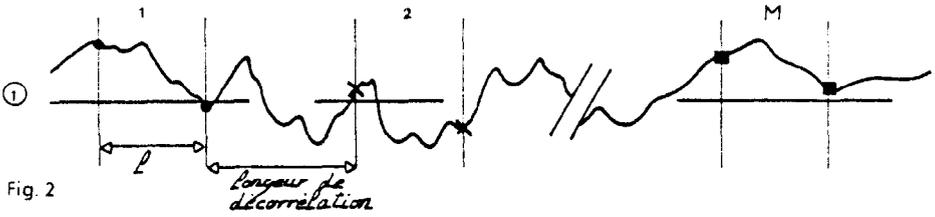
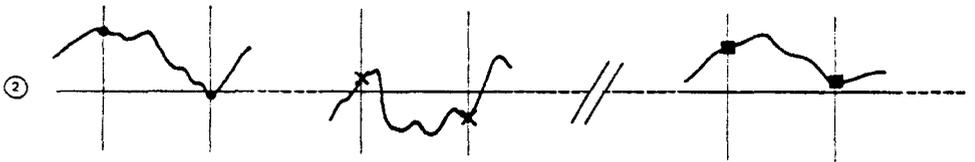
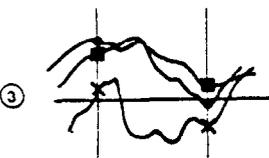


Fig. 2

M tronçons de M réalisations :



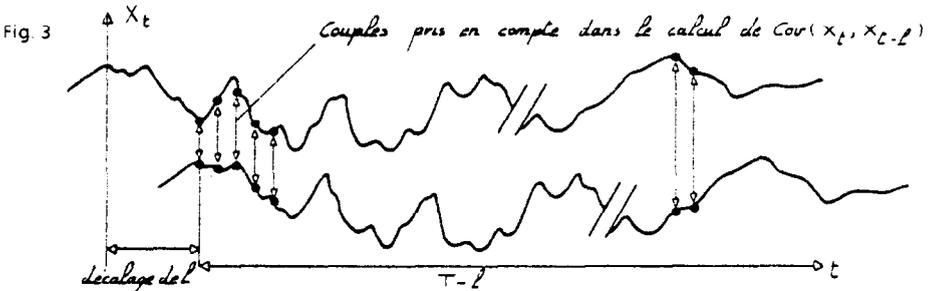
M tronçons de M réalisations :



En fait on considère que le deuxième tronçon 2 de ① pourrait tout aussi bien provenir d'une autre réalisation indépendante ②. De même pour les autres. On peut alors "superposer" ces M tronçons "indépendants" et calculer sur ces M couples en ③ la covariance au décalage l.

De manière équivalente, cela reviendrait à prendre la réalisation, à la décaler de l pas et à sélectionner M couples indépendants, c'est à dire suffisamment éloignés les uns des autres dans la réalisation.

En pratique, on prendra tous les couples disponibles, soit T-l



En sachant bien qu'ils ne sont plus indépendants et que le nombre "équivalent" de couples indépendants est $\ll T$. On calcule alors une valeur approchée de la covariance, par exemple :

$$Cov(X_t, X_{t-l}) = \frac{1}{T-l} \sum_{t=1}^T (X_t - \bar{X}_t) (X_{t-l} - \bar{X}_{t-l}) = C_l \quad (17)$$

Il existe différentes expressions de cet estimateur (Cf. Salas et Delleur, 1978) mais on sait que si $T \rightarrow \infty$ elles convergent vers la vraie valeur de C_l .

On a donc fait l'inférence de la covariance dans une seule réalisation, en admettant qu'elle est suffisamment longue et représente bien l'ensemble de la population. Sans être trop rigoureux mathématiquement, on peut dire que l'on a ainsi supposé le processus ergodique.

D'autre part, les valeurs empiriques ainsi calculées pour C_l , $l = 1, \dots, L$, sont entachées d'effets d'échantillonnage (en particulier, quand le décalage devient grand, il y a de moins en moins de couples communs, d'où les oscillations inopportunes). On est donc conduit à lisser ces valeurs empiriques à l'aide de fonctions adéquates constituant des modèles acceptables de corrélogramme. En particulier, pour certains processus "simples" (autorégressifs par exemple) on connaît même la forme analytique du corrélogramme théorique. On en ajustera donc un aux valeurs empiriques,

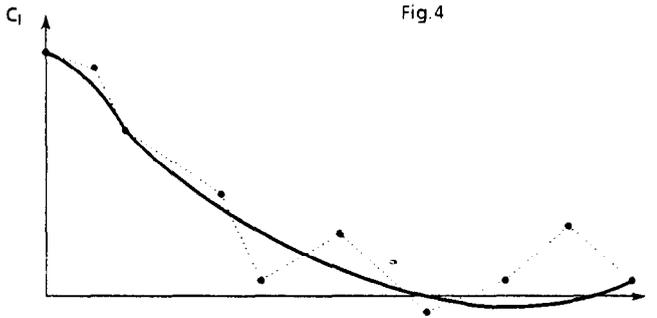


Fig. 4

et c'est avec les valeurs lissées que l'on calculera, grâce au système (5) ou (6), les coefficients de l'équation (14).

Néanmoins, si ces hypothèses plus ou moins vérifiées permettent de proposer un modèle, même en ne disposant que d'une chronique, elles apportent avec elles quelques problèmes annexes... Par exemple, dans la formule (17), on a besoin pour centrer les données d'une estimation de la moyenne $E[X_t]$.

On comprend aisément, sur la figure 5a que même si le processus est stationnaire de moyenne nulle, une chronique trop courte donnera une estimation erronée de la moyenne.

Dans le cas, pas forcément invraisemblable, où le processus n'est stationnaire que par morceaux (changements de lois de tarage dans une chronique de débits par exemple) on voit aussi Fig. 5.b que la moyenne d'ensemble n'est qu'un artefact, qui jouera cependant un rôle essentiel dans le calcul de C_l : c'est par rapport à cette moyenne que seront calculés les écarts et leurs produits croisés. C'est ainsi qu'un signal complètement corrélé (bruit blanc) par morceaux, apparaîtra globalement corrélé !

I.3. Conclusions de la première partie :

On a repris dans cette partie la notion d'estimation fournie par un modèle linéaire et ajustée par des techniques de moindres carrés. Celles-ci peuvent s'appliquer sur un échantillon empirique d'observations, auquel cas on minimise la somme des carrés des erreurs obtenues en appliquant le modèle à l'échantillon, ou bien en espérance mathématique auquel cas on minimise une variance résiduelle théorique, celle du terme ϵ introduit dans le modèle.

On a vu comment, dans le domaine temporel, cette technique s'appliquait à un processus aléatoire échantillonné. Et on a pu constater les différents contextes pratiques que l'on rencontre :

- plusieurs observations disponibles \Rightarrow contexte multiréalisations. Comme ces observations correspondent en pratique à des instants différents du passé, on parle de contexte "climatologique"
- une seule observation disponible \Rightarrow contexte monoréalisation. Il faut alors inférer toute la structure statistique du processus sur cette seule réalisation. Pour cela, on la suppose stationnaire, afin de pouvoir la "replier" sur elle-même.

On a vu que, dans ce dernier cas surtout, on lisait les covariances empiriques à l'aide d'un modèle de covariance, ce qui revenait à effectuer l'ajustement du modèle du processus en espérance mathématique.

Complémentairement, on a fait quelques incursions dans l'espace à deux dimensions. On considère alors les observations comme des réalisations d'un champ aléatoire échantillonné. Si les propriétés statistiques sont identiques en tout point du champ, on dit qu'il est stationnaire, ou encore homogène dans l'espace. On a vu que l'inférence pouvait se faire dans un contexte multiréalisations, mais plus difficilement dans un contexte monoréalisation.

Or ce cas est assez courant en hydrologie (par exemple, le champ des transmissivités hydrauliques en hydrogéologie - Delhomme, 1976). Et même quand on dispose de plusieurs observations d'un champ, on n'est pas toujours sûr qu'elles soient comparables, c'est à dire correspondent à des tirages dans la même population.

Enfin, à l'inverse du cas temporel où l'échantillonnage se fait en général à pas constant, on trouve souvent dans l'espace des réseaux assez irréguliers, implantés plutôt sur des critères d'accessibilité ou de disponibilité de site. D'où la nécessité de développer ce cas.

Et ce faisant, sans y faire encore référence, on a pratiquement manipulé tous les ingrédients de la Géostatistique. L'utilisateur potentiel la perçoit souvent comme une nouvelle technique qu'il doit appréhender à partir de rien. Mais il en connaît déjà, sous d'autres dénominations, les principaux outils...

II LE KRIGEAGE ET SES OUTILS :

II.1. Les principes et les différentes formulations des conditions de non biais

II.1.1. Les principes :

La méthode cherche à estimer de manière optimale un processus $Z(t)$ au point t^0 . On souhaite un estimateur basé sur les valeurs brutes du processus, que l'on appellera plutôt $Z(t^i)$

En effet, l'utilisation de données centrées $X(t^i)$ pose parfois un délicat problème, sur des séries courtes ou des champs peu étendus, d'estimation de la moyenne.

t^i représente l'abscisse le long d'une ligne, ou les coordonnées x_i, y_i , dans le plan, etc.

Donc on cherche un modèle de la forme :

$$Z^*(t^0) = \sum \lambda_i Z(t^i) \quad (20)$$

en valeurs brutes et sans terme constant. On le souhaite non biaisé, c'est à dire que l'espérance de l'estimateur $E\{Z^*(t^0)\}$ est identique à celle de la valeur vraie $E\{Z(t^0)\}$

$$E\{Z(t^0) - Z^*(t^0)\} = 0 \quad (21)$$

et optimal au sens de la variance du résidu :

$$E\{(Z(t^0) - Z^*(t^0))^2\} = \text{Min} \quad (22)$$

II 1.2. Les différentes conditions de non biais :

a) Phénomène de moyenne nulle : On suppose connu que $E\{Z(t)\} = 0 \quad \forall t$

Dans ce cas, la condition de non biais (21) s'écrit :

$$E\{Z^*(t^0)\} = E\left\{\sum \lambda_i Z(t^i)\right\} = \sum \lambda_i E\{Z(t^i)\} = 0 \text{ donc } = E\{Z(t^0)\}$$

et elle est toujours vérifiée.

La condition d'optimalité (22) s'obtient en dérivant par rapport aux paramètres inconnus λ_i :

$$\frac{\delta}{\delta \lambda_i} E\{(Z(t^0) - \sum \lambda_i Z(t^i))^2\} = 0 \quad i = 1 \dots n$$

d'où, avec $E\{Z(t^i) \cdot Z(t^j)\} = \text{cov}(X_i, X_j)$ le système linéaire (5'), tout à fait similaire à (5)

$$(5) \quad \begin{bmatrix} C(t^1, t^1) \\ C(t, t) \quad C(t, t) \\ C(t, t) \quad C(t, t) \\ C(t^n, t^n) \end{bmatrix} \begin{bmatrix} \lambda_1(t^0) \\ \vdots \\ \lambda_i(t^0) \\ \vdots \\ \lambda_j(t^0) \\ \vdots \\ \lambda_n(t^0) \end{bmatrix} = \begin{bmatrix} C(t^1, t^0) \\ C(t, t^0) \\ C(t, t^0) \\ C(t^n, t^0) \end{bmatrix}$$

On est donc strictement dans le cas du modèle linéaire de régression multiple, mais écrit en espérance sur la population des champs possibles.

En particulier, la variance de l'erreur d'estimation, ou variance résiduelle s'écrit comme en (9) :

$$E\{(Z(t^0) - Z^*(t^0))^2\}_{Z(t^0), Z(t^1), \dots, Z(t^n)} = \sigma^2 \quad (23)$$

$$= C(t^0, t^0) - \sum \lambda_i(t^0) \cdot C(t^i, t^0)$$

b) Phénomène de moyenne constante mais inconnue : $E[Z^*(t)] = m \quad \forall t$

La condition (21) s'écrit alors :

$$E\{Z^*(t^0)\} = \sum \lambda_i E\{Z(t^i)\} = \sum \lambda_i \cdot m \quad (24)$$

et cela ne sera égal à $E\{Z(t^0)\} = m$ que si $\sum \lambda_i = 1$ (25)

Il faudra donc optimiser (22) en respectant la contrainte (25), et on sait que cela revient à minimiser avec un paramètre supplémentaire v :

$$E\{(Z(t^0) - \sum \lambda_i Z(t^i))^2\} + 2v(\sum \lambda_i - 1) = \text{Min}_{\lambda_i, v} \quad (26)$$

On dérive donc par rapport aux λ_i et v (appelé multiplicateur de Lagrange).

La dérivation est évidente mais pour faire réapparaître les covariances, car les données ne sont plus centrées, on réinjecte dans l'équation (26) la relation (24) réécrite :

$$E\{Z(t^0)\} - \sum \lambda_i E\{Z(t^i)\} = 0$$

d'où :

$$E\{(Z(t^0) - E\{Z(t^0)\}) - \sum \lambda_i (Z(t^i) - E\{Z(t^i)\})\}^2 + 2v \left(\sum \lambda_i - 1 \right)$$

et la dérivation conduit à un système de dimension $n + 1$:

La mise en évidence des covariances nécessite, comme en (b) la réinjection des équations (31) et (24) dans l'expression (26) avant dérivation.

On notera que pour les trois cas ci-dessus (a, b et c) on suppose que le processus centré, c'est à dire $X(t) = Z(t) - m(t)$ possède une covariance, ou encore que la covariance $E[X(t).X(t')] = C(t, t')$ existe et reste bornée $\forall t, t'$.

On peut de plus supposer que le processus est homogène c'est à dire $C(t, t') = C(|t-t'|)$ mais ce n'est pas nécessaire pour écrire le système.

Un exemple courant du cas c) consiste à choisir pour $f(t)$ une base polynômiale et à supposer la moyenne localement linéaire : c'est à dire que la tendance locale, on dit encore la dérivée, dans le voisinage des points t_1, \dots, t_n , est à peu près approchée par une portion de plan.

Dans ce cas on a :

$$\left\{ \begin{array}{l} f^0(t) = 1 \quad f^1(t) = f^1(x, y) = x \quad f^2(t) = f^2(x, y) = y \\ \text{et } m(t) = a^0 + a^1 x + a^2 y \end{array} \right. \quad (33)$$

II.2 Formulation du krigeage en variogramme et interprétation du variogramme

II.2.1. Ecriture du système en variogramme :

Dans le paragraphe précédent, II.1, en particulier dans II.1 b) et c), on a vu comment, dans un estimateur (20) basé sur les valeurs brutes, on arrivait à éliminer l'estimation de la moyenne m inconnue, au prix d'une ou plusieurs contraintes.

Néanmoins, on arrivait au système (28) qui utilise les covariances :

$$C(t, t') = E\{(Z(t) - m).(Z(t') - m)\}$$

lesquelles nécessitent... l'estimation de la moyenne m !

En fait ceci n'est pas nécessaire si on utilise un autre outil que la covariance, à savoir le variogramme, défini par :

$$\gamma(t, t') = \frac{1}{2} E\{(Z(t) - Z(t'))^2\} \quad (34)$$

Si le processus est homogène, on peut tout de suite écrire :

$$\gamma(t, t' + d) = \gamma(d) = \frac{1}{2} E\{(Z(t) - Z(t + d))^2\} \quad \text{où } d \text{ est l'interdistance entre les deux points } t \text{ et } t'$$

cela donne, en fonction des covariances :

$$\gamma(d) = C(0) - C(d) \quad (36)$$

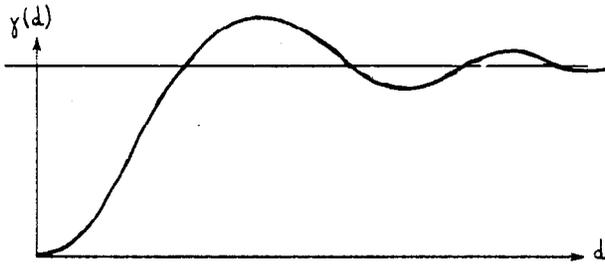
Et si l'on introduit cela dans le système de krigeage en covariance (28) ; soit

$$Y_{ij} = C(0) - C_{ij}$$

On vérifie qu'il se décompose en une partie triviale contenant les $C(0)$, et en un système :

Il a pour avantage de faire ressortir explicitement une "portée" (distance au-delà de laquelle la corrélation devient nulle). Or il est tout à fait concevable d'imaginer des variogrammes oscillant de façon amortie.

Fig. 11



cas fréquent en hydrométéorologie, par exemple de la forme

$$\gamma(d) = 1 - e^{-\alpha d} \cdot \cos \beta d$$

Le calage de ces modèles, à l'aide de leurs paramètres α , β ..., est fait soit à l'oeil, soit par une procédure de moindres carrés assez élémentaire. Ensuite, le modèle calibré obtenu est généralement utilisé sans plus tenir compte de la procédure d'ajustement, ni des particularités de l'échantillon qui a servi à l'établir.

Or, rappelons que les classes de distances :

- agrègent des distances d_i parfois assez différentes (en valeur relative)
exemple : classe des couples compris entre 3 et 5 km, puis entre 5 et 10 km...
- dont la moyenne n'est pas forcément au centre de la classe
- dont le nombre de couples varie significativement d'une classe à l'autre, créant des fluctuations d'échantillonnage.
- et que l'estimateur : "moyenne arithmétique des carrés des écarts" oublie la distribution parfois très dissymétrique de ces écarts

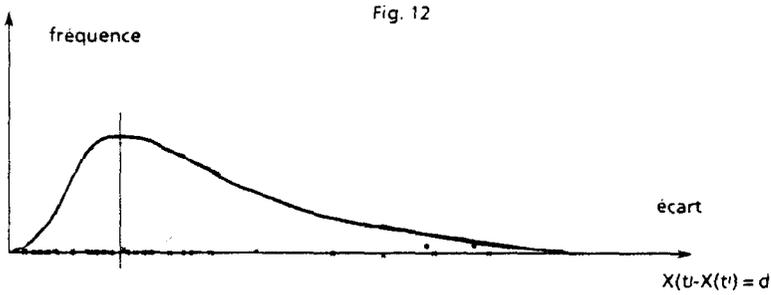
Un point du variogramme empirique (Figure 9) correspond à la valeur moyenne :

$$\frac{1}{Nd} \sum |X(t^i) - X(t^i + d)|^2$$

Mais la distribution des Nd écarts individuels peut être très dissymétrique, et celle des carrés plus encore. (Figure 12).

Une valeur anormale (erronée ?) peut alors créer un écart anormal dans tous les couples de toutes les classes d'interdistances où elle intervient.

L'analyse des variogrammes empiriques requiert donc un certain soin et un grand savoir-faire expérimental.



Plus récemment, on a préconisé un calage automatique des variogrammes. Pour cela, on prend une forme paramétrée, sphérique ou puissance par exemple :

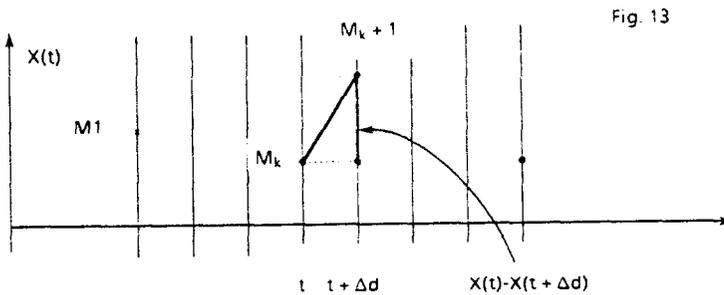
$$\gamma(d_{ij}) = \alpha \cdot d_{ij}^\beta \quad (41)$$

On considère alors chaque station tour à tour, on l'occulte momentanément et on la reconstitue à l'aide de ses voisines. On optimise la reconstitution globale à l'aide des paramètres α et β , optimaux sur l'échantillon disponible. Il n'est pas possible ici d'entrer dans les détails de ces techniques (cf. Lebel et Bastin, 1985). On ne développera pas non plus les hypothèses intrinsèques d'ordre 1 ou plus (F.A. 1...k), où le processus peut être de variance infinie, car cela est assez rarement nécessaire en hydrologie de surface

II.2.3. Le variogramme : son interprétation

a) Où l'on retrouve Pythagore :

Une première interprétation consiste à se placer à une dimension (par exemple en temporel) avec un échantillonnage à pas constant Δd



Dans ce cas, on voit que pour le pas Δd (pour un multiple de Δd , le raisonnement est identique) :

$$\gamma(\Delta d) = \frac{1}{2N} \sum_{k=1}^N |X(t+k\Delta d) - X(t+(k+1)\Delta d)|^2$$

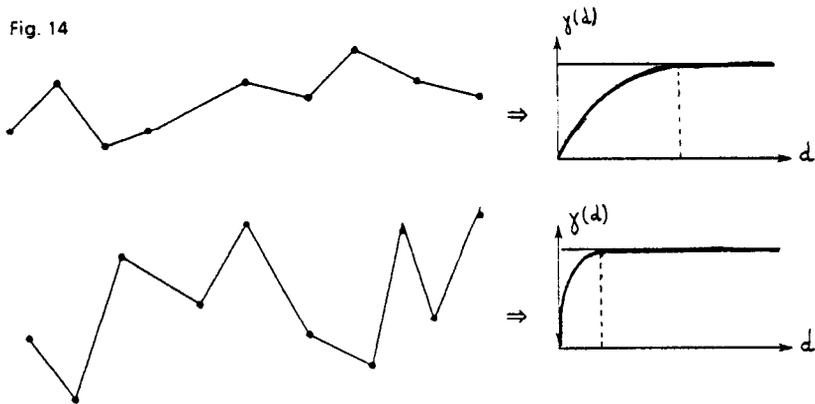
Or, on voit que, si on considère un écart $X(t+k\Delta d) - X(t+(k+1)\Delta d)$, qu'on l'élève au carré, et qu'on lui ajoute Δd^2 , on obtient, selon le théorème de Pythagore, le carré de la longueur du segment $M_k M_{k+1}$.

Donc, à $N \cdot \Delta d^2$ près, l'expression $\gamma(\Delta d)$ est la somme des carrés des segments $M_k M_{k+1}$, donc indirectement une fonction de la longueur de la ligne brisée complète $M_1 M_2 \dots M_N$.

Et si le processus est "lisse", cette longueur s'écarte peu de $N \cdot \Delta d$, et la quantité $\gamma(\Delta d)$ est faible. Au contraire, si le processus est chaotique, elle augmente très fortement. On trouvera donc :

- pour les processus "lisses", très autocorrélés positivement, un variogramme $\gamma(d)$ qui monte lentement quand d augmente.
- et au contraire un variogramme qui monte très vite quand le processus est peu autocorrélé et se rapproche d'un bruit blanc.

Fig. 14



On notera, en poursuivant cette interprétation sur les processus temporels, qu'elle vaut surtout pour une autocorrélation positive (c'est plus compliqué, par exemple, pour un processus markovien $X_t = -0.8X_{t-1} + \epsilon_t$ négativement corrélé, que l'on pourrait croire chaotique). Mais en Géostatique, on ne considère souvent que les corrélations positives.

b) un drôle de pistolet

Une autre interprétation du variogramme consiste à se rappeler la relation (36) entre variogramme et covariance

$$\gamma(d) = C(0) - C(d) = \text{var}[X_t] - \text{cov}[X_t, X_{t+d}]$$

Dans le cas où la covariance est normée, c'est à dire où l'on travaille en corrélation, $\text{var}[X_t] = 1$ et $\gamma(d)$ est le complément à 1 du corrélogramme.

II 2.4. Variogramme et expression des variances :

a) si on suppose connu le processus $X(t)$ (centré pour simplifier) alors toute combinaison linéaire

$$Y = \sum_i \lambda_i X(t^i) \quad (43)$$

a pour variance :

$$\begin{aligned} E[Y^2] &= E\left[\sum_i \lambda_i X(t^i)\right]^2 = \sum_i \sum_j \lambda_i \lambda_j \text{cov}(X(t^i)X(t^j)) \\ &= \sum_i \lambda_i \lambda_j (\gamma(d_{ij}) - C(0)) \end{aligned} \quad (44)$$

deux cas peuvent se présenter :

- la covariance existe (le processus est stationnaire)
- elle n'existe pas mais on se limite à des combinaisons linéaires telles que :

$$\sum_i \lambda_i = 0 \Rightarrow \sum_i \lambda_i C(0) = 0 \quad (45)$$

et dans ce cas :

$$\text{var}[Y] = \sum_i \sum_j \lambda_i \lambda_j \gamma_{ij} \quad (46)$$

On notera au passage que $\text{var}[Y]$ est forcément positive ou nulle. Alors $\gamma(d)$ doit être telle que $\forall d_{ij} \lambda_i \lambda_j$ les valeurs $\gamma(d_{ij}) = \gamma_{ij}$ dans 46 donnent un résultat positif ou nul. Cela réduit $\gamma(d)$ à une certaine classe de fonctions.

b) Cas particulier : la variance d'estimation du krigage.

L'écart d'estimation s'écrit :

$$Z(t^0) - Z^*(t^0) = Z(t^0) - \sum_i \lambda_i Z(t^i)$$

$$\text{avec } \sum_i \lambda_i = 1$$

c'est à dire que la variable "écart" est bien centrée (d'espérance nulle).

Si on pose $\lambda_0 = 1$, cela donne la combinaison linéaire :

$$\begin{aligned} Y &= Z(t^0) - \sum_{i=1}^n \lambda_i Z(t^i) = \lambda_0 Z(t^0) - \sum_{i=1}^n \lambda_i Z(t^i) \\ &= \sum_{j=0}^n \lambda_j Z(t^j) \quad \text{avec } \sum_{j=0}^n \lambda_j = 0 \end{aligned}$$

et l'on se retrouve dans le cas du a) ci-dessus. D'où :

$$\text{var} Y_i = \sum_{j=0}^n \sum_{l=0}^n \lambda_i \lambda_j Y_{ij}$$

Mais comme $Y_{00} = Y_{ii} = Y(0) = 0 \forall i$ et que les valeurs de λ_i satisfont le système (37), on retrouve bien alors l'expression (38) donnée sans démonstration :

$$\text{var} \left[Z(t^0) - Z^*(t^0) \right] = \sigma_k^2(t^0) = \sum \lambda_i Y_{i0} + v^i$$

II.2.5. Carte krigée et écart type d'estimation :

Naturellement, cette valeur de $\sigma_k(t^0)$ est fonction de la position de t^0 par rapport aux points mesurés t^i , par $Y_{i0} = Y(d_{i0}) = Y(t^i, t^0)$, ce qui conduit aux valeurs $\lambda_i(t^0)$ et $v(t^0)$.

En particulier, quand on s'éloigne des points t^i au point que la covariance, ou la corrélation, devienne nulle alors

$$\sigma_k^2(t^0) \rightarrow \text{var} Z$$

c'est à dire la variance libre du phénomène.

D'autre part, il faut bien voir que les valeurs krigées expriment l'espérance du processus, et donc sa moyenne en t^0 sur l'ensemble des réalisations qui passent par les valeurs $Z(t^1)$, ou encore l'espérance en t^0 , conditionnée par ces seules informations disponibles : les valeurs $Z(t^i)$ soit $E[Z(t^0)|Z(t^1), Z(t^2), \dots, Z(t^n)]$

Cette courbe est donc anormalement "lisse" par rapport à une quelconque des réalisations possibles qui passe par les valeurs observées $Z(t^1), \dots, Z(t^n)$. En fait, la "vraie" réalisation est probablement plus chaotique que la ligne "krigée". Elle oscille autour de celle-ci "dans" la zone de confiance définie par l'écart type d'estimation.

On voit plusieurs configurations de cela sur la figure 17. Un exemple concret de simulations conditionnelles extrait de J.P. Delhomme (1976) montre différentes réalisations possibles ayant les mêmes valeurs aux point t^i donc la même surface krigée. (Fig. 18).

Fig. 17

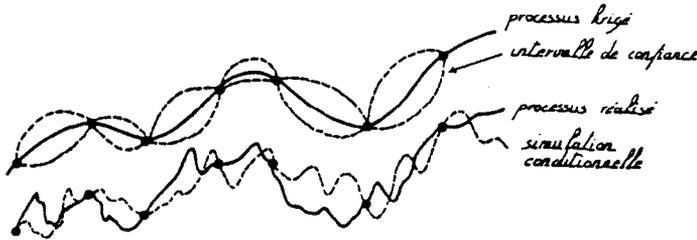
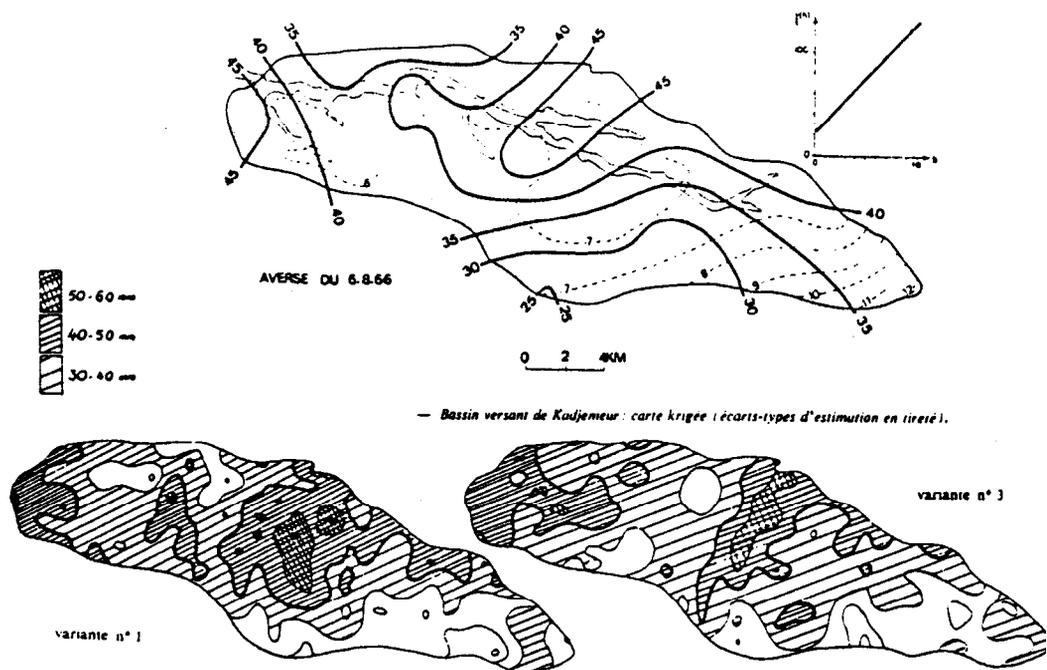


Fig 18 : Analyse d'une averse. Carte krigée et simulations conditionnelles (D'après Delhomme, 1978 ; Bulletin du BRGM (2)-III-4, 1978).



II.3. TRAITEMENT DES ERREURS ET EFFET DE PÉPITE :

On peut reprocher aux géostatisticiens de promouvoir un langage parfois ésotérique, mais il est indéniable qu'ils ont su parfois trouver des termes très images. Ainsi en est-il du fameux "effet de pépite", souvent interprété à tort comme prenant en compte les erreurs de mesure. Nous allons d'abord l'explicitier puis regarder ce que cela peut apporter dans nos applications hydrologiques.

II.3.1. Effet de pépite.

Supposons que le processus $Z(t)$ soit la superposition (Fig. 19) de deux processus (que l'on peut supposer centrés pour simplifier les calculs) :

- un processus $X(t)$ de fonction de corrélation $C_{xx}(t, t')$
- et un bruit blanc $BB(t)$ dont la fonction de corrélation est tout simplement :

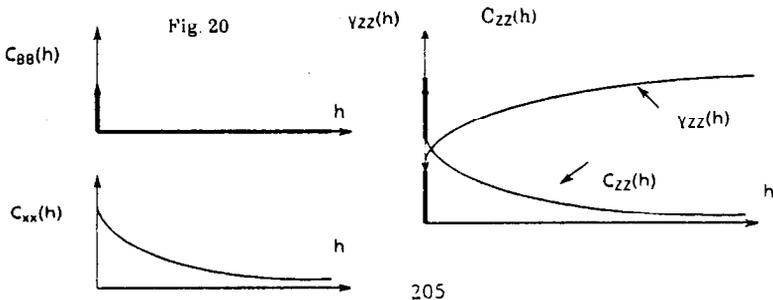
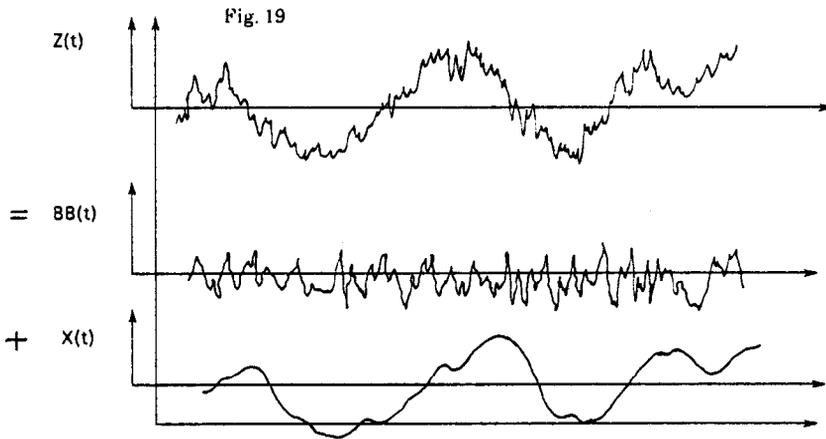
$$E[BB(t).BB(t')] = \begin{cases} \text{var}[BB] = c & \text{si } t = t' \\ 0 & \text{si } t \neq t' \text{ car le processus est décorréle} \end{cases} \quad (47)$$

On montre alors simplement que la covariance de $Z(t) = X(t) + BB(t)$ est

$$C_{ZZ}(t) = \begin{cases} C_{xx}(t, t') & \text{si } t \neq t' \\ \text{var}[X] + \text{var}[BB] = C_{xx}(0) + C & \text{si } t = t' \end{cases} \quad (48)$$

soit encore

$$C_{ZZ}(t, t') = (\delta_{tt'})C + C_{xx}(t, t') \quad \text{avec} \quad \begin{cases} \delta_{tt'} = 1 & \text{si } t = t' \\ = 0 & \text{sinon} \end{cases}$$



De même en variogramme : pour $h \rightarrow 0$, le variogramme de Z ne tend plus vers l'origine.

Cela ne signifie pas pour autant que les mesures soient entâchées d'erreur, au contraire, mais plutôt qu'elles perçoivent à la fois la composante structurée $X(t)$ et une composante sans structure spatiale $BB(t)$.

Dans le système de krigage, dont le développement est inchangé, on trouvera donc, comme d'habitude, 0 sur la diagonale, car $\gamma(0) = 0$, et $\gamma_{ZZ}(h) = C + \gamma_{XX}(h)$ sur les autres termes.

En fait, on sent quand même que plus le terme C est grand, plus il tend à modifier le système. Cela se voit mieux encore si on l'écrit en covariance car alors le système s'écrit :

$$\begin{pmatrix} C + \text{var}[X] & & & & 1 \\ & \text{Cov}_{XX}(t^1, t^2) & & & \vdots \\ & \text{Cov}_{XX}(t^1, t^3) & & & \vdots \\ & & \text{Cov}_{XX}(t^1, t^n) & & \vdots \\ & & & C + \text{var}[X] & 1 \\ 1 & \dots & \dots & \dots & 0 \end{pmatrix} \begin{matrix} \lambda_1 \\ \vdots \\ \lambda_n \\ v \end{matrix} = \begin{pmatrix} \vdots \\ C \cdot \delta(t^i, t^0) + \text{Cov}_{XX}(t^i, t^0) \\ \vdots \\ 1 \end{pmatrix} \quad (50)$$

Et on voit que plus C augmente par rapport à $\text{var}[X]$, et donc à $\text{Cov}[X(t^i), X(t^j)]$, plus le système devient à diagonale dominante. Et à la fin, la solution tend vers :

$$\lambda_1 = \lambda_2 = \dots = \lambda_n = \frac{1}{n} \quad \text{si } t^0 \neq t^i \forall i$$

car le bruit blanc est tellement dominant que la meilleure estimation, en dehors des points observés, est simplement la moyenne !... Mais les points observés sont toujours interpolés exactement car ils ne comportent pas d'erreur.

Ce même système s'écrit, en variogramme :

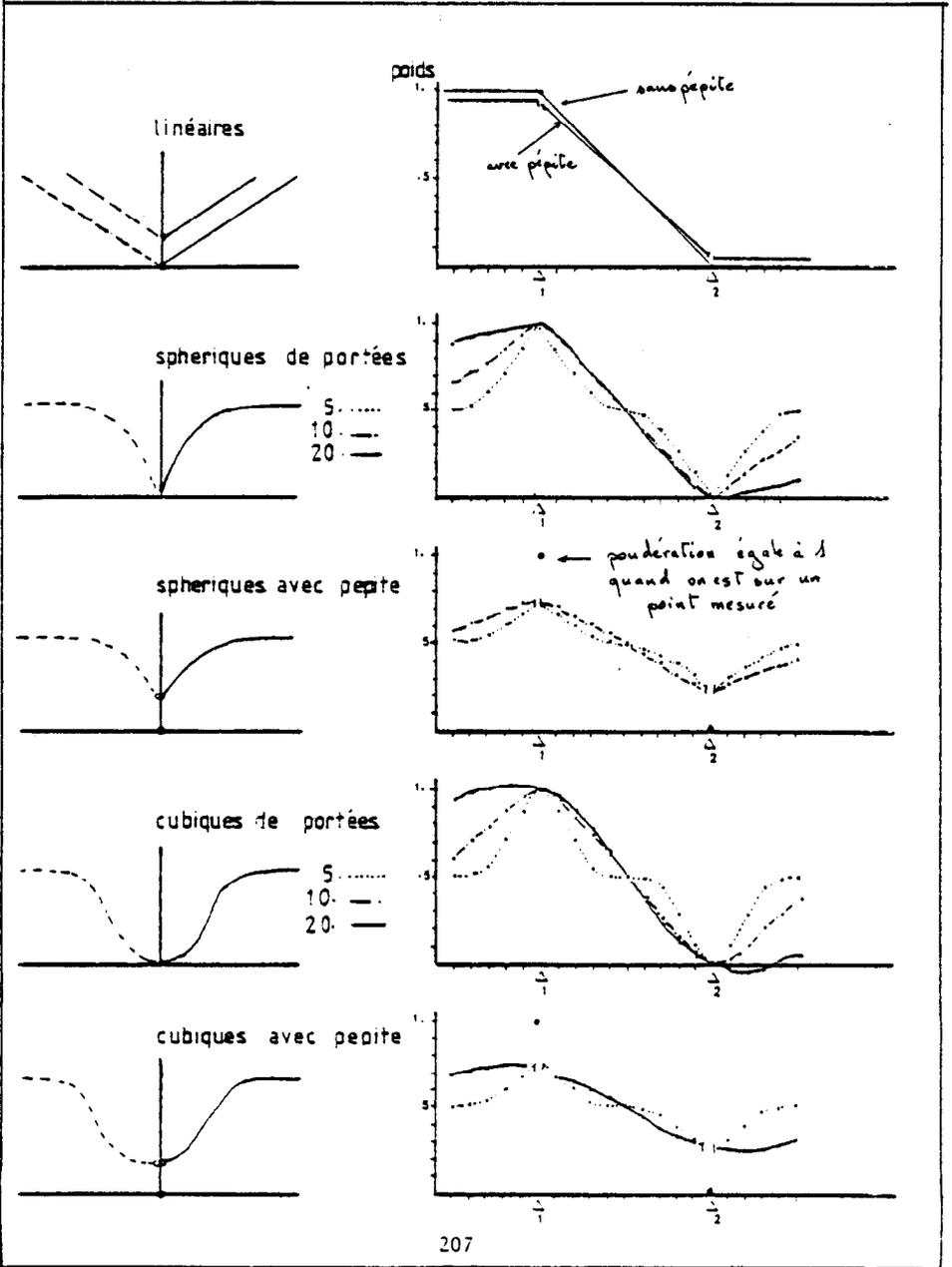
$$\begin{pmatrix} 0 & & & & 1 \\ & C + \gamma_{XX}(t^1, t^2) & & & \vdots \\ & C + \gamma_{XX}(t^1, t^3) & & & \vdots \\ & & C + \gamma_{XX}(t^1, t^n) & & \vdots \\ & & & 0 & 1 \\ 1 & \dots & \dots & \dots & 0 \end{pmatrix} \begin{matrix} \lambda_1 \\ \vdots \\ \lambda_n \\ v \end{matrix} = \begin{pmatrix} C \cdot \delta(t^1, t^0) + \gamma_{XX}(t^1, t^0) \\ \vdots \\ C \cdot \delta(t^i, t^0) + \gamma_{XX}(t^i, t^0) \\ \vdots \\ C \cdot \delta(t^n, t^0) + \gamma_{XX}(t^n, t^0) \\ 1 \end{pmatrix} \quad (50)$$

C'est celui qu'on utilisera dans les exemples suivants.

On notera aussi que l'on peut "théoriquement", à l'aide des points observés, inférer correctement les variogrammes de Z, de X et la "pépite" c.

On montre sur la figure 21 extraite de H. Darricau (1981) l'effet de l'introduction d'un terme de pépite dans le variogramme initial.

Fig. 21 : Pondération sur une droite à partir de deux points observés, selon le variogramme choisi et la présence ou non de pépite (D'après H. Darricau, 1981).

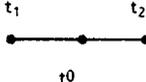


Exemple :

Soit à kriger le point t^0 à partir des deux points t^1 et t^2 , à l'aide d'un variogramme

$$\gamma_{zz}(d) = C(1 - \delta(d)) + \gamma_{xx}(d)$$

a) Si on prend t^0 au centre, le système s'écrit :



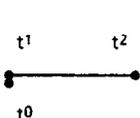
$$\begin{pmatrix} 0 & C + \gamma_{xx,21} & 1 \\ C + \gamma_{xx,12} & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ v \end{pmatrix} = \begin{pmatrix} C + \gamma_{xx,01} \\ C + \gamma_{xx,02} \\ 1 \end{pmatrix} \quad (51)$$

avec $\gamma_{xx,01} = \gamma_{xx,02}$, et on vérifie que $\lambda_1 = \lambda_2 = \frac{1}{2}$

et accessoirement

$$v = \gamma_{xx,01} + \frac{1}{2}(c - \gamma_{xx,21})$$

b) Si par contre le point t^0 coïncide avec un point d'interpolation : le système devient :

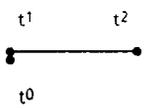


$$\begin{pmatrix} 0 & C + \gamma_{xx,21} & 1 \\ C + \gamma_{xx,12} & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ C + \gamma_{xx,12} \\ 1 \end{pmatrix} \quad (52)$$

On trouve alors $\lambda_1 = 1$, $\lambda_2 = 0$ et $v = 0$

c) Mais si le point t^0 se rapproche de t^1 à une distance très petite ϵ , le système reste différent :

Dans ce cas, pour $t^0 = t^1 + \epsilon$



$$\begin{pmatrix} 0 & C + Y_{xx,21} & 1 \\ C + Y_{xx,12} & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \nu \end{pmatrix} \begin{pmatrix} C \\ C + Y_{xx,12} \\ 1 \end{pmatrix} \leftarrow \text{au lieu de } 0 \quad (53)$$

$$\lambda_1 = 1 - \frac{1}{2} \frac{C}{C + Y_{xx,21}} \quad \text{et} \quad \lambda_2 = \frac{1}{2} \frac{C}{C + Y_{xx,21}} \quad (54)$$

On voit que si $C = 0$ on retrouve le cas b) et que si C est petit devant la variance de $X(t)$, donc devant $Y_{xx,21}$, le terme correctif est faible, mais non nul.

11.3.2. Données entachées d'erreur :

On considère cette fois que l'on a un "vrai" processus $Z(t)$, que l'on cherche à estimer en un point quelconque, mais en sachant que l'on a mesuré, en des points t_i , $i = 0, \dots, n$

- la variable $Y(t_i) = Z(t_i) + e_{r_i}$,
- où e_{r_i} est une erreur, évidemment inconnue, mais qui vérifie, par exemple :
- $E[e_{r_i}] = 0$
- $\text{cov}[e_{r_i}, e_{r_j}] = \text{cov}[e_{r_i}, Z(t)] = 0 \quad \forall i, j, t$
- i.e elles sont non corrélées entre elles, et non corrélées avec la variable exacte $Z(t)$
- $\text{var}[e_{r_i}] = \sigma_r^2$

On suppose de plus, pour simplifier, que $Z(t)$ et donc $Y(t)$ sont stationnaires d'ordre 2. On vérifie alors que :

$$E[Y(t)] = E[Z(t)] = m_z = m$$

$$\text{Var}\{Y(t_i)\} = \text{Var}\{Z(t_i)\} + \text{Var}\{e_{r_i}\} = C_{ZZ}(0) + \sigma_r^2$$

$$\text{Cov}\{Y(t_i), Y(t_j)\} = \text{cov}\{Z(t_i), Z(t_j)\} = C_{ZZ}(t_i, t_j)$$

Mais on veut cette fois estimer, à l'aide des données erronnées $Y(t_i)$, le processus

$$Z^*(t^0) = \sum \lambda_i Y(t^i)$$

Le système s'écrit alors, après dérivation comme pour le système (28) :

$$\begin{pmatrix} C_{ZZ}(0) + \sigma_1^2 & & & 1 \\ & C_{ZZ}(t, t) & & \vdots \\ & & C_{ZZ}(0) + \sigma_2^2 & \vdots \\ C_{ZZ}(t, u) & & & \vdots \\ & & & C_{ZZ}(0) + \sigma_n^2 & 1 \\ \hline 1 & \dots & 1 & & 0 \end{pmatrix} \begin{array}{c} \lambda_1 \\ \vdots \\ \lambda_n \\ \hline v \end{array} = \begin{pmatrix} \vdots \\ \vdots \\ \text{Cov}_{ZZ}(t^i, t^0) \\ \vdots \\ \vdots \\ \hline 1 \end{pmatrix} \quad (56)$$

Cela peut s'écrire aussi en variogramme, avec $\gamma_{ZZ}(t^i, t^j) = \frac{1}{2}E[(Z(t^i) - Z(t^j))^2]$ variogramme du processus exact $Z(t)$:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma_{ZZ}^{ij} - \lambda_i \sigma_i^2 + v = \gamma_{ZZ}^{i0} & i = 1 \dots n \\ \sum \lambda_j = 1 \end{cases} \quad (57)$$

Par contre, ici, le variogramme "brut" que nous pouvons inférer des valeurs observées $Y(t)$ représente en fait :

$$\begin{aligned} \gamma_{YY}(t^i, t^j) &= \frac{1}{2}E[(Y(t^i) - Y(t^j))^2] = \frac{1}{2}E[Z(t^i) + e_{r_i} - Z(t^j) - e_{r_j}]^2 \\ &= \gamma_{ZZ}(t^i, t^j) + \frac{1}{2}(\sigma_i^2 + \sigma_j^2) \end{aligned} \quad (58)$$

Mais le système (57) doit être rempli avec $\gamma_{ZZ}(t^i, t^j)$:

$$\begin{pmatrix} -\sigma_1^2 & & & 1 \\ & \gamma_{ZZ}(t, t) & & \vdots \\ & & -\sigma_2^2 & \vdots \\ \gamma_{ZZ}(t, u) & & & \vdots \\ & & & -\sigma_n^2 & 1 \\ \hline 1 & \dots & 1 & & 0 \end{pmatrix} \begin{array}{c} \lambda_1 \\ \vdots \\ \lambda_i \\ \vdots \\ \lambda_n \\ \hline v \end{array} = \begin{pmatrix} \vdots \\ \gamma_{ZZ}(t^1, t^0) \\ \vdots \\ \gamma_{ZZ}(t^i, t^0) \\ \vdots \\ \gamma_{ZZ}(t^n, t^0) \\ \hline 1 \end{pmatrix} \quad (57)$$

et

$$\sigma_E(t^0) = \sum \lambda_i \gamma_{ZZ}(t^i, t^0) + v$$

Il y a donc lissage, d'autant plus important que la variance d'erreur σ_1^2 sur $Z(t^1)$ est forte : On ne passe plus par les points mesurés car on sait qu'ils sont entachés d'erreur.

Ici encore, on peut voir que si les variances d'erreurs deviennent "grandes" par rapport à celle du processus Z , le système (56) devient à diagonale dominante, et la meilleure estimation de Z , en tous points y compris les points mesurés, tend vers la moyenne arithmétique des valeurs mesurées.

II.3.3. Comparaison entre les deux approches précédentes.

Les deux cas II.3.1. et II.3.2. ne sont pas vraiment fondamentalement différents mais correspondent à des pratiques expérimentales différentes.

Supposons que l'on dispose d'une réalisation du processus $Z(t)$; par exemple, la profondeur d'une couche géologique, dans laquelle on a fait des sondages en t^1, \dots, t^n .

- a) Si on mesure plusieurs fois Z en t^1 et que l'on trouve la même valeur $Z(t^1)$, on s'est assuré ainsi qu'il n'y avait pas d'erreur de mesure.

Ensuite, on fait un trou en $t^1 + \varepsilon$, et on constate que $Z(t^1 + \varepsilon) \neq Z(t^1)$. De même en $t^1 - \varepsilon$. On soupçonnera alors le vrai processus $Z(t)$ de comporter, de manière intrinsèque, un bruit blanc, un "effet de pépite", qui fait partie intégrante de $Z(t)$ et traduit une variabilité locale.

Mais on ne remettra pas en cause les $Z(t^i)$ et on cherchera, en t^0 , à estimer le processus $Z(t^0)$ à l'aide des $Z(t^i)$.

- b) Si par contre, en mesurant plusieurs fois la profondeur au point t^1 , on obtient des valeurs variant de ± 5 cm, à cause du jeu des appareils de mesure, des incertitudes de lecture etc., on peut conclure, en gardant la première mesure
- que celle-ci, soit $Y(t^1)$, ne représente pas $Z(t^1)$ mais $Z(t^1) + e_r(t^1)$
 - avec une erreur $e_r(t^1)$ inconnue, mais dont on pourrait (par exemple en répétant la mesure) connaître la variance σ_r^2 . (En fait, dans ce cas de mesures répétées, on pourrait prendre la moyenne des $Y(t^1)$ qui serait meilleure qu'une quelconque des mesures)
 - Et si par exemple : $\sigma_r^2 = \sigma_r^2 = \dots \forall i, j$ (On utilise le même appareil, avec la même variance d'erreur, partout). alors la série obtenue $Y(t^i)$ est en fait une mesure des $Z(t^i)$, à laquelle on aura superposé un bruit blanc externe, de variance σ_r^2 .

Mais ce que l'on veut estimer en tout point t^0 (y compris en t^1, \dots, t^n), c'est la valeur de $Z(t^0)$ et non celle de $Y(t^0) = Z(t^0) + e_r(t^0)$, mais cette estimation ne peut se faire qu'à partir des seules valeurs erronées $Y(t^i)$.

On cherche donc, contrairement au cas de la "pépite", à filtrer les données observées.

Avec ces considérations, on a fait pratiquement le tour des développements vraiment nécessaires aux applications cartographiques en hydrométéorologie. On retiendra néanmoins, à l'issue de ce chapitre sur les erreurs mais plus généralement, pour tout ce qui précède sur le krigeage, que celui-ci, à l'inverse de l'approche régression multiple-analyse des données, s'appuie sur des espérances mathématiques théoriquement optimisées. En particulier, s'il a fortement contribué à relativiser les résultats en insistant sur l'écart type d'estimation, qui est encore une espérance mathématique, il a un peu ignoré les fluctuations dues à l'échantillonnage.

On trouve peu de références (Muñoz, 1987) sur l'incertitude d'échantillonnage dans l'estimation du variogramme, et pratiquement pas de travaux sur les répercussions de cette incertitude sur les coefficients de krigeage λ_i , ou sur l'écart type d'estimation. C'est probablement le domaine où des développements utiles sont à attendre.

III EXTENSIONS DE LA METHODE

Dans ce chapitre, on s'attachera plutôt aux extensions présentant un caractère de généralisation, plutôt qu'au traitement de cas particuliers.

Par rapport au chapitre II, les démonstrations seront esquissées de manière moins complète, car il s'agit plus de fournir des idées qui enrichissent l'interprétation que d'être rigoureux mathématiquement.

III.1. Krigage d'une moyenne spatiale :

Jusqu'ici, on a cherché à estimer une valeur ponctuelle, en un point non mesuré, $Z^*(t^0)$. Désormais, on souhaite estimer, de manière optimale, une nouvelle variable :



$Y(D) = \int_D Z(t) dt$ ou $1/D \int_D Z(t) dt$ (63)

où Y est soit l'intégrale, soit la valeur moyenne de Z sur un domaine D quelconque.

Ici encore, on va chercher à "expliquer" Y à l'aide des valeurs ponctuelles disponibles $Z(t)$, d'où l'estimateur :

$$Y^*(D) = \sum \lambda_i(D) \cdot Z(t^i) \quad (64)$$

On minimise donc, sous des contraintes éventuelles selon les hypothèses faites sur la moyenne :

$$E\left\{\left[\frac{1}{D} \int_D Z(t) \cdot dt - \sum \lambda_i(D) \cdot Z(t^i)\right]^2\right\}$$

Dans laquelle vont apparaître des termes en :

$$\begin{aligned} E\left\{\left[\frac{1}{D} \int_D Z(t) \cdot dt\right]^2\right\} &= E\left\{\frac{1}{D^2} \left[\int \int Z(t) \cdot Z(t') \cdot dt dt'\right]\right\} = \frac{1}{D^2} \left[\int \int E\{Z(t) \cdot Z(t')\} \cdot dt dt'\right] \\ &= \frac{1}{D^2} \left[\int \int_D C(t, t') \cdot dt dt'\right] = \text{var } Y(D) \end{aligned}$$

et des termes :

$$\begin{aligned} E\left\{\frac{1}{D} \int_D Z(t) \cdot dt \times Z(t^i)\right\} &= E\left\{\frac{1}{D} \int_D Z(t) \cdot Z(t^i) \cdot dt\right\} = \frac{1}{D} \int_D E\{Z(t) \cdot Z(t^i)\} \cdot dt \\ &= \frac{1}{D} \int_D C(t, t^i) \cdot dt \end{aligned}$$

Ces derniers sont les seuls qui apparaîtront lors de la dérivation par rapport aux λ_i , et on obtiendra le même système qu'en (28 ou 32) sauf au second membre où les $C(t^0, t^1)$ sont remplacés par l'intégrale sur le domaine

$$\int_D C(t, t^1) dt,$$

de même pour les $f(t^0)$ remplacés par

$$\int_D f^1(t) dt$$

On ne s'étendra pas sur le calcul numérique de ces intégrales, qui peut parfois réserver quelques difficultés.

La démarche est identique si, au lieu de partir de (32) on part de (39) en choisissant le variogramme pour exprimer la structure.

On remarquera que, très globalement, il s'agit encore d'une espèce de corrélation multiple, mais où la variable à expliquer Y est un peu plus compliquée, et nécessite donc un calcul un peu plus compliqué des "corrélations" entre la variable à expliquer et les variables explicatives.

On peut bien sûr toujours calculer un écart type théorique d'estimation :

$$\sigma_E^2 = E[(Y - Y^*)^2] \text{ avec } Y^* \text{ conditionné par } Z(t^1) \dots Z(t^N) : \quad (67)$$

$$\sigma_E^2 = \frac{1}{D^2} \left[\int \int C(t, t^1) dt dt^1 - \sum_i \lambda_i \frac{1}{D} \int_D C(t, t^1) dt + \sum \mu_l \frac{1}{D} \int_D f^1(t) dt \right]$$

Il s'agit bien d'une variance résiduelle théorique, car elle dépend du modèle choisi $C(t, t^1)$, plus ou moins bien ajusté au processus, et des hypothèses faites sur la moyenne.

On peut imaginer que, si l'on avait plusieurs réalisations $k = 1 \dots P$, avec pour chacune une valeur observée Y_k , alors l'erreur empirique de reconstitution :

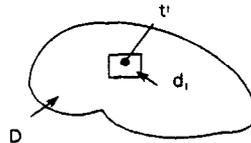
$$\frac{1}{P} \sum_k (Y_k - Y_k^*)^2 \rightarrow \sigma_E^2$$

Tous les développements décrits ci-dessus en covariance peuvent se faire en variogramme (cf. Matheron in Lafitte 1972).

Enfin, il peut arriver que les variables explicatives elles-mêmes soient le résultat d'une intégration. Dans ce cas,

$$Y = 1/D \int_D Z(t) dt \quad \text{sera expliquée par les}$$

$$V_i = 1/d \int_{d_i} Z(t) dt$$



i.e la variable explicative n'est pas ponctuelle, mais représente elle-même une moyenne autour du point t . C'est le cas... du pluviomètre (qui représente l'intégrale sur les quelques décimètres carrés du cône collecteur !). C'est celui des mesures de neige à l'aide d'un coussin à neige de quelques m^2 . C'est le cas aussi de la transmissivité d'une aquifère, mesurée par un essai de pompage, et qui représente en fait les propriétés de quelques dizaines de m^2 environnants, voire plus.

Cependant, il n'est pas nécessaire de prendre en compte ce raffinement dans la mesure où la fonction de structure utilisée est déjà celle de la donnée intégrée sur d (seule accessible à l'observation) et où les d sont de plusieurs ordres de grandeur inférieurs à D .

On trouvera une application de cette approche dans T. Lebel (1984) pour le calcul de pluie moyenne sur des bassins. Il existe bien sûr des extensions à des domaines volumiques, surtout dans le secteur minier. Mais on ne négligera pas les applications dans le domaine temporel. Il est fréquent en effet que la mesure d'un processus temporel (exemple : la vitesse du vent) soit l'objet d'une intégration sur un temps plus ou moins long, et que l'on cherche ensuite à estimer des moyennes ou interpoler des valeurs manquantes.

III.2. Liaison avec les autres techniques d'interpolation.

Le physicien ou l'ingénieur connaît souvent des méthodes élémentaires d'interpolation auxquelles il aimerait pouvoir se rattachier pour "assimiler" la Géostatistique.

Et ici encore, il est bon de commencer à une dimension, par des méthodes très couramment utilisées.

III .2.1 L'interpolation polynômiale

Prenons le cas de l'interpolation polynômiale sur n points. On sait que le seul polynôme qui interpole ces n points est de degré $n-1$, mais qu'il peut s'exprimer de différentes façons :

La première consiste à l'écrire à l'aide des monômes habituels $1, t, (t)^2 \dots (t)^{n-1}$, soit :

$$Z^*(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_{n-1} t^{n-1} \quad (68)$$

et on obtient les coefficients a_i en écrivant les n équations :

$$i = 1 \dots n \quad \{ Z^*(t^i) = a_0 + a_1(t^i) + a_2(t^i)^2 + \dots + a_{n-1}(t^i)^{n-1} \quad (69)$$

exprimant que le polynôme passe par chacun des points observés $[t^i, Z(t^i)]$.

Dans ce cas, on fait bien apparaître dans l'opération (68) les fonctions génératrices choisies $1, t, (t)^2 \dots (t)^{n-1}$.

On parlera d'une expression en interpolateur : les outils d'interpolation apparaissent clairement (ce sont les monômes) tandis que les valeurs observées, pour le polynôme particulier considéré, $(Z(t^1) \dots Z(t^n))$ sont contenues de manière peu explicites dans les coefficients $-a_0 - a_1 \dots - a_{n-1}$. Or il existe une seconde écriture du même polynôme, qui s'appuie sur des fonctions de base assez différentes ; toutes, de degré $n-1$:

$$i = 1, \dots, n \quad \left\{ L_i(t) = \frac{(t-t^1) \dots (t-t^{i-1})(t-t^{i+1}) \dots (t-t^n)}{(t^i-t^1) \dots (t^i-t^{i-1})(t^i-t^{i+1}) \dots (t^i-t^n)} \right. \quad (70)$$

On constate que ce polynôme de base a des propriétés particulières :

$$\begin{aligned} L_i(t_j) &= 0 & \text{si } j \neq i \\ L_i(t_j) &= 1 & \text{si } j = i \end{aligned}$$

Il vaut donc 1 au point t_i , 0 aux $n-1$ points d'interpolation t_j , et... n'importe quoi au point t quelconque. Ces $L_i(t)$ constituent la base de Lagrange. Et le polynôme d'interpolation, unique et strictement identique à (68), s'écrit alors :

$$Z^*(t) = Z(t^1)L_1(t) + Z(t^2)L_2(t) + \dots + Z(t^n)L_n(t) = \sum L_i(t).Z(t^i) \quad (71)$$

On vérifie d'ailleurs que si $t = t_i$, $Z^*(t) = Z(t)$

Il y a bien interpolation, en pondérant les valeurs observées par un coefficient polynomial $L_i(t)$.

On ne manquera pas de relever l'analogie avec le krigeage en un point t quelconque :

$$Z^*(t) = \sum \lambda_i(t).Z(t^i)$$

ou avec toute méthode (type pondération par la moyenne de l'inverse des carrés des distances de t aux points t^i).

Nous parlerons alors d'une expression de l'estimateur $Z^*(t)$ sous forme de pondérateur. C'est la plus classique utilisée en krigeage. Néanmoins, il peut être intéressant de regarder ce que cela devient sous une forme (il peut y en avoir plusieurs) d'interpolateur.

III.2.2 Le krigeage en interpolateur (krigeage dual) :

Reprenons pour cela le système (32) (en covariance ou en variogramme, les développements restent analogues). On constate qu'il consiste à rechercher, pour interpoler au point t , un vecteur de poids $\{\lambda_1(t) \dots \lambda_n(t)\} = \Lambda^T(t)$ (l'exposant T indique la transposition au sens matriciel) solution du système (32) réécrit :

$$G \cdot \begin{bmatrix} \Lambda(t) \\ M(t) \end{bmatrix} = \begin{bmatrix} \Gamma & F \\ F^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \Lambda(t) \\ M(t) \end{bmatrix} = \begin{bmatrix} \Gamma^*(t) \\ f^*(t) \end{bmatrix} = g_0^*(t) \quad (72)$$

où G ne dépend que des points connus t^i et pas du point courant t .

On en déduit le vecteur des pondérations (et des multiplicateurs de Lagrange) :

$$\begin{bmatrix} \Lambda(t) \\ M(t) \end{bmatrix} = G^{-1} \cdot g_0^*(t)$$

D'où l'expression de l'estimateur :

$$Z^*(t) = \begin{bmatrix} Z(t^1) \dots Z(t^n) 0 \dots 0 \end{bmatrix} \cdot \begin{bmatrix} \Lambda(t) \\ M(t) \end{bmatrix}$$

Qu'on peut encore écrire :

$$\begin{aligned} Z^*(t) &= \begin{bmatrix} Z(t^1) \dots Z(t^n) 0 \dots 0 \end{bmatrix} \cdot G^{-1} \cdot g_0^*(t) = B \cdot g_0^*(t) \\ &= [B]^T \cdot g_0(t) \end{aligned}$$

où B calculé une fois pour toutes, ne dépend que des points observés $t^1 \dots t^N$.

Soit encore :

$$Z^*(t) = [b_1, \dots, b_n | c_0, \dots, c_k] \begin{bmatrix} K(t^1, t) \\ \dots \\ K(t^N, t) \\ \hline f_0(t) \\ f_1(t) \\ \dots \\ f_k(t) \end{bmatrix} \quad (73)$$

D'où les expressions équivalentes :

$$Z^*(t) = \sum_{i=1}^N \lambda_i Z(t^i) \quad (74)$$

$$Z^*(t) = \sum_{i=1}^N b_i K(t^i, t) + \sum_{l=1}^k c_l f_l(t) \quad (75)$$

La seconde formulation est du type interpolateur puisqu'elle fait apparaître explicitement les outils de l'interpolation, à savoir :

- les fonctions de base $f_l(t)$ utilisées pour représenter la partie "tendance", ou "dérive", due à la non stationnarité de la moyenne (Fig. 22-1 à 3).
- et des fonctions $K(t, t) = K_l(t)$ correspondant au processus lui-même et dérivées directement de sa structure $-K(t, t')$, exprimée en covariance, ou en variogramme (Fig. 22-4 à 6)

Les valeurs des $Z(t)$ observées étant directement contenues dans les coefficients b_i et c_l .

Desormais on appelle cette formulation en interpolateur le krigeage dual

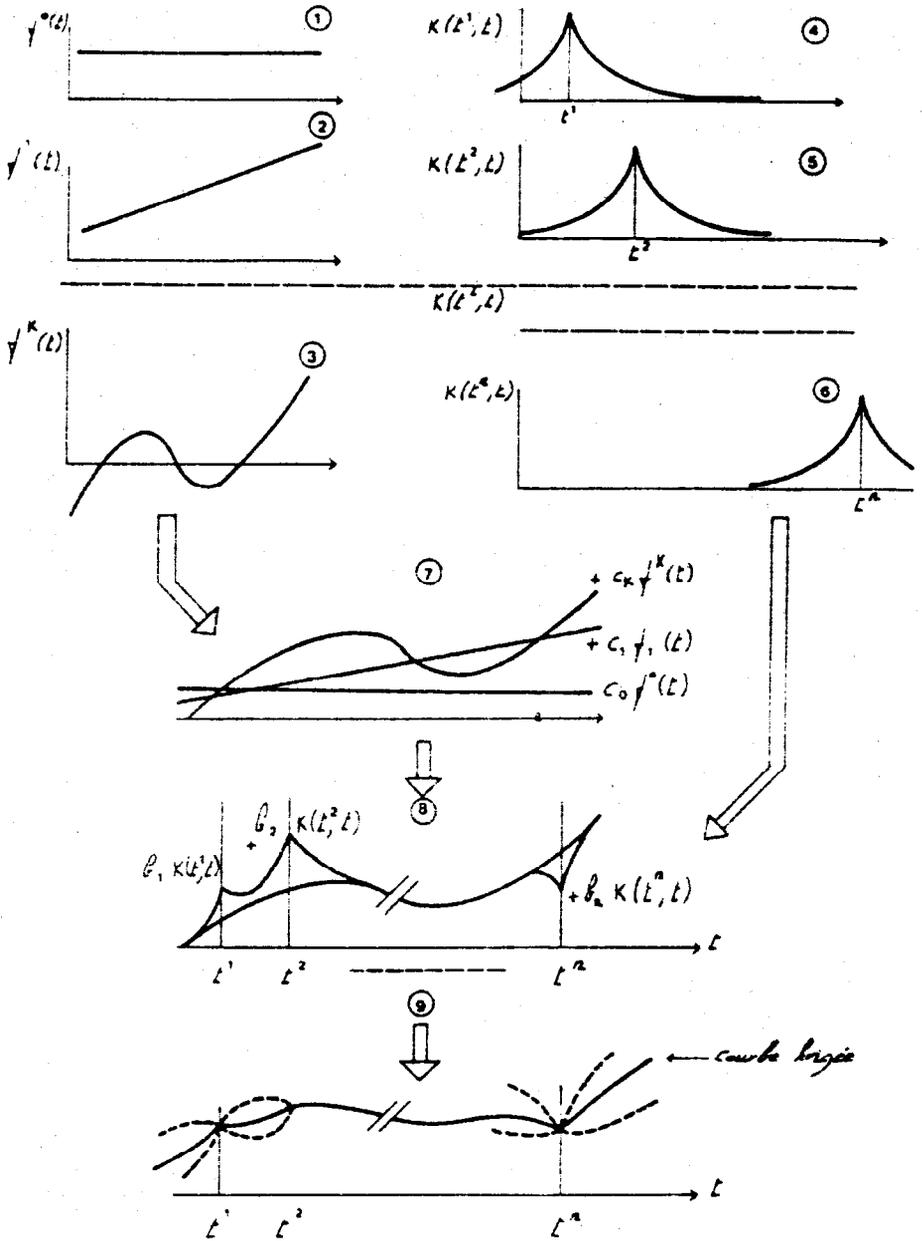
Cela permet une interprétation assez imagée du krigeage, qui nous ramène à l'approche classique en interpolation. En effet, pour reconstituer le champ estimé,

on pondère d'abord les fonctions de base $f_l(t)$ de la moyenne, d'où (Fig. 22-7)

$$\sum c_l f_l(t)$$

puis on balaye chaque point observé t^i , qui, à l'aide de la fonction de structure $K(t, t')$, préalablement adaptée à ce point d'où $K(t^i, t)$, apportera sa contribution $b_i K(t^i, t)$ sur l'ensemble du champ. Et on somme le tout (Fig. 22-8)

Figure 22 : Interpretation du krigeage dual en somme de fonctions globales et locales



III. 2.3 Cas particulier des fonctions splines :

Il s'agit d'une technique très intéressante, qui évite les inconvénients couramment rencontrés en interpolation polynômiale ou trigonométrique, essentiellement les ondulations intempestives.

La technique des fonctions splines est assez facile à mettre en oeuvre et donne en général des surfaces très lisses, bien que passant par les points d'interpolation. Il n'est pas question ici d'en exposer la théorie (voir pour cela Chenin *et al*, 1985), mais de voir les analogies avec le krigeage.

Si on considère que la fonction spline est une plaque élastique mince, infinie, qui passe par les points $Z(t)$ en minimisant son énergie de flexion, son équation $S(t) = S(x,y)$ minimise :

$$\iint_{R^2} \left(\frac{\delta^2 S}{\delta x^2} \right)^2 + \left(\frac{\delta^2 S}{\delta x \delta y} \right)^2 + \left(\frac{\delta^2 S}{\delta y^2} \right)^2 dx dy = \text{Min}$$

avec $S(t) = Z(t) \quad i = 1, \dots, n$

parmi toutes les fonctions possibles, l'intégration étant étendue au plan complet R^2 .

On montre alors que cette équation s'écrit :

$$S(x,y) = \sum_{i=1}^n \lambda_i K(x - x_i, y - y_i) + \alpha x + \beta y + \gamma \quad (76)$$

avec

$$K(x,y) = \frac{1}{2}(x^2 + y^2) \text{Log}(x^2 + y^2)$$

Les réels $\lambda_i, \alpha, \beta, \gamma$ étant définis par le système

$$\begin{aligned} \sum_{j=1}^n K(x_i - x_j, y_i - y_j) - \lambda_j + \alpha x_i + \beta y_i + \gamma &= 0 \quad i=1, n \\ \sum \lambda_j y_j &= 0 \\ \sum \lambda_j x_j &= 0 \\ \sum \lambda_j &= 0 \end{aligned} \quad (77)$$

Et on reconnaît ici la formulation en interpolateur présentée en III. 2.2, où l'hypothèse sur la moyenne serait qu'elle est localement linéaire, avec :

$$f_0(t) = 1 \quad f_1(t) = x \quad f_2(t) = y$$

et où la fonction de structure serait de la forme

$$K(d) = d^2 \cdot \text{Log } d$$

Cette forme n'est d'ailleurs pas la seule possible et on montre qu'en changeant le critère à optimiser, on trouve des fonctions $K(d)$ de la forme

$$K(d) = d^\alpha \quad \alpha = >0 \text{ et non entier pair.}$$

Or, on utilise, dans l'hypothèse intrinsèque d'ordre 1 des variogrammes de ce type.

On peut donc en conclure que l'interpolation spline n'est qu'un cas particulier de krigeage.

Or la formulation spline n'a fait, du moins explicitement, aucune hypothèse de "processus"... Accessoirement, elle appelle la fonction $K(d)$ le "noyau reproduisant". Et on voit bien dans l'expression (76) que c'est bien ce noyau qui est l'outil de reproduction de la fonction $S(t)$ qui interpole les $Z(t_i)$. D'où une convergence entre les méthodes que l'utilisateur peut certes ignorer, mais qui enrichit l'interprétation et relativise les querelles d'école....

CONCLUSIONS

Au terme de cet article, j'espère que les hydrologues accepteront plus volontiers de considérer l'approche géostatistique comme appropriée pour l'interprétation de leurs données souvent distribuées dans l'espace.

Mathématiquement, elle présente des caractères semi-arides assez marqués, d'où sans doute sa première utilisation en zone sahélienne (Delhomme et Delfiner, 1973). Néanmoins, sa mise en oeuvre élémentaire n'est pas très compliquée. Des programmes sont donnés (David, 1977) ou sont accessibles à des prix modiques pour micro-ordinateur même s'ils n'ont pas l'exhaustivité de la Formule 1 du genre, le logiciel Blue pack.

En pratique, on commence à voir de nombreuses applications en hydrologie de surface et en hydrométéorologie, qui prennent alors en compte la répétition des champs dans le temps, par exemple les champs pluviométriques (Lebel et al à paraître). L'application en hydrologie souterraine, proche du secteur minier, est déjà ancienne, mais on voit de plus en plus d'applications en hydropédologie de la zone non saturée.

Enfin l'avenir est probablement à l'utilisation conjointe de plusieurs sources d'informations, par exemple des mesures sol et des valeurs télédéteçtées qui débouchent sur les techniques de cokrigeage (Creutin J.D., 1987) que nous n'avons pu aborder ici.

REMERCIEMENTS

L'auteur remercie J.P. Delhomme, à l'époque au Centre d'Informatique Géologique à Fontainebleau, qui a guidé nos premiers pas en Géostatistique. Il remercie aussi les membres, passés ou présents, de l'équipe d'Hydrologie de Surface et Hydrométéorologie de l'IMG Grenoble, pour les emprunts à leurs travaux et pour leurs suggestions dans la réalisation de cet article. Et surtout, l'auteur voue une reconnaissance sincère à ceux qui ont assuré la frappe des notations et la mise en forme de cet article...

BIBLIOGRAPHIE

- CHENIN P., COSNARD, M. GARDAU Y., ROBERT F., ROBERT Y., WITOWSKI P. (1985) "Méthodes de Base. Mathématiques et CAO". Hermès Pub. Ed. 170 p.
- CREUTIN J.D. et OBLED Ch. (1982) "Objective Analyses and Mapping Techniques for Rainfall Fields: An Objective Comparison". Wat. Res. Res. vol. 18 n°2 p 413-431.
- CREUTIN J.D. (1987) "Validation et étalonnage d'images de télédétection à l'aide de mesures ponctuelles selon une approche géostatistique. Application à la mesure des précipitations et de l'insolation". Thèse d'Etat. INPG Grenoble. Mars 1987.
- DUBAND D. (1982) "Cours d'hydrologie statistique" ENS d'Hydraulique Grenoble. Polycopié - 200p.
- DARRICAU-BEUCHER H. (1981) "Approche géostatistique du passage des données de terrain aux paramètres des modèles en hydrogéologie". Thèse de D.I. ENS des Mines de Paris.
- DAVID M. (1977) "Geostatistical Ore reserve estimation". Elsevier Pub. Company. 364 p.
- DELHOMME J.P. et DELFINER P. (1973) "Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride" Symposium Unesco WMO-IAHS. Publi. AIHS n°
- DELHOMME J.P. (1976) "Applications de la théorie des variables régularisées dans les sciences de l'eau". Thèse de Docteur Ingénieur. Ecole des Mines de Paris. Université de Paris VII.
- DELFINER P., DELHOMME J.P., CHILES J.P., RENARD D., IRIGOIN F. (1982). BLUEPACK 3D (Manuel d'utilisation régulièrement mis à jour). Centre de Géostatistique et de Morphologie Mathématique, 35, rue St Honoré - 77305 Fontainebleau.
- GANDIN L.S. (1965) "objective analysis of meteorological fields". Leningrad, 1963. Traduit du russe par Israel. Program for Scientific Translation. Jerusalem, 1965. 242 p.
- HERGE, (1954) "On a marché sur la lune". Casterman. Ed. 63 p.
- JOHNTSON J. (1972) "Econometric methods" International Student Edition. 2nd Ed. Mac Graw Hill. 437 p.
- JOWETT (1955) "The comparison of means of sets of observations from sections of independent stochastic series". J. of Roy. Statist. soc. serie B. Vol. 17, n° 2, pp. 208-227
- LAFITTE P. (1972) traité d'informatique géologique. Ouvrage collectif, Masson et Cie Ed. 624 p.
- LEBEL T. (1984) Moyenne spatiale de la pluie sur un bassin versant : estimation optimale, génération stochastique, et gradex des valeurs extrêmes. Thèse de Docteur Ingénieur. Institut National Polytechnique de Grenoble (1984).
- LEBEL T. et BASTIN G. (1985) "Variogram identification by the Mean Square Interpolation Error method with application to hydrology fields". J. of Hydrology. Vol. 77. n°..... p. 31-56
- LEBEL T., BASTIN G., OBLED Ch., CREUTIN J.D (1987) "On the accuracy of Areal Rainfall estimation : a case study" A paraître dans Wat. Res. Res.
- LEBEL T. et BASTIN G. (1985) "Variogram identification by mean squared interpolation error method with applications to hydrologic fields" J. of Hydrology.

LEBEL T., OBLED Ch., CREUTIN J.D. (à paraître) "Application de la géostatistique à la cartographie des phénomènes pluvieux : aspects pratiques". Soumis à la Rev. d'Hydrol. Continentale en 1987.

MATHERON G. (1965) "Variables généralisées et leur estimation". Masson et Cie, Editeur.

MUNOZ-PARDO J.F. 1987 "Approche géostatistique de la variabilité spatiale des milieux géophysiques - Application à l'échantillonnage de phénomènes bidimensionnels par simulation d'une fonction aléatoire". Thèse de Docteur Ingénieur INPG-USMG. Février 1987. 350 p.

ORFEUIL J.P. (1976) "Géostatistique et analyse des données" in *Advanced Geostatistics in the Mining Industry*. p. 423-434. D. Reidel Publishing Company Ed.

PHILIP G.M., et WATSON D.F. (1986) "Matheronian geostatistics. Quo Vadis ?" *Mathematical geology*. Vol 18 n° 1. pp. 93-117

SALAS J., DELLEUR J.W., YEVJEVICH V. and LANE W. (1980) "Applied Modelling of hydrologic time series". *Wat. Res. Publ.* 484 p.

SLIMANI M. (1985) "Etude des pluies de fréquence rare à faible pas de temps sur la région Cévennes Vivarais: estimation, relation avec le relief, et cartographie synthétique". Thèse de Doctorat. INPG. Grenoble. Septembre 1985.

SRIVASTAVA R.M. (1986) "Philip and Watson - Quo Vadis ?". *Mathematical geology*. Vol. 18, pp. 141-146.