APPLICATIONS DE LA GÉOSTATISTIQUE A QUELQUES PROBLÈMES DE FILTRAGE

THÈSE

présentée à

l'Ecole Nationale Supérieure des Mines de Paris

par

Colin DALY

pour obtenir le titre de

Docteur en Géostatistique

Soutenue le 28 Juin 1991

devant le Jury composé de :

M. G. MATHERON	Président
M. P. HAWKES	Rapporteur externe
M. Ch. LAJAUNIE	Rapporteur interne
M^{me} D. BENOIT	Examinateur
M. J. HASLETT	Examinateur
M. D. JEULIN	Examinateur

For Aisling and Caia.

REMERCIEMENTS

Je voudrais remercier le Professeur Georges MATHERON, Directeur du Centre de Géostatistique de l'Ecole des Mines de Paris et Président de mon jury, qui m'a donné la possibilité de travailler au Centre de Géostatistique. Ses travaux scientifiques ainsi que ses conseils et commentaires ont profondément influencé ce mémoire d'une manière qui sera évidente au lecteur.

Je souhaite exprimer ma reconnaissance à M. Dominique JEULIN, mon Directeur de Thèse, qui m'a constamment incité à écrire les résultats de mon travail dans des délais raisonnables. Sa remarquable capacité de travail et la diversité des idées qu'il a produites durant nos nombreuses discussions ont certainement laissé leur trace.

Je voudrais remercier M. Peter HAWKES, Directeur de Recherche au Laboratoire d'Optique Electronique du CNRS à Toulouse, qui a accepté d'être le Rapporteur Externe de cette thèse. Ses commentaires et ses remarques sur le texte ont été très utiles et très instructives.

Je remercie également M. Christian LAJAUNIE, Rapporteur Interne, avec lequel j'ai eu des discussions intéressantes et fructueuses. Sa clarté et sa rigueur intellectuelle ont beaucoup aidé ce travail.

Je tiens à remercier Mme Danielle BENOÎT, chef de service à l'IRSID, qui en tant que métallurgiste a largement influencé le choix des méthodes utilisées dans cette thèse, méthodes développées dans le but d'avoir des applications routines pour les images microsondes.

Si je suis venu à Fontainebleau, c'est dans une large mesure grâce aux conseils de M. John HASLETT, Senior Lecturer in Statistics, Trinity College, Dublin, qui m'a fait connaître la Géostatistique. Je voudrais témoigner ici mon respect pour ses connaissances de la statistique, son regard critique et finalement sa passion évidente pour cette discipline.

A l'ensemble des chercheurs du Centre de Géostatistique, j'offre mes remerciements les plus vifs. Je pense que personne n'a échappé aux questions péniblement exprimées et souvent mal formulées (quand elles n'étaient pas carrément mal posées). Je pense en particulier à M. Christian LANTUÉJOUL et à mes camarades de bureau, M. Xavier FREULON et M. Pierre PETITGAS.

Je voudrais remercier très sincèrement Mme Françoise POIRIER qui a traduit ce mémoire de franglais en français et qui a partagé le travail d'organisation pratique de ce texte.

Je remercie enfin M. Jacques LAURENT pour ses illustrations et pour la patience avec laquelle il a dû à maintes reprises récupérer les fichiers que j'avais perdus.

TABLE DES MATIÈRES

Chapitre 0 : Introduction *1*

Chapitre I : La randomisation *3*

1.0. — INTRODUCTION 3
1.1. — LES DONNÉES 4
1.2. — LA RANDOMISATION 7
1.3. — LES FONCTIONS ALÉATOIRES 9
1.4. — COVARIANCES 12
1.4.1. — Estimation de la fonction de covariance 16
1.5. — LES LOIS SOUS-JACENTES 26

Chapitre II : Le filtrage linéaire *29*

2.0. — INTRODUCTION 29
2.1. — LE KRIGEAGE UNIVARIABLE 31
2.1.1. — Filtrage univariable 31
2.1.2. — L'estimation des composantes spatiales 46
2.2. — LE CAS MULTIVARIABLE 53
2.3. — ESTIMATION D'UN ÉLEMENT NON OBSERVÉ 58

Chapitre III : Estimation de l'histogramme sous-jacent *63*

3.0. — INTRODUCTION 63
3.1. — POISSONISATION D'UNE LOI 64
3.2. — ESTIMATION PAR MOINDRES CARRÉS 65
3.2.1. — Problèmes mal posés 66
3.2.2. — Maximum de vraisemblance 68
3.3. — LES SOLUTIONS RÉGULARISÉES 68
3.3.1. — Le krigeage et la régularisation 69
3.3.2. — La validation croisée 72
3.3.3. — Deux acquisitions 76
3.3.4. — La norme d'erreur 76
3.3.5. — Estimation de K 78
3.4. — EXEMPLES ET CAS D'ÉTUDE 81

Chapitre IV : Les problèmes mal posés *103*

4.0. — INTRODUCTION 103
4.1. — DÉFINITIONS ET THÉORÈMES 106

4.2. — POISSONISATION ET CONTINUITE DANS ℓ^1 111
4.2.1. — Notation 111
4.2.2. — Existence et continuité de P^{-1} 111
4.3. — POISSONISATION ET CONTINUITE DANS ℓ^2 114
4.4. — ESTIMATION PAR MAXIMUM DE VRAISEMBLANCE 119
4.4.1. — Les conditions de Kuhn et Tucker 119
4.4.2. — Le maximum de vraisemblance 120
4.4.3. — Le χ^2 à point fixe 122
4.5. — NOTES ET COMMENTAIRES 124

Chapitre V : La régularisation *127*

5.0. — INTRODUCTION 127
5.1. — RÉGULARISATION DANS UN ESPACE MÉTRIQUE 127
5.2. — DÉFINITIONS ET THÉORÈMES 129
5.3. — RÉGULARISATION ET KRIGEAGE 133
5.3.1. — La régularisation dans un espace de Hilbert 133
5.3.2. — Le krigeage dans un espace de Hilbert 135
5.4. — THÉORÈMES D'ÉQUIVALENCE DE LA RÉGULARISATION ET DU KRIGEAGE 138

Chapitre VI : Le filtrage non linéaire *157*

6.0. — INTRODUCTION 157
6.1. — ESTIMATEURS PROVENANT D'UNE PROJECTION 158
6.1.1. — Krigeage simple 159
6.1.2. — Le krigeage disjonctif 160
6.1.3. — L'espérance conditionnelle 161
6.2. — LE KRIGEAGE DISJONCTIF 162
6.2.1. — Un exemple simple 163
6.3. — MODÈLES ISOFACTORIELS 172
6.3.1. — Modèles isofactoriels hermitiens 175
6.3.2. — Modèles isofactoriels de Laguerre 177
6.3.3. — Modèle isofactoriel binomial discret 180
6.3.4. — Un modèle mixte 181
6.4. — MISE EN OEUVRE D'UN KRIGEAGE DISJONCTIF 182
6.4.1. — L'anamorphose empirique 182
6.4.2. — Ajustement des lois dissymétriques 184
6.4.3. — Filtrage et histogramme local 186
6.5. — L'ESPÉRANCE CONDITIONNELLE 187
6.6. — EXEMPLES 189

Chapitre VII : Conclusion *229*

Références *231*

CHAPITRE 0

INTRODUCTION

Dans un grand nombre d'applications, la variable d'intérêt ne peut pas être observée en soi, car elle est cachée par un bruit expérimental lié aux conditions d'acquisitions. Les exemples les plus connus se rencontrent dans le domaine du traitement du signal ou celui de l'analyse d'images. Un exemple particulier est la visualisation d'une carte chimique d'un élément dans un acier à l'aide d'une microsonde électronique où la concentration vraie est partiellement obscurcie par un bruit poissonien.

Il est souvent d'un intérêt capital de pouvoir connaître certaines informations sur les données sous-jacentes à partir des données bruitées. Par exemple, on peut vouloir obtenir une carte moins bruitée pour mieux interpréter la structure sous-jacente, ou bien on peut vouloir connaître la loi de la variable sous-jacente. Ce sont là des problèmes de filtrage.

L'introduction du filtre linéaire optimal au sens de A.N. Kolmogorov et N. Wiener dans les années 30 et 40 a constitué un pas très important dans la théorie du filtrage. Le filtrage géostatistique linéaire, ou krigeage linéaire, que nous étudierons au chapitre II, n'est autre mathématiquement qu'une réécriture du filtre de Wiener, du moins dans le cas stationnaire. La différence entre le krigeage linéaire et le filtre classique réside dans l'approche méthodologique. En géostatistique, le rôle donné à la fonction structurale, le variogramme, est central. Une interprétation physique de cette fonction nous conduit à l'idée que l'hypothèse de stationnarité n'est pas essentielle pour l'application d'un filtre linéaire, et l'on peut introduire l'idée de stationnarité locale et utiliser des voisinages glissants pour appliquer le filtre linéaire dans les cas ne présentant pas une stationnarité idéale. Ces idées seront abordées dans ce travail, mais il ne sera pas possible de les expliciter complètement sans trop nous éloigner du sujet à traiter, à savoir le filtrage. Une introduction complète détaillée à la modélisation géostatistique est donnée dans le livre de G. Matheron [3].

Depuis les années soixante, une littérature abondante s'est développée sur le filtrage statistique non linéaire. Malgré une mathématique parfois très intéressante - ou peut être à cause de celle-ci, ces idées n'ont pas été aussi bien reçues que

le filtrage linéaire de Wiener et ses dérivées telles que le filtre de Kalman. Une des causes principales de cette réticence est le fait que chaque estimateur non linéaire est généralement adapté à une classe spécifique de problèmes et peut donner de très mauvais résultats s'il est appliqué en dehors de cette classe. Comme ces estimateurs sont souvent difficiles à ajuster aux données expérimentales, ils apparaissent peu robustes. De plus, pour des problèmes de filtrage simples, par exemple la visualisation d'une carte moins bruitée, il est souvent vrai que des filtres sophistiqués n'apportent qu'une amélioration marginale par rapport aux méthodes linéaires même quand ils sont appliqués correctement.

Il existe néanmoins des problèmes de filtrage qui sont intrinsèquement non linéaires, par exemple l'estimation d'un histogramme local, pour lesquels les méthodes linéaires ne sont vraisemblablement pas adaptées et il faut alors construire un filtre non linéaire. Dans le chapitre VI, nous proposons un filtre non linéaire basé sur la théorie du krigeage disjonctif. Ces filtres constituent une généralisation naturelle des filtres linéaires, en ce sens qu'ils n'utilisent que des hypothèses absolument nécessaires pour estimer une fonction non linéaire quelconque du signal sous-jacent, soit les lois bivariables. On peut donc imaginer qu'ils seront plus faciles à ajuster aux données expérimentales et plus robustes que des filtres plus puissants comme l'espérance conditionnelle qui repose sur des hypothèses très fortes mais qu'on ne sait pas tester avec les seules données. Le filtre du krigeage disjonctif, comme les autres filtres non linéaires, n'est pas adapté à tous les problèmes de filtrage. Par exemple, il ne semble pas que les problèmes de classification puissent se résoudre facilement avec un filtrage disjonctif, et l'approche des champs markoviens, bien qu'elle ne soit pas très satisfaisante en termes des hypothèses utilisées, semble mieux adaptée à ces problèmes [4]. Le filtrage disjonctif a cependant un certain nombre d'applications potentielles, la plus évidente d'entre elle qui sera abordée dans ce travail étant l'estimation de la probabilité qu'un phénomène dépasse un seuil donné dans le voisinage d'un point x.

CHAPITRE I

LA RANDOMISATION

1.0. — INTRODUCTION

Si l'on bombarde un échantillon d'acier sur un certain nombre de points ou de pixels avec un faisceau d'électrons et qu'on compte le nombre de photons X qui sont émis avec une énergie fixe caractéristique d'un élément, alors pour chaque pixel le nombre de photons est modélisé par une variable de Poisson ayant pour paramètre une variable proportionnelle à la concentration de l'élément. La carte ou l'image qu'on observe donne une représentation de la répartition spatiale de la concentration de notre élément mais contient un bruit poissonien qui à chaque pixel dépend de la concentration vraie. Il est souvent intéressant d'essayer de filtrer cette image pour obtenir un meilleur estimateur de la concentration, ou d'essayer d'estimer la loi sous-jacente à partir de l'histogramme observé. Nous avons là des exemples de randomisation où l'on observe non pas les données sous-jacentes mais les données bruitées et où l'objectif est d'estimer certaines des caractéristiques des données sous-jacentes. Nous n'essayerons pas de traiter ces problèmes de façon générale, et nous nous limiterons en fait essentiellement à quelques problèmes liés au problème de la poissonisation, en notant cependant que les méthodes qui seront appliquées ici peuvent être utilisées pour d'autres types de randomisation.

L'exemple classique de randomisation est un signal observé avec un bruit additif et indépendant. Le filtre de Wiener a été développé pour estimer le signal dans ce cas [1]. Le krigeage linéaire, que nous appliquerons aux problèmes de randomisation dans le chapitre II, est formellement équivalent au filtre de Wiener dans le cas stationnaire. Classiquement le filtre de Wiener est mise en œuvre dans le domaine fréquentiel et utilise le spectre des signaux. Le krigeage linéaire est mise en œuvre dans le domaine spatial et utilise le variogramme ou, dans le cas stationnaire, la fonction de covariance. Le variogramme est une fonction qui présente un intérêt en soi, parce qu'il peut donner des renseignements sur la structure de l'image.

Nous avons un autre exemple de randomisation quand les données réelles sont des taux $\varrho(x)$, $0 \leq \varrho(x) \leq 1$, comme par exemple le pourcentage théorique d'une

population au point x ayant une propriété donnée. A un certain moment, la population au point x est de $N(x)$ (supposée connue), et l'on observe $Z(x)$ cas avec cette propriété. Le taux expérimental $r(x) = Z(x) / N(x)$ peut être considéré comme une randomisation de $\varrho(x)$. Dans certains cas, il sera raisonnable de modéliser $Z(x)$ comme une variable binomiale $Z(x) \sim B(N(x), \varrho(x))$. Cette randomisation peut être intéressante pour certains problèmes d'environnement (par exemple l'analyse structurale des données sur le cancer [2]).

Nous avons indiqué que le problème de randomisation le plus important ici était celui de la poissonisation dans le contexte des images microsondes obtenues sur des aciers. A la section suivante, nous donnerons une brêve description des données.

1.1. — LES DONNÉES

Comme pour la formation des flocons de neige, un acier ne se solidifie pas de manière homogène et on observe la formation de cristaux ou dendrites. Les premières dendrites se forment à l'extérieur du liquide et elles croissent rapidement vers l'intérieur (contre le flux de chaleur) d'une manière plus ou moins linéaire. On appelle cette zone de métal la zone basaltique. Ensuite le métal liquide au centre refroidit et atteint une température inférieure à la température critique de fusion. Il se forme alors des dendrites de type équiaxe qui ont une morphologie très différente de celles de la zone basaltique étant donné qu'elles croissent un peu dans tous les sens et ne présentent plus la structure linéaire de la zone basaltique.

Quand une dendrite se forme, certains éléments ont tendance à quitter le solide pour rester dans le liquide le plus longtemps possible. Pour d'autres éléments, c'est l'inverse qui se produit. Ceci nous donne une répartition non uniforme de la concentration chimique, qu'on appelle la microségrégation. A une échelle plus grande, on trouve aussi une ségrégation (macroségrégation) qui n'est plus liée cette fois aux structures dendritiques mais plutôt aux forces plus macroscopiques telles que la convection et la procédure de solidification choisie par le métallurgiste [3].

Pour étudier les propriétés des métaux produits dans certaines conditions, on utilise une microsonde ou une macrosonde afin d'obtenir des cartes chimiques pour quelques éléments d'un échantillon du matériau. Pour obtenir ces cartes, on bombarde l'échantillon avec un faisceau d'électrons sur un certain nombre de pixels (souvent $(256)^2$ ou $(512)^2$) et on compte le nombre de photons qui ont

l'énergie caractéristique de l'élément étudié. Malheureusement le nombre de photons ne donne pas une mesure exacte de la concentration vraie au pixel x, mais c'est une variable de Poisson de paramètre proportionnelle à la concentration.

Sur la figure 1.1 nous voyons un exemple de carte de concentration de l'élément Mn faite lors d'une étude de microségrégation. La taille de l'image est de 512 × 512 et la taille réelle est de 1.536 mm, soit 3 μm par pixel.

Cet élément est fortement ségrégé en ce sens qu'il y a des poches de concentration très fortes entre les dendrites. Sur la figure 1.2, nous avons le même élément mais ici avec une coupe dans une autre partie de l'échantillon. On voit que la concentration de l'élément est beaucoup plus diffuse, ce qui nous montre que la ségrégation n'est pas constante le long de l'échantillon.

Sur la figure 1.3, on voit une étude de macroségrégation (image $(512)^2$, taille réelle 12.7 cm). Ici on voit clairement la zone équiaxe à gauche, la zone basaltique à droite et des Vés ségrégés autour du centre de la plaque.

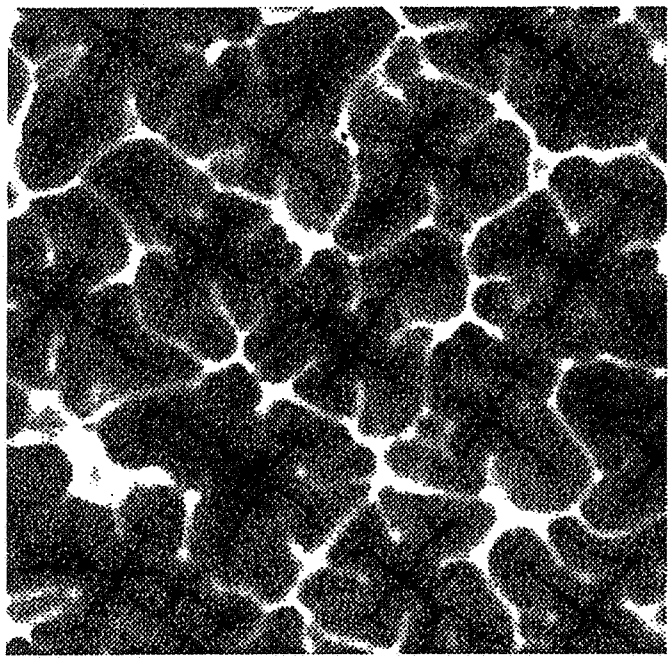


Figure 1.1 Microségrégation. 512 x 512 Carte de concentration de Mn . Taille réelle de l'image 1.536 mm

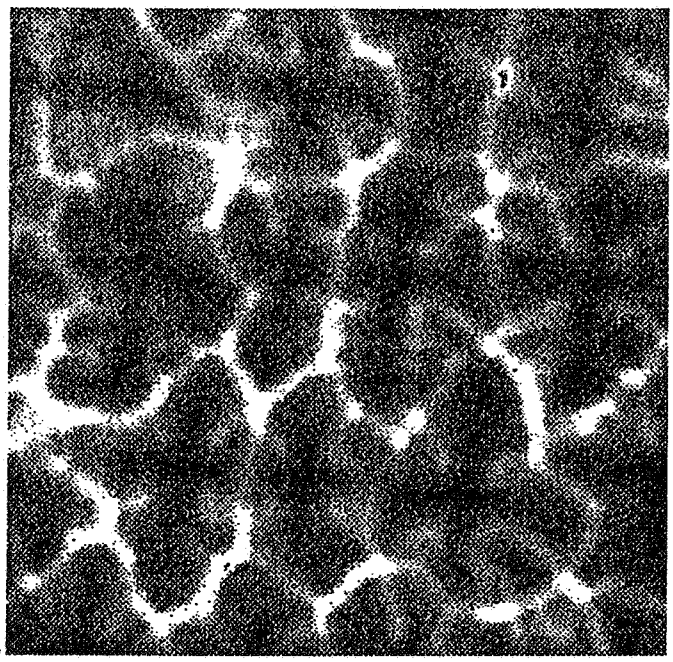


Figure 1.2 Microségrégation. 512 x 512 Carte de concentration de Mn . Taille réelle de l'image 1.536 mm

Pour cette étude, il peut être intéressant de séparer les structures de courte portée, qui sont de la taille de quelques amas de dendrites, de celles ayant une plus grande variabilité. Pour cela, il faut trouver une décomposition du variogramme selon des structures élémentaires, puis estimer ces dernières séparément. Nous aborderons cette procédure au chapitre II.

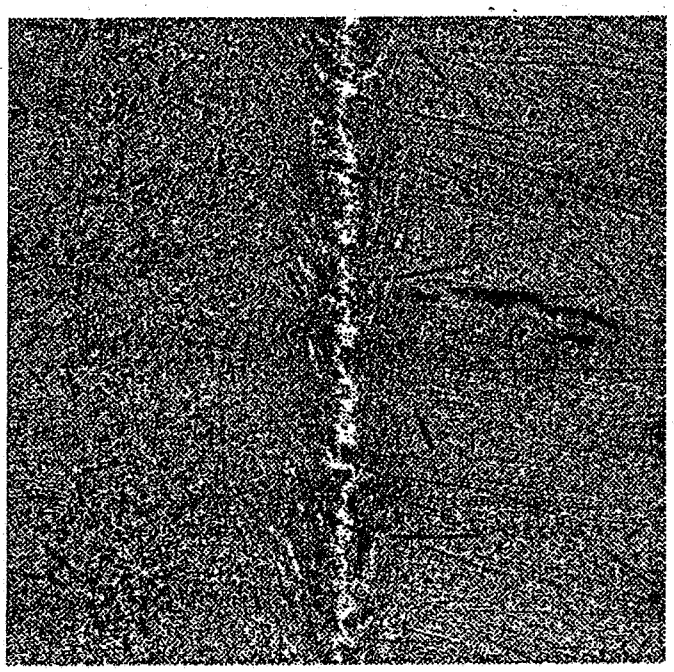


Figure 1.3 Macroségrégation. 512 x 512
Carte de concentration de Mn. Taille
réelle de l'image 12.288 mm

Jusqu'à présent, nous avons vu qu'il y avait un intérêt manifeste à avoir une méthode pour le filtrage des images bruitées, et qu'on pouvait être intéressé par une estimation individuelle de chaque structure. Mais il est aussi d'un intérêt capital d'avoir un estimateur du taux de ségrégation. Pour un taux de ségrégation global, il faut connaître l'histogramme sous-jacent à partir duquel on pourra calculer le nombre de valeurs élevées et faibles. Pour un taux de ségrégation local, on est encore plus exigeant, et il faut connaître l'histogramme local. On traitera de l'estimation de l'histogramme global au chapitre III, et on étudiera une méthode d'estimation de l'histogramme local au chapitre VI.

1.2. — LA RANDOMISATION

Supposons qu'on ait une variable aléatoire X_θ dépendant d'un paramètre θ. Soit $F(x,\theta)$ sa fonction de répartition

$$P\,[X_\theta < x] = F(x,\theta)$$

et

$$\lim_{x\to\infty} F(x,\ \theta) = 1\,,\ \lim_{x\to-\infty} F(x,\ \theta) = 0 \qquad \forall\ \theta$$

et $F(x,\theta)$ continue à gauche, non-décroissant.

Si Y est une autre variable aléatoire avec la fonction de répartition F^Y, et la mesure associée μ^Y, $\mu^Y(]a,b]) = F^Y(b) - F^Y(a)$, on dit que Y est randomisée par X_θ, et qu'elle donne une nouvelle variable aléatoire Z de loi F^Z si

$$F^Z(x) = \int F(x,\theta)\ d\mu^Y(\theta) \tag{1.1}$$

Intuitivement Z est la variable aléatoire X_Y c'est-à-dire X avec pour paramètre aléatoire Y, et l'équation (1.1) est une conséquence du théorème de probabilité totale.

La loi $F(x,\theta)$ est la loi conditionnelle de Z connaissant Y et on l'écrit parfois $F^{Z|Y}$.

Si $F(x,\theta)$ a une densité $f(x,\theta) = \dfrac{\partial F}{\partial x}(x,\theta)$ pour chaque θ, alors $G(x)$ a une densité $g(x)$ et

$$g(x) = \int f(x,\theta)\ d\mu(\theta)$$

Si $F(x,\theta)$ est une loi de probabilité discrète qui charge les entiers avec la probabilité $p(n,\theta)$, alors Z est discrète avec des probabilités $q(n)$

$$q(n) = \int p(n,\theta)\ d\mu(\theta)$$

Avant de passer aux exemples, notons la définition de l'espérance conditionnelle. Si $F^{Z|Y}(x)$ est la loi conditionnelle de Z connaissant Y, alors l'espérance conditionnelle de Z connaissant Y est

$$\mathrm{E}\,[Z|Y] = \int x\, F^{Z|Y}(dx)$$

et on a la relation importante

$$\mathrm{E}\big[E[Z|Y]\big] = E[Z]$$

Pour une définition rigoureuse de la randomisation et d'autres propriétés utiles, voir [4].

EXEMPLES :

1. — Supposons que ϵ soit une variable aléatoire de loi de probabilité $F(x)$ et Y une variable aléatoire de loi F^Y et de mesure μ^Y, indépendante de ϵ. Alors

$$Z = Y + \epsilon$$

est une randomisation de Y de loi F^Z donnée par

$$F^Z(x) = \int F(x-\theta)\, d\mu^Y(\theta)$$

Si ϵ a une densité, alors Z a une densité et

$$f^Z(x) = \int f(x-\theta)\, d\mu^Y(\theta)$$

Par symétrie, il en est de même si Y a une densité.

2. — L'exemple le plus intéressant pour nous est quand la loi de randomisation $p(n,\theta)$ est poissonienne

$$p(n,\theta) = e^{-\theta}\,\frac{\theta^n}{n!}$$

et on trouve

$$q(n) = \int e^{-\theta}\,\frac{\theta^n}{n!}\, d\mu(\theta)$$

On appelle $q(n)$ la loi poissonisée de $p(n,\theta)$.

3. — Si ϱ est une variable aléatoire qui prend des valeurs entre 0 et 1, et si N est un entier positif, on peut considérer une randomisation de ϱ en prenant

une variable binomiale des paramètres N et ϱ . Soit μ la mesure de probabilité associée à ϱ . Alors la loi de $Z \sim B(N, \varrho)$ est discrète et prend des valeurs entre 0 et N avec la probabilité

$$q(n) = \binom{N}{k} \int_0^1 \varrho^k \, (1-\varrho)^{n-k} \, d\mu(\varrho)$$

Donc $\frac{Z}{N}$ est une variable aléatoire entre 0 et 1 qui randomise ϱ et est telle que $E[\frac{Z}{N}] = \varrho$.

4. — Dans le dernier exemple, on peut considérer N comme une variable aléatoire au lieu de ϱ . Z est donc la variable aléatoire discrète avec la loi

$$q(k) = \sum_{n=0}^{\infty} \binom{n}{k} \varrho^k (1-\varrho)^{n-k} \, r_n$$

où $r_n = P[N = n]$. Cette randomisation est connue sous le nom de décimation de N.

1.3. — LES FONCTIONS ALÉATOIRES

Comme le filtrage d'une image bruitée ne peut pas être adapté de manière indépendante pour chaque pixel x (on ne dispose pas de suffisamment d'information pour pouvoir caractériser le signal et le bruit à une échelle très locale), on est obligé de chercher un filtre qui soit adapté "en moyenne" sur l'image ou sur une région de l'image. On voit donc qu'il y a un passage, peu précis et implicite, à un aspect statistique ou probabiliste avec tout type de filtrage. Notre filtre sera bon "en moyenne", mais on ne pourra pas exclure que localement les résultats ne soient pas bons. Un moyen d'expliciter cette hypothèse consiste à considérer l'image réelle ou sous-jacente comme une fonction aléatoire et à l'observer comme une randomisation de l'image réelle.

On dit que $Y(x)$ est une fonction aléatoire d'ordre deux sur $\mathbb{R}^2$ si pour chaque $x \in \mathbb{R}^2$, $Y(x)$ est une variable aléatoire avec une espérance et une variance finies.

La fonction aléatoire est stationnaire d'ordre deux si

1) $M = \mathrm{E}[Y(x)]$ est constant

2) $C(h) = \mathrm{Cov}[Y(x),\ Y(x+h)]$ dépend uniquement de h et non de x.

Le variogramme est défini par $\gamma(x,\ x+h) = \frac{1}{2}\ \mathrm{Var}\ (Z(x+h) - Z(x))$. Dans le cas stationnaire d'ordre deux, le variogramme ne dépend que de h, et on trouve que

$$\gamma(h) = C(0) - C(h)\,.$$

En particulier, le variogramme de la fonction aléatoire stationnaire est borné.

Les images que nous allons étudier ne semblent pas stationnaires d'ordre deux, mais en fait pour effecteur un filtrage par krigeage, nous n'avons pas vraiment besoin de stationnarité. L'hypothèse plus faible de stationnarité locale est suffisante [5], [6]. En effet, pour effectuer un filtrage en un point x, on n'utilise pas généralement l'image entière mais seulement un voisinage V autour de x, et il suffit que la moyenne soit à peu près constante à l'échelle de V et que le variogramme dépende uniquement de la distance entre les points, c'est-à-dire de h, à l'échelle de V. Le variogramme d'une fonction aléatoire localement stationnaire aura en général un palier (ou plutôt un changement de pente) associé à des structures à la petite échelle qui nous intéresse pour le filtrage.

Pour les cas qui sont vraiment non stationnaires, on utilise des modèles de fonctions aléatoires intrinsèques d'ordre k (FAI-k). Dans le cas le plus simple, la FAI-0, $Y(x)$ n'est pas stationnaire, mais les incréments $(Y(x+h) - Y(x))$, eux, sont stationnaires. Par conséquent on peut travailler avec le variogramme $\gamma(h)$, bien que la fonction de covariance ne soit pas stationnaire. Evidemment les combinaisons linéaires des données $\Sigma\ \lambda_i\ Y_i$ ne sont généralement pas stationnaires, mais on peut démontrer [6] que quand $\Sigma\ \lambda_i = 0$, alors $\Sigma\ \lambda_i\ Y_i$ est stationnaire, et on peut calculer sa variance à l'aide du variogramme. Dans ce cas, on dit que $\{\lambda_i\}$ est une combinaison linéaire autorisée. Le mouvement brownien est un exemple de processus très bien modélisé avec une FAI-0 de variogramme $\gamma(h) = |h|$. Plus généralement, si les incréments d'ordre k sont stationnaires, on dit que Y est une FAI-k. Ici la famille des combinaisons linéaires pour lesquelles on peut calculer les variances sont encore plus réduites. Il faut que

$$\sum \lambda_i\ f^{\ell}(x_i) = 0 \qquad |\ell| = 0,\ldots,k$$

où les f^ℓ sont des polynômes usuels. Donc pour une FAI-1, il faut que $\Sigma\, \lambda_i = 0$ et $\Sigma\, \lambda_i\, x_i = 0$ dans une dimension, et que de plus $\Sigma\, \lambda_i\, y_i = 0$ dans deux dimensions. Pour les FAI-k, les incréments $(Y(x+h) - Y(x))$ ne sont plus stationnaires, mais il existe une fonction, qu'on appelle la covariance généralisée $K(h)$, telle que si $\{\lambda_i\}$ est une combinaison linéaire autorisée, alors

$$\text{Var}\left(\sum \lambda_i\, Y(x_i)\right) = \sum \lambda_i\, \lambda_j\, K(x_i - x_j)$$

Pour une introduction aux FAI-k, voir [7]. Leur définition rigoureuse et une étude de leurs propriétés sont proposées dans [8].

Pour cette étude, nous n'avons pas besoin des FAI-k pour modéliser les images. Elles se modélisent facilement avec des fonctions aléatoires localement stationnaires.

Puisque dans la pratique on traite assez souvent des données multivariables (notamment lors d'une étude de solidification), nous rappelons ici la définition du variogramme croisé qui joue le rôle de fonction structurale entre deux éléments. Si les deux éléments sont modélisés par des fonctions aléatoires localement stationnaires $Y_i(x)$ et $Y_j(x)$ alors le variogramme croisé $\gamma_{ij}(h)$ est défini par

$$\gamma_{ij}(h) = \tfrac{1}{2}\, \mathrm{E}\left[(Y_i(x+h) - Y_i(x))\,(Y_j(x+h) - Y_j(x))\right]$$

On peut maintenant écrire les trois problèmes qui nous intéressent, et qui ont déjà été décrits dans § 1.1, en termes de fonctions aléatoires. Soit $Y(x)$ la fonction aléatoire sous-jacente avec laquelle on modélise l'image sous-jacente. A chaque pixel x, on observe une occurrence de la variable aléatoire $Z(x)$ qui est la randomisée de $Y(x)$. Le premier problème est le suivant. Ayant une réalisation de la fonction aléatoire $Z(x)$, est-ce qu'on peut estimer la réalisation correspondante de $Y(x)$? Pour le deuxième problème, on admet que $Y(x)$ a une loi stationnaire donnée par sa mesure de probabilité μ^Y. On veut estimer μ^Y connaissant seulement un histogramme expérimental de la loi de Z. Donc si

$$F^Z(x) = \int F(x,\theta)\, d\mu^Y(\theta)$$

et si nous n'avons qu'un histogramme expérimental pour $Z(x)$, par exemple l'histogramme de l'image bruitée $g(x)$, est-ce qu'on peut alors estimer μ^Y à partir de $g(x)$? La réponse est en général négative, mais nous verrons que dans certains

cas, dont la poissonisation, on peut obtenir des pseudo-solutions ou solutions régularisées, et que la méthode utilisée est équivalente à un krigeage. Le troisième problème consiste à estimer l'histogramme local de Y connaissant l'image bruitée $Z(x)$. Autrement dit, dans une région V, on veut connaître la probabilité que $Y(x)$ dépasse un certain seuil, $P[Y(x) > y | x \in V]$. Ce problème est traité par la technique du krigeage disjonctif.

Pour la mise en oeuvre d'un krigeage linéaire, il est nécessaire de connaître les fonction de covariance $\mathrm{Cov}(Z(x), Z(y))$, $\mathrm{Cov}(Z(x), Y(y))$ et $\mathrm{Cov}(Y(x), Y(y))$. A la section suivante, nous discuterons du lien existant entre covariances observées et covariances sous-jacentes.

1.4. — COVARIANCES

Soient $Y_1(x)$ et $Y_2(x)$ deux fonctions aléatoires sous-jacentes et $Z_1(x)$, $Z_2(x)$ leurs randomisées. Nous faisons deux hypothèses sur la forme de la randomisation

$$1)\ \mathrm{E}\ [Z_i(x)\ Z_j(y) | Y_i(x)\ Y_j(y)] = \mathrm{E}\ [Z_i(x) | Y_i(x)\ Y_j(y)]\ \mathrm{E}\ [Z_j(y) | Y_i(x)\ Y_j(y)] \quad \text{si} \quad x \neq y \quad \text{ou} \quad i \neq j \qquad (1.1)$$

$$2)\ \mathrm{E}\ [Zi(x) | Y_i(x)\ Y_j(y)] = \mathrm{E}\ [Z_i(x) | Y_i(x)] \qquad \forall\ i,j$$

Intuitivement, la première hypothèse demande que les bruits aux deux points différents x et y soient indépendants. La deuxième hypothèse demande que le bruit en x dépende uniquement de la valeur de $Y(x)$, et conditionnellement soit indépendant des autres points $Y(y)$, $y \neq x$. Le lemme classique suivant est très utile pour le calcul des fonctions de covariance observées.

LEMME 1.1.

1) Soit $Y(x)$ une fonction aléatoire et $Z(x)$ sa randomisée. Alors

$$\mathrm{Var}\ Z(x) = \mathrm{E}[\mathrm{Var}(Z(x) | Y(x))] + \mathrm{Var}\ (\mathrm{E}[Z(x) | Y(x)])$$

2) Si les hypothèses (1.1) sont valables

$$\mathrm{Cov}\ (Z(x),\ Z(y)) = \mathrm{Cov}\ (\mathrm{E}[Z(x) | Y(x)],\ \mathrm{E}[Z(y) | Y(y)]) \qquad x \neq y$$

3) Si l'on a deux fonctions aléatoires $Y_i(x)$ et $Y_j(x)$ et si $Z_i(x)$ et $Z_j(x)$ sont leurs randomisées, alors si (1.1) est vrai

$$\mathrm{Cov}\big(Z_i(x), Z_j(x)\big) = \mathrm{Cov}\big(\mathrm{E}[Z_i(x)\,|\,Y_i(x)], \mathrm{E}[Z_i(x), Y_j(x)]\big) \quad \forall\ x,\ y$$

DÉMONSTRATION. – On sait que

$$\mathrm{E}[Z] = \mathrm{E}[\mathrm{E}[Z|Y]]$$

et donc

$$\mathrm{E}[Z^2] = \mathrm{E}[\mathrm{E}[Z^2|Y]] = \mathrm{E}\ [\mathrm{Var}(Z|Y) + \mathrm{E}[Z|Y]^2]$$

Alors

$$\mathrm{Var}\ Z = \mathrm{E}[\mathrm{Var}\ (Z|Y)] + \Big(\mathrm{E}\big[\mathrm{E}[Z|Y]^2\big] - \mathrm{E}[\mathrm{E}[Z|Y]]^2\Big)$$

et

$$\mathrm{Var}\ Z = \mathrm{E}[\mathrm{Var}(Z|Y)] + \mathrm{Var}\ (\mathrm{E}[Z|Y])$$

ii) $\mathrm{Cov}\ [Z(x), Z(y)]$

$= \mathrm{E}[Z(x)\ Z(y)] - \mathrm{E}[Z(x)]\ \mathrm{E}[Z(y)]$

$= \mathrm{E}[\mathrm{E}[Z(x)|Y(x)]\ \mathrm{E}[Z(y)|Y(y)]] - \mathrm{E}[\mathrm{E}[Z(x)|Y(x)]]\ \mathrm{E}[\mathrm{E}[Z(y)|Y(y)]]$ si $x \neq y$

$= \mathrm{Cov}(\mathrm{E}[Z(x)|Y(x)],\ \mathrm{E}[Z(y)|Y(y)])$ si $x \neq y$

et iii) se démontre de la même manière. ■

NOTA 1. – De la même manière, on démontre que $\mathrm{Cov}(Z(x),\ Y(y)) = \mathrm{Cov}(\mathrm{E}[Z(x)|Y(x)],\ Y(y)) \quad \forall\ x,y$.

NOTA 2. – Si de plus $\mathrm{E}[\ Z(x)\ |\ Y(x)\ ] = Y(x)$ et on appelle $f(x) = \mathrm{E}[\mathrm{Var}(Z(x)|Y(x))]$, alors les deux premières propositions du lemme donnent

$$\mathrm{Cov}\big(Z(x), Z(y)\big) = \mathrm{Cov}\big(Y(x), Y(y)\big) + f(x)\delta_{xy}$$

Appliquons maintenant ce lemme aux exemples déjà considérés.

1. — BRUIT ADDITIF ET INDÉPENDANT

Il existe beaucoup d'applications dans lesquelles on peut considérer le signal $Y(x)$ comme contaminé par une erreur ou un bruit indépendant $\epsilon\,(x)$. Si $y(x)$ et

$\epsilon(x)$ sont modélisés par des fonctions aléatoires localement stationnaires, alors l'observé $Z(x)$ donné par

$$Z(x) = Y(x) + \epsilon(x)$$

est aussi localement stationnaire. Quand le bruit est de moyenne nulle, alors

$$\mathrm{E}[Z] = \mathrm{E}[\mathrm{E}[Z|Y]] = \mathrm{E}[Y]$$

Par le lemme 1.1, ou simplement par indépendance

$$\mathrm{Var}\ Z = \mathrm{Var}\ Y + \mathrm{Var}\ \epsilon$$

et

$$\mathrm{Cov}\ (Z(x), Z(y)) = \mathrm{Cov}\ (Y(x), Y(y)) \qquad \text{pour } x \neq y.$$

2. — POISSONISATION

$Y(x)$ est maintenant la variable sous-jacente, et l'on observe $Z(x)$ qui est une variable de Poisson avec comme paramètre $Y(x)$, donc $(Z|Y)$ est une variable de Poisson de paramètre Y et l'on a

$$\mathrm{E}[Z|Y] = Y$$

d'où

$$\mathrm{E}[Z] = \mathrm{E}[Y]$$

Il est vrai aussi que $\mathrm{Var}\ (Z|Y) = Y$ et le lemme 1.1 donne

$$\mathrm{Var}\ Z = \mathrm{Var}\ Y + \mathrm{E}[Y]$$

et

$$\mathrm{Cov}\ (Z(x), Z(y)) = \mathrm{Cov}\ (Y(x), Y(y)) \qquad \text{si } x \neq y.$$

Comme pour le bruit indépendant, la seule différence entre la covariance sous-jacente et la covariance observée est un "saut" à l'origine, ou un effet de pépite. Ici l'effet de pépite est fixe, et il est égal à $\mathrm{E}[Y] = m$. Le variogramme de Z est

$$\gamma^Z(h) = \gamma^Y(h) + m$$

3. — PROBLÈME DE TAUX

Le taux $\varrho(x)$ est modélisé comme une fonction aléatoire. Pour chaque point x, on connaît la population $N(x)$, et on observe $Z(x)$ cas favorables, avec $Z(x) \sim B(N(x), \varrho(x))$.

Alors $E[Z(x)|\varrho(x)] = N(x)\,\varrho(x)$ et $\text{Var}\,(Z(x)|\varrho(x)) = N(x)\,(\varrho(x)(1-\varrho(x))$.

Si l'on considère alors le taux expérimental $r(x) = \dfrac{Z(x)}{N(x)}$, on trouve que

$$E[r(x)|\varrho(x)] = \varrho(x)$$

et

$$\text{Cov}\,(r(x),\ r(y)) = \text{Cov}\,(\varrho(x),\ \varrho(y)) \qquad \text{pour } x \neq y$$

et

$$\text{Var}\,(r(x)) = \frac{1}{N(x)}\ E\ [\varrho(x)(1-\varrho(x))] + \text{Var}\,(\varrho(x))\,.$$

Notons que si $N(x)$ n'est pas constant, $r(x)$ n'est pas une fonction aléatoire stationnaire, mais si l'on connaît la fonction de covariance de ϱ, et si $N(x)$ est connue aussi, alors on connaît la fonction de covariance non stationnaire de $r(x)$.

J'ignore si l'autre randomisation par un binomial, celle de N, a des applications dans un problème de randomisation. Par contre elle a une application en géostatistique minière dans le contexte d'un changement de support. Si un gisement de pierres précieuses est divisé en régions de taille V, et que dans chaque région, on échantillonne un morceau de taille v, alors le nombre de pierres n_i dans l'échantillon v_i est une décimation du nombre de pierres N_i dans V_i. Il existe même un modèle isofactoriel pour les lois bivariables avec la décimation comme changement de support isofactoriel (modèle binomial négatif). Nous en discuterons brièvement au chapitre VI.

Pour des FAI-k, la situation est un petit peu plus compliquée. Par exemple dans le cas d'un bruit poissonien, même quand $Y(x)$ est une FAI-0, il n'est pas vrai que $Z(x)$ soit une FAI-0. Soit $m(x) = E[Y(x)]$. Alors

$$\gamma^Z(x,y) = \frac{1}{2}\text{Var}\,(Z(y)-Z(x)) = \gamma^Y(x-y) + \frac{1}{2}\,(m(x)+m(y))$$

et par conséquent γ^Z dépend de x et de y. Malheureusement, la fonction $m(x)$ n'est pas connue, et donc $\gamma^Z(x,y)$ ne peut pas être connue exactement. Mais si l'on connait $\gamma^Y(x,y)$, on peut estimer $m(x) = E[Y(x)] = E[Z(x)]$ par moindres carrés ou par une autre méthode, pour obtenir une estimation non biaisée de $\gamma^Z(x,y)$. Plus généralement, pour une randomisation telle que $E[\ Z(x)\ |\ Y(x)\ ] = Y(x)$, on sait que (nota 2 après lemme 1.1)

$$\text{Cov}\,(Z(x),\ Z(y)) = \text{Cov}\,(Y(x),\ Y(y)) + f(x)\ \delta_{xy}$$

est vraie indépendamment des questions de stationnarité, et donc si l'on peut considérer $Y(x)$ comme une FAI-k de covariance généralisée $K(h)$, la covariance généralisée non stationnaire pour $Z(x)$ sera alors

$$K^Z(x,y) = K^Y(x-y) + f(x)\ \delta_{xy}$$

Nous n'insisterons pas sur ce détail parce que toutes les applications du filtrage qui nous concernent ici sont réalisables à l'aide d'une covariance localement stationnaire.

1.4.1. — Estimation de la fonction de covariance

L'estimateur évident de la fonction de covariance d'une fonction aléatoire stationnaire d'ordre 2, $Y(x)$ est donnée par

$$\hat{C}(h) = \frac{1}{N_h} \sum_{i \in S_h} Z(x_i)\ Z(x_i + h) - (\hat{m})^2$$

où S_h est l'ensemble des points x, tel que nous avons des données en x et en $x+h$, où $N_h = \text{Card}\ (S_h)$ et où $\hat{m} = \frac{1}{N} \Sigma\ Z(x_i)$. Il est bien connu que cet estimateur est biaisé, dans le sens que $\text{E}[\hat{C}(h)] \neq C(h)$, mais il est vrai que le biais est peu important quand N_h est relativement grand. L'estimateur du variogramme

$$\hat{\gamma}(h) = \frac{1}{N_h} \sum_{x \in S_h} (Z(x+h) - Z(x))^2$$

est tel que $\text{E}[\hat{\gamma}(h)] = \gamma(h)$, même quand $Y(x)$ est une FAI-0.

Il est facile de démontrer (par les méthodes du lemme 1.1) que pour le problème de poissonisation

$$\text{E}[\hat{\gamma}^Z] = \gamma^Y(h) + m \qquad \text{avec } m = \text{E}[Y] = \text{E}[Z].$$

Donc, avec le variogramme expérimental de Z, $\hat{\gamma}^Z$, on peut obtenir un estimateur non biaisé de γ^Y, soit

$$\hat{\gamma}^Y = \hat{\gamma}^Z - \hat{m} \qquad \text{avec } \hat{m} = \sum_i \frac{Z(x_i)}{N}.$$

En général, on ne peut pas malheureusement obtenir un estimateur de γ^Y avec une seule acquisition de données bruitées. En particulier dans le cas du bruit additif indépendant de variance σ_ϵ^2 on ne peut pas obtenir un estimateur de γ^Y, sauf si – par exemple pour des raisons physiques – on peut dire que $Y(x)$ doit être une fonction aléatoire continue. Dans ce cas, l'effet de pépite apparent de $\hat{\gamma}^Z$ peut être considéré comme égal à σ_ϵ^2.

Pour le problème de randomisation du taux $\varrho\ (x)$ par un binomial, le problème est plus délicat. Après un peu de calcul, on peut démontrer que

$$\mathrm{E}\ [\hat{\gamma}_r(h)] = \gamma_\varrho(h) + \frac{\mathrm{E}\ [\varrho(x)(1-\varrho(x))]}{N_h}\left[\sum_{x\in S_h}\left(\frac{1}{N(x)} + \frac{1}{N(x+h)}\right)\right]$$

Ici le problème est que le deuxième terme dépend du moment d'ordre deux de $\varrho\ (x)$, qui est essentiellement le terme qu'on veut estimer. Il ne semble donc pas facile d'estimer le variogramme sous-jacent à l'aide d'une acquisition bruitée.

On a une solution au problème de l'estimation du variogramme sous-jacent ou de la covariance sous-jacente dès qu'on a deux acquisitions de données bruitées. Si l'on suppose que le bruit est indépendant pour chaque acquisition, alors les hypothèses (1.1) s'appliquent pour tout x et tout y et pas uniquement quand $x \neq y$. Il est facile de démontrer que

$$\gamma^Y(h) = \frac{1}{2}\ \mathrm{E}\left[(Z^1(x+h) - Z^1(x))(Z^2(x+h) - Z^2(h))\right]$$

et que le variogramme croisé expérimental est un estimateur non biaisé pour γ^Y. Puisque $C^Y(h) = C^Y(0) - \gamma^Y(h)$, nous avons un estimateur pour $C^Y(h)$. Finalement, nous avons déjà vu que

$$\mathrm{Cov}\ (Y(x),\ Z(x)) = \mathrm{Cov}\ (Y(x),\ Y(y))$$

quand $\mathrm{E}[Z|Y] = Y$ et nous avons donc un estimateur pour Cov $(Y(x), Z(x))$.

Avant d'examiner quelques exemples de variogrammes calculés sur des images bruitées, notons que les variogrammes expérimentaux doivent être ajustés par un modèle théorique. La raison en est qu'une fonction de covariance ne peut pas être une fonction quelconque, elle doit être positive en ce sens que pour chaque ensemble fini de points $\{x_i\}_{i=1}^n$ et chaque ensemble de pondérateurs $\{\lambda_i\}_{i=1}^n$,

$$\text{Var}\left(\sum \lambda_i\, Y(x_i)\right) = \sum \lambda_i\, \lambda_j\, C^Y(x_i - x_j) \geq 0$$

est positive. En fait, avec une application du krigeage en vue, nous admettons une hypothèse légèrement plus forte ; il faut que $C^Y(h)$ soit strictement positive

$$\sum \lambda_i\, \lambda_j\, C^Y(x_i - x_j) \geq 0$$

et soit nulle si et seulement si $\lambda_i = 0 \quad \forall\ i$.

De la même manière, le variogramme $\gamma^Y(h)$ doit être conditionnellement négatif défini. Pour des points $\{x_i\}_{i=1}^n$ et des pondérateurs $\{\lambda_i\}_{i=1}^n$ tels que $\Sigma\, \lambda_i = 0$, alors

$$-\sum \lambda_i\, \lambda_j\, \gamma(x_i - x_j) \geq 0$$

et nous demandons que ceci soit strict, c'est-à-dire qu'il soit nul si et seulement si $\lambda_i = 0 \quad \forall\ i$.

EXEMPLES

1. — Nous reprenons ici une des images de microségrégation que nous montrons à nouveau à la figure 1.4.

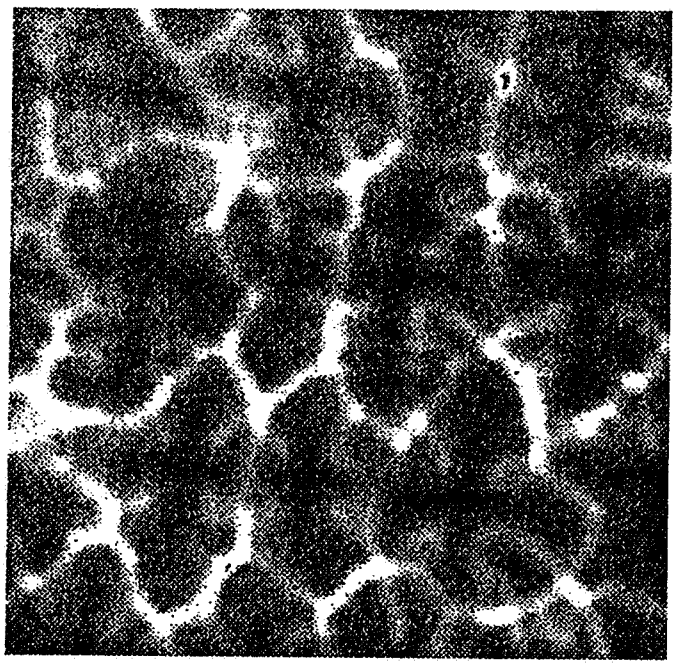


Figure 1.4 Microségrégation.

Son variogramme expérimental a été calculé dans les directions horizontale et verticale pour les 250 premiers pas (fig. 1.5).

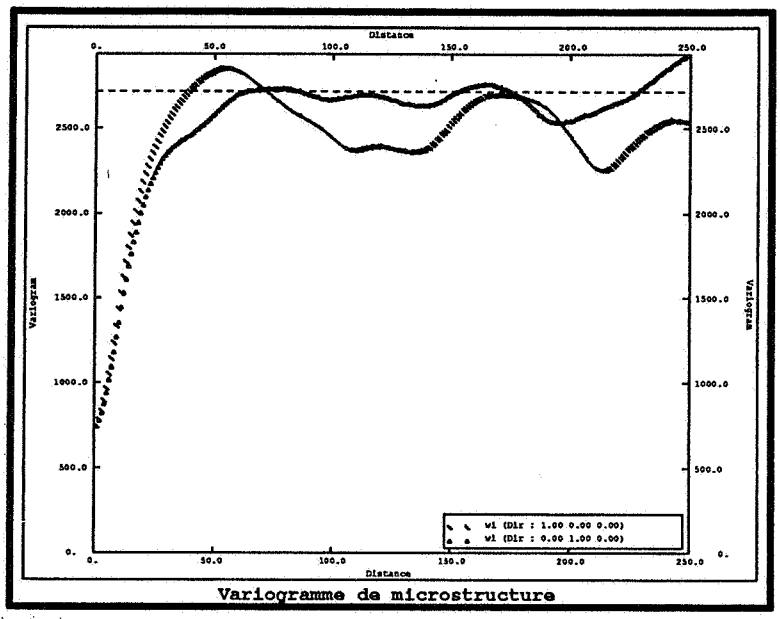


Figure 1.5 Variogramme de microségrégation

On voit que le variogramme monte très rapidement dans les deux directions, puis les deux directions divergent et commencent à onduler autour d'un palier. Ces ondulations sont liées à la taille des dendrites qui sont différentes suivant les deux directions sur cette image. Pour le filtrage, la seule partie du variogramme qui nous intéresse est celle qui se trouve près de l'origine (Fig. 1.6). A cette échelle on constate que l'image est à peu près isotrope et on voit qu'un modèle isotrope avec palier sera raisonnable. Nous faisons alors une hypothèse de stationnarité locale pour cette image. L'ajustement du variogramme se fait au moyen de deux modèles cubiques (qui sont différentiables à l'origine) d'une portée de 25 et 45 et un effet de pépite (voir fig. 1.7).

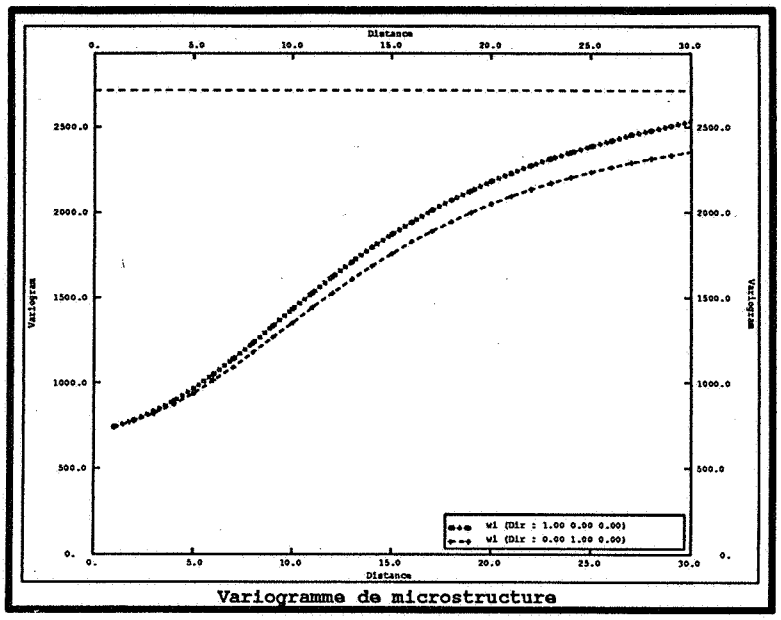


Figure 1.6 Variogramme de microségrégation

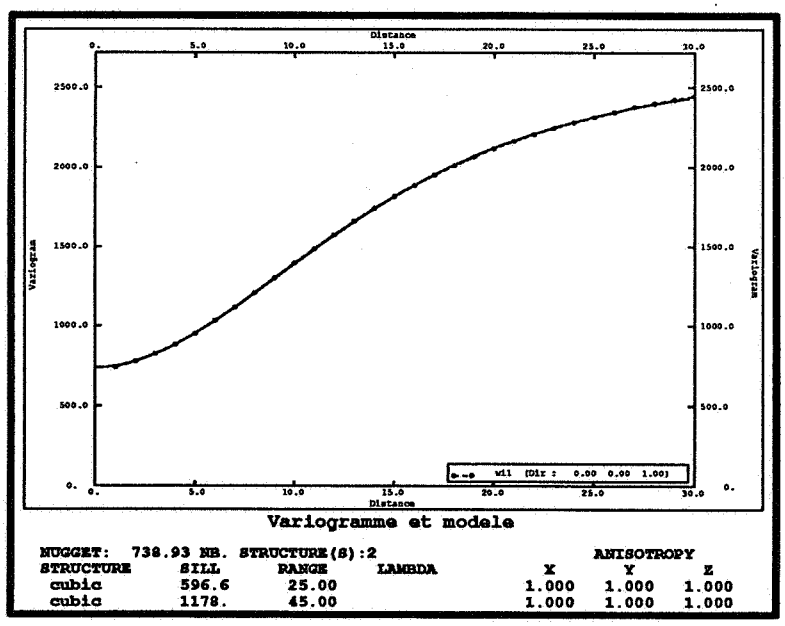


Figure 1.7 Modèle pour microségréation

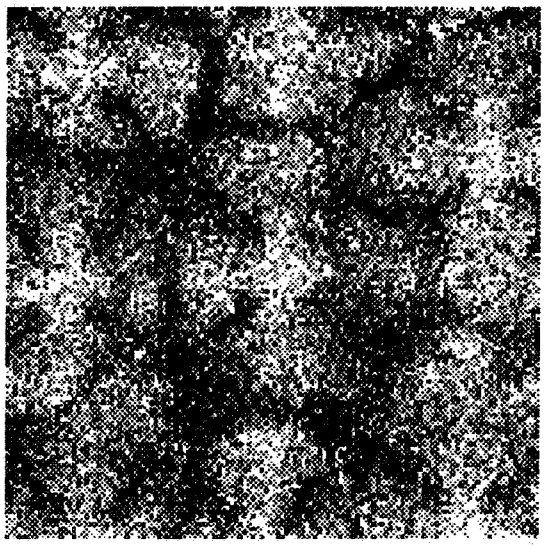


Figure 1.8 Cobalt

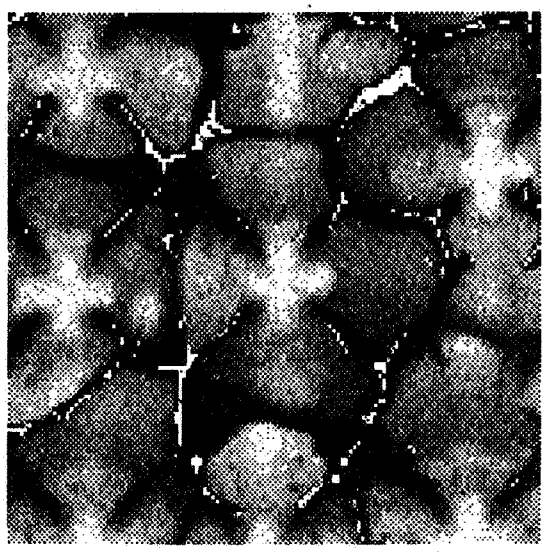


Figure 1.9 Tungstène

2. — Dans cet exemple, nous considérons une étude multivariable de microségrégation dans un superalliage [9]. La concentration pour les deux éléments Co et W est montrée aux figures 1.8 et 1.9. On remarque l'existence de petites phases (lignes noires ou blanches) qui ont des concentration chimiques très différentes de celle de la phase principale. Il est alors intéressant de masquer ces phases avant de calculer les variogrammes afin de travailler sur un champ plus homogène (fig. 1.10). Pour ces données, le bruit n'est plus poissonien parce qu'avec ce type de microscope à balayage, on retire du signal un bruit de fond, et donc ce qu'on observe est la différence de deux variables de Poisson de paramètres différents. De plus, les données sont regroupées dans des classes. On se trouve donc ici dans un cas où l'usage de deux acquisitions est nécessaire pour pouvoir calculer les variogrammes sous-jacents. Les résultats de ces calculs avec deux acquisitions sont montrés pour les trois éléments Co, W et Ti sur la figure 1.11.

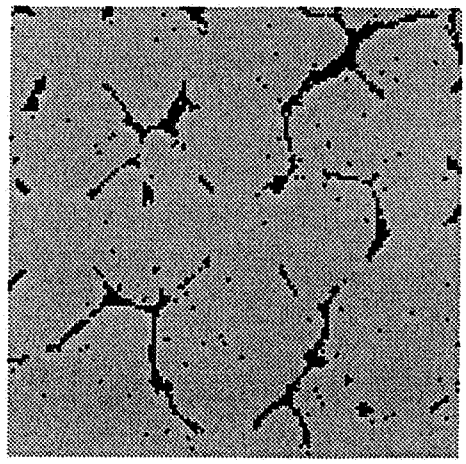


Figure 1.10 Le masque appliqué.

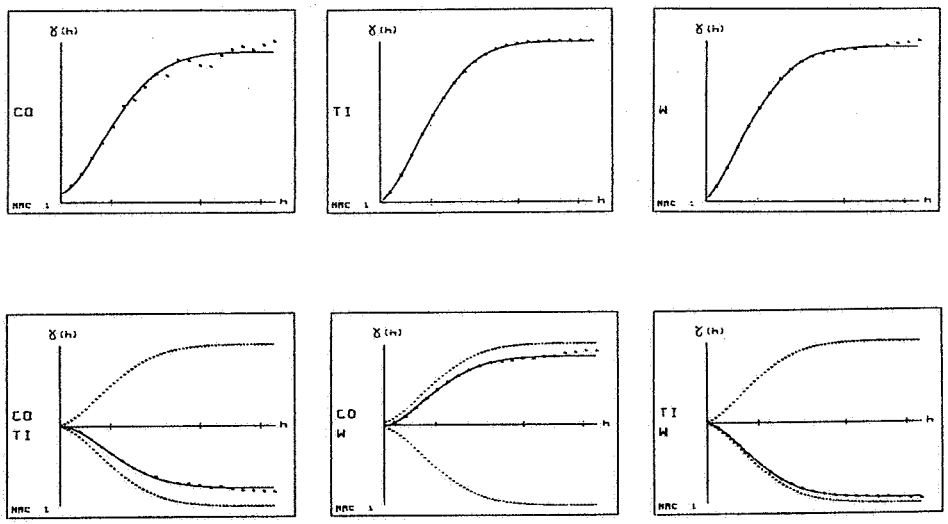


Figure 1.11 Les variogrammes et les variogrammes croisées pour Co, Ti et W.

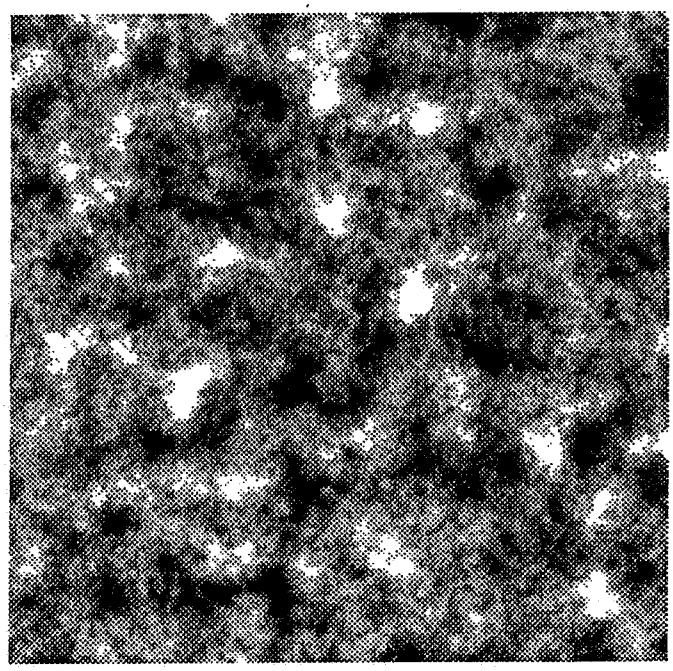


Figure 1.12 Image de taux ρ(x) simulée

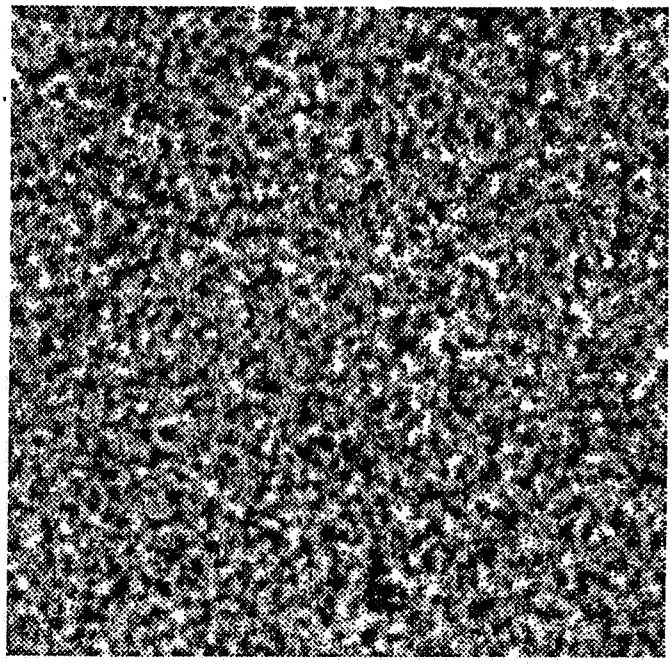


Figure 1.13 Image de "population", $N(x)$ simulée.

3. — Il est intéressant de considérer un exemple autre que la poissonisation. Sur la figure 1.12, on voit une simulation d'un taux régional $\varrho\ (x)$. Il s'agit en effet d'une simulation gaussienne avec un modèle sphérique d'une portée de 20 pixels. La moyenne est de 0.3 et l'écart-type de 0.06. Sur la figure 1.13, on voit une simulation de répartition de population $N(x)$ (sphérique, de portée = 5, moyenne = 100, écart-type = 20). On randomise maintenant $\varrho\ (x)$ par un binomial $Z(x) \backsim B(N(x),\ \varrho(x))$, et on trouve la carte de $Z(x)$ sur la figure 1.14. Son variogramme (fig. 1.15) a deux structures de portées 5 et 20. On voit donc que le variogramme des valeurs brutes est très influencé par la structure géométrique de la répartition de la population. Le taux expérimental $(Z(x))/(N(x))$ n'a pas de structure parasite à 5 pixels (fig. 1.16), mais en revanche il existe un effet de pépite difficile à estimer. Avec deux acquisitions du taux expérimental $(Z^1(x))/(N(x))$ et $(Z^2(x))/(N(x))$, on peut calculer leur variogramme croisé, et on trouve un estimateur satisfaisant pour le variogramme sous-jacent (fig. 1.17).

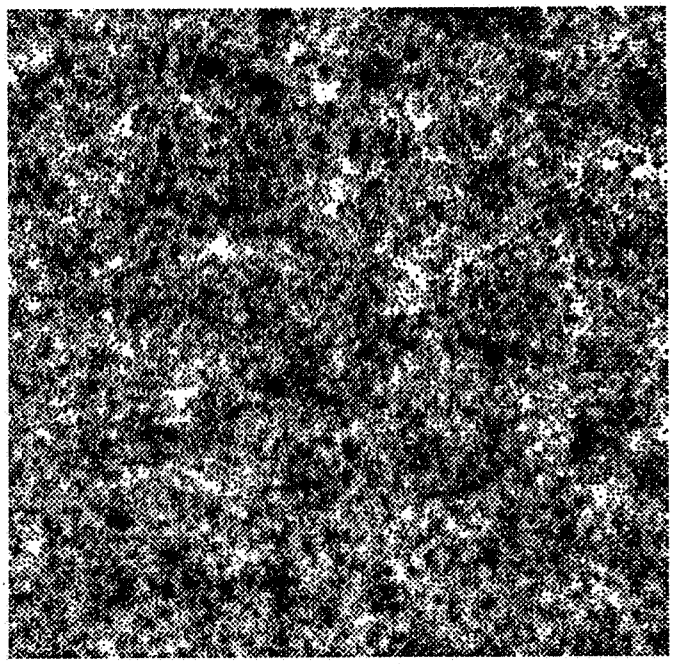


Figure 1.14 Image observée $Z(x)$

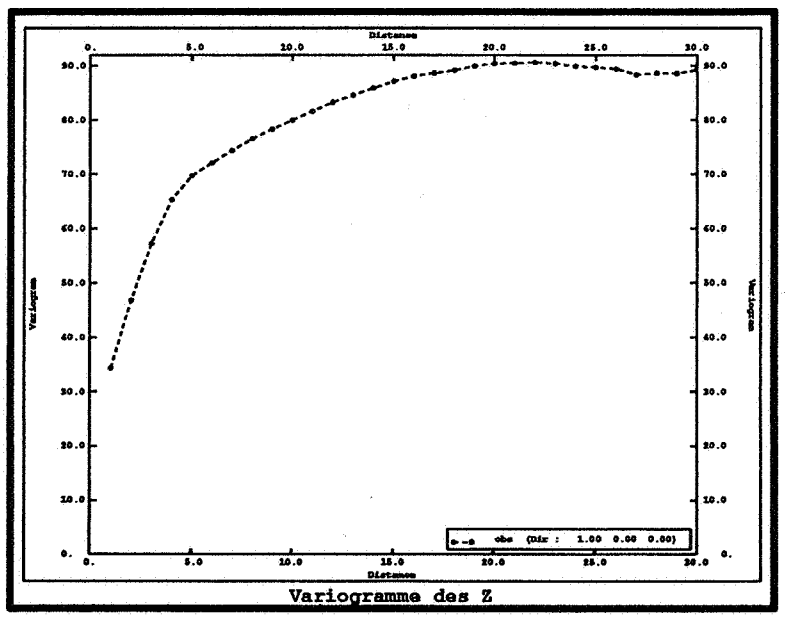


Figure 1.15 Variogramme des données observées $Z(x)$.

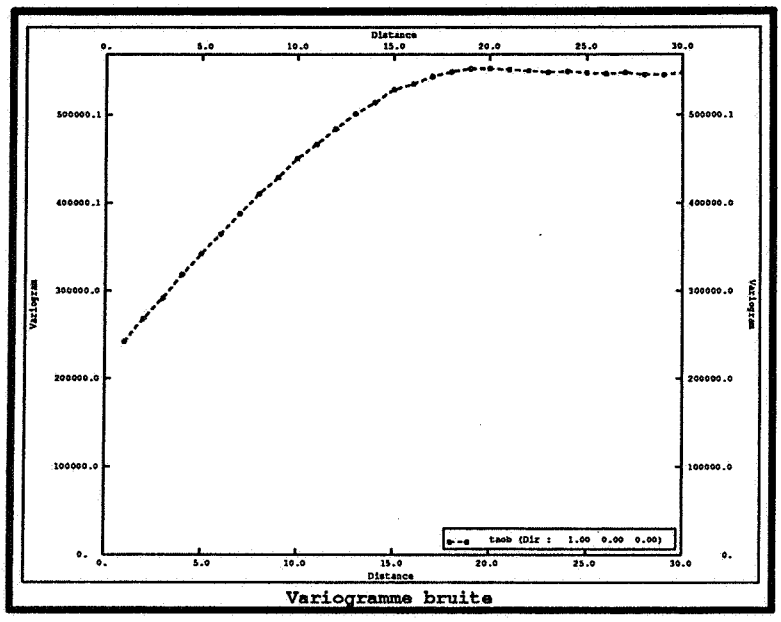


Figure 1.16 Variogramme de taux expérimental

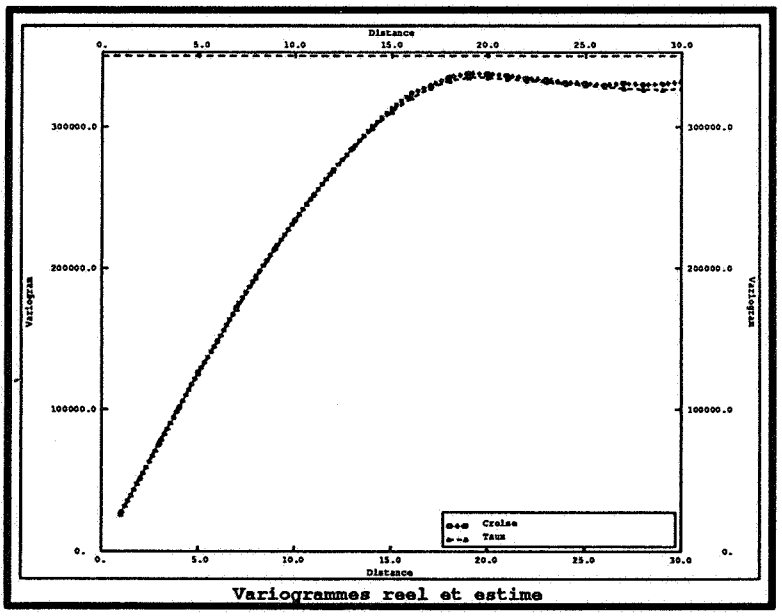


Figure 1.17 Variogramme croisé entre deux aquisitions de taux expérimental et variogramme de taux réel

1.5. — LES LOIS SOUS-JACENTES

En général, il n'est pas possible de trouver la loi sous-jacente $F^Y(x)$ connaissant la loi $F^Z(x)$. Mais dans le cas où il existe une solution unique, les méthodes des transformées sont souvent très utiles pour démontrer l'existence. Nous n'essayerons pas de donner une théorie générale, mais nous nous limiterons aux exemples traités jusqu'à présent.

EXEMPLES

1. — BRUIT ADDITIF ET INDÉPENDANT

Soit $Z(x) = Y(x) + \epsilon(x)$ avec ϵ indépendant de Y. Soit F^Z, F^Y et F^ϵ les lois de Z, Y et ϵ respectivement. Puisque Y et ϵ sont indépendants, F^Z est la convolution de Y et ϵ

$$F^Z = F^Y * F^\epsilon$$

et si l'on prend la transformée de Fourier

$$\varphi^Z = \varphi^Y \varphi^\epsilon$$

d'où on trouve $\varphi^Y = \dfrac{\varphi^Z}{\varphi^\epsilon}$ et la solution unique pour F^Y est la transformée inverse de $\dfrac{\varphi^Z}{\varphi^\epsilon}$.

2. — PROBLÈME DE DÉCIMATION

Nous avons une fonction aléatoire discrète $N(x)$ de loi stationnaire $p_N(i)$ et à chaque point x, on échantillonne $Z(x)$ selon une loi binomiale $Z(x) \backsim B(N(x), \varrho)$. Si la loi $p_Z(i)$ est connue, on peut trouver la loi de N, $p_N(i)$. En effet la fonction génératrice de p_Z est

$$G_Z(s) = \mathrm{E}[s^Z] = \mathrm{E}\left\{\mathrm{E}[s^Z|N]\right\} = \mathrm{E}\left[(q + ps)^N\right] = G_N(q + ps)$$

où $q = 1 - p$.

Quand $p \neq 0$, nous connaissons G_N sur l'intervalle non dégénéré [q,1]. Il suit alors d'un lemme classique de l'analyse complexe (voir l'énoncé dans § 4.1) que G_N a une extension unique sur [0,1] et par l'unicité des fonctions génératrices, il existe une loi unique p_N, telle que sa randomisée par un binomial est égale à p_Z.

3. — PROBLÈME DE TAUX

Pour la randomisation de $\varrho\,(x)$, il n'existe pas de théorème d'inversion de ce type, même dans le cas le plus simple, quand $N(x) = N$ est constant pour chaque x. En effet, pour N constant, la loi stationnaire des observées Z est donnée par

$$F^Z(k) = \binom{N}{k} \int_0^1 \varrho^k (1-\varrho)^{N-k}\, d\mu(\varrho)$$

Choisissons μ une mesure discrète qui charge les points 0, 1/3, 2/3 et 1 avec les poids μ_0, $\mu_{\frac{1}{3}}$, $\mu_{\frac{2}{3}}$ et μ_1, tel que $\mu_0 + \mu_{\frac{1}{3}} + \mu_{\frac{2}{3}} + \mu_1 = 1$. Alors si $N = 1$ nous avons $F^Z(0) = \mu_0 + \frac{2}{3}\mu_{\frac{1}{3}} + \frac{1}{3}\mu_{\frac{2}{3}}$ et $F^Z(1) = \frac{1}{3}\mu_{\frac{1}{3}} + \frac{2}{3}\mu_{\frac{2}{3}} + \mu_1$, un système de trois équations à quatre inconnues et par conséquent il existe un nombre infini de 'solutions' μ_0, $\mu_{\frac{1}{3}}$, $\mu_{\frac{2}{3}}$ et μ_1, avec les probabilités randomisées $F^Z(0)$ et $F^Z(1)$. Il n'existe donc pas de loi sous-jacente unique pour F^Z. Plus généralement, pour N fixe, si l'on choisit une mesure discrète qui charge les points $\{x_i\}_{i=1}^{N'}$, avec $N' > N + 2$, il existera toujours un nombre infini de "solutions" $\varrho\,(x)$ ayant pour randomisé F^Z parce que les équations linéaires qui définissent F^Z ne seront pas de plein rang. On voit ici donc un exemple d'une randomisation pour laquelle il n'existe pas de solution unique pour le problème de la restauration de l'histogramme.

4. — POISSONISATION

Considérons la fonction génératrice des données observées $Z(x)$, pour $0 \leq s \leq 1$, nous avons

$$G_Z(s) = \mathrm{E}[s^Z] = \mathrm{E}\big[\mathrm{E}[s^Z \mid Y]\big] = \mathrm{E}[e^{(s-1)Y}] = L_Y(1-s)$$

où L_Y est la transformée de Laplace de la variable sous-jacente Y. Donc la transformée de Laplace est connue sur [0,1] et on démontrera au chapitre IV que ceci est suffisant pour spécifier la transformée de Laplace partout et donc la loi de Y, $F^Y(x)$, est uniquement déterminée.

Nous avons vu que dans certains cas, la loi de la fonction randomisée suffit pour déterminer la loi sous-jacente. Dans la pratique, les choses se compliquent un peu. En général nous n'avons qu'un nombre fini de données randomisées et par conséquent, on ne connaît pas la loi randomisée mais seulement un estimateur de cette loi randomisée. Il se trouve que dans bon nombre de cas, même si la loi sous-jacente peut être déterminée univoquement par la loi randomisée, elle est très sensible aux petites variations. En particulier, il est souvent difficile d'estimer la loi sous-jacente à partir d'un échantillon tiré selon la loi randomisée. Nous aborderons ce problème au chapitre III en insistant plus particulièrement sur le problème de la poissonisation.

CHAPITRE II

LE FILTRAGE LINÉAIRE

2.0. — *INTRODUCTION*

Les filtres linéaires sont ceux qui cherchent un estimateur $Y^*(x)$ du signal $Y(x)$ en fonction des données $\{Z(x_i)\}$ de la forme suivante

$$Y^*(x) = \sum_a \lambda^a \, Z(x_a)$$

et le filtre du krigeage est le filtre linéaire optimal dans le sens que si l'on considère $Y(x)$ et $Z(x)$ comme des fonctions aléatoires stationnaires d'ordre deux, alors le filtre de krigeage choisit les $\{\lambda^a\}$ tels que $\mathrm{Var}[Y^*(x) - Y(x)]$ est un minimum. Ceci est essentiellement un problème de type moindres carrés et il se trouve qu'il est équivalent à l'espérance conditionnelle si $(Y(x), \{Z(x_a)\})$ est multigaussienne et si le bruit est additif et indépendant de Y. En dehors de ces conditions, le krigeage n'est plus équivalent à l'espérance conditionnelle, mais il satisfait toujours un critère d'optimalité, celui d'être le meilleur filtre linéaire. Ce critère en soi est assez fort et il se trouve que le krigeage donne de bons résultats dans bon nombre de situations. En effet, pour des images de type diffusif, il est assez difficile de trouver des cas où des techniques non-linéaires telles que le krigeage disjonctif ou l'espérance conditionnelle sont les mieux adaptées au problème du filtrage (Chapitre VI). Mais on note que l'intérêt pour le filtrage non linéaire ne se limite pas à la simple estimation du signal sous-jacent (ce qui peut assez souvent se faire tout aussi bien par krigeage linéaire), mais qu'il peut porter sur des problèmes explicitement non linéaires, tels que l'estimation d'un histogramme local.

Un des principes de modélisation en géostatistique, consiste à ne pas utiliser des modèles plus compliqués, et donc plus difficiles à ajuster aux données, qu'il ne faut pour résoudre le problème. Le filtre linéaire de krigeage dépend seulement des fonctions de covariance $\mathrm{Cov}(Z(x),Z(y))$, $\mathrm{Cov}(Z(x),Y(y))$ et $\mathrm{Cov}(Y(x),Y(y))$ qu'on peut estimer facilement à partir des données pour $x - y$ petit (du moins si l'on a deux acquisitions, ou si la forme du bruit le permet). De plus, il est bien connu

que le krigeage linéaire est robuste par rapport aux petites variations dans le choix des fonctions de covariance quand le comportement à l'origine de la fonction de covariance n'est pas trop régulier. Il semble donc qu'il nous donne une méthode fiable et peu compliquée pour filtrer une variété importante d'images bruitées.

Avant de passer à l'étude des équations du krigeage linéaire, il faut souligner les cas pour lesquels le krigeage linéaire n'est pas adapté. Puisqu'il s'agit d'un filtre linéaire, on voit que le krigeage lisse les données. Si nous avons une image qui contient un nombre important de pics, le krigeage risque de lisser ces pics (de plus l'estimation du variogramme pourra être difficile). C'est pour cette raison que, dans le filtrage linéaire des images de solidification, on est parfois obligé d'appliquer un masque aux valeurs fortes. Par exemple, dans une microségrégation, on masque les valeurs fortes dans la partie liquide. Il est légitime d'appliquer une telle procédure ici. Les caractéristiques de la solidification de la zone liquide sont souvent différentes de celle des dendrites. De plus, le bruit poissonien, ayant un écart-type en $\sqrt{Y(x)}$, est relativement plus faible sur les valeurs fortes (dans le sens que le bruit relatif au signal décroit en $\frac{1}{\sqrt{Y(x)}}$), et le filtrage de ces dernières est moins important. Mais pour les autres types d'image, il peut arriver qu'on ait une image homogène qui contienne des pics importants. Il semble arbitraire d'appliquer un masque dans un tel cas, et il pourra être intéressant d'envisager un filtrage non linéaire selon les méthodes décrites au chapitre VI.

Un autre cas pour lequel le krigeage n'est pas très bien adapté, est lorsque nous avons une image sous-jacente en forme de mosaïque, et qu'il existe alors des petites régions où l'image sous-jacente est plus ou moins constante. Si le saut moyen entre ces régions est plus important que le bruit, un filtre linéaire risque soit de ne pas filtrer suffisamment, soit de trop filtrer et donc de lisser les frontières entre les régions. Un filtre non linéaire de type séquentiel alterné qui est basé sur les outils morphologiques d'ouverture et de fermeture [1] peut être mieux adapté dans ce cas.

Il existe d'autres types de problèmes importants qui ne sont pas étudiés dans ce texte, notamment le problème de classification. Si nous avons une image, divisée en N composants qui ne sont pas forcément connexes, nous voulons savoir à quel pixel le composant appartient. On rencontre souvent ce type de problème dans l'imagerie satellite où on cherche à classer les pixels selon le type de végétation par exemple. La théorie des champs markoviens a été appliquée avec quelque

succès sur des données simulées pour ce type de problème [2], [3]. Il est à noter que cette méthode est lourde en termes d'hypothèses de départ, parce qu'elle dépend explicitement de la loi spatiale complète, laquelle, en toute rigueur, n'est pas accessible avec les seules données expérimentales. Une autre méthode, assez proche de celle qui sera décrite dans ce chapitre, a été appliquée au problème de classification [4], et bien qu'on puisse supposer qu'elle n'est pas adaptée au problème, elle donne des solutions comparables à celles qu'on obtient avec la méthode des champs markoviens pour le problème à deux composantes.

2.1. — *LE KRIGEAGE UNIVARIABLE*

Dans cette section, nous considérons le filtrage d'une seule variable, et à la section 2.2, nous généraliserons ce résultat aux données multivariables.

2.1.1. — *Filtrage univariable*

Soit $Y(x)$ un modèle stationnaire d'ordre deux pour l'image sous-jacente et $Z(x)$ un modèle pour les données randomisées. Nous admettons que les fonctions de covariance stationnaires $C^Y(h)$ et $C^{Y,Z}(h)$ sont connues, et que la fonction de covariance $C^Z(x, x+h)$ est également connue. Admettons que la moyenne de Z soit stationnaire et que $E(Z) = E(Y)$ (après changement d'échelle si nécessaire). Sans restreindre la généralité, on peut prendre $E(Y) = 0$. (En effet on peut prendre $Y - E(Y)$ et $Z - E(Y)$ comme fonctions aléatoire. L'objection que dans le cas poissonien Y et Z doivent être positives n'est pas très importante. Le filtrage linéaire ne fait pas la distinction entre les différents types de bruit. L'hypothèse d'un bruit poissonien intervient seulement pour la modélisation de la fonction de covariance de Y à partir de celle de Z). On cherche un estimateur linéaire de $Y(x)$

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

où N est le nombre de points dans le voisinage de α, V_α. Les pondérateurs $\{\lambda^\alpha\}$ sont choisis tels que $\mathrm{Var}[Y(x) - Y^*(x)]$ est un minimum.

$$Q(\lambda_1,\ldots,\lambda_N) = \mathrm{Var}(Y(x) - Y^*(x))$$

$$= \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N} \lambda^\alpha \lambda^\beta C^Z(x_\alpha, x_\beta) - \sum_{\alpha=1}^{N} \lambda^\alpha C^{Y,Z}(x_\alpha - x) + C^Y(0) \qquad (2.1)$$

On minimise par $\frac{\partial Q}{\partial \lambda_i} = 0 \quad \forall \ i$, et on trouve les équations du krigeage [5]

$$\sum_{\alpha=1}^{N} \lambda^\alpha C^Z(x_\alpha, x_\beta) = C^{Y,Z}(x_\beta - x) \qquad (2.2)$$

Puisque nous avons fait l'hypothèse que C^Z est strictement positif, on peut résoudre ces équations pour $\{\lambda^\alpha\}$.

Alors la variance de krigeage, ou erreur résiduelle, est donnée par la substitution de (2.2) dans (2.1)

$$\sigma_K^2 = C^Y(0) - \sum_{\alpha=1}^{N} \lambda^\alpha C^{Y,Z}(x_\alpha - x)\,.$$

Dans les applications à l'analyse d'images, on utilise un rapport signal sur bruit (SSB). Le rapport SSB initial (avant filtrage) est défini par

$$(\mathrm{SSB})_\mathrm{I} = \frac{\mathrm{Var}\ Y(x)}{\mathrm{Var}\ Z(x) - \mathrm{Var}\ Y(x)}$$

et après filtrage, nous avons le rapport SSB final

$$(\mathrm{SSB})_\mathrm{F} = \frac{\mathrm{Var}\ Y(x)}{\sigma_K^2}$$

On voit ici un des avantages du krigeage comme filtre : il ne donne pas seulement un estimateur du signal sous-jacent mais également un estimateur de l'erreur résiduelle.

L'hypothèse que $Y(x)$ est stationnaire avec une moyenne connue est souvent un peu trop forte. Une hypothèse moins forte dont on a vu qu'elle était raisonnable pour les images de solidification est celle de stationnarité locale pour $Y(x)$. Ici, la moyenne $m(x)$ n'est pas constante partout dans l'image, mais peut être considérée comme constante sur le voisinage V_x du point x où l'on veut faire un filtrage. Dans ce cas, il nous faudra une condition de non biais

$$E[Y(x) - Y^*(x)] = 0$$

donc

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

et le krigeage doit être soumis à cette condition. La méthode de Lagrange nous donne les équations du krigeage correspondantes [5]

$$\begin{aligned} \sum_{\alpha=1}^{N} \lambda^\alpha \; C^Z(x_\alpha, \; x_\beta) + \mu \quad &= C^{Y,Z}(x_\beta - x) \\ \sum_{\alpha=1}^{N} \lambda^\alpha \qquad\qquad\qquad\quad &= 1 \end{aligned} \tag{2.3}$$

où μ est le paramètre de Lagrange. La variance de krigeage σ_K^2 est donnée par

$$\sigma_K^2 = C^Y(0) - \sum_\alpha \lambda^\alpha \; C^{Y,Z}(x_\alpha - x) - \mu$$

Pour compléter cette section, on écrira les équations dans le cas où l'on a une FAI-k. Supposons qu'une covariance généralisée pour Y soit $K^Y(h)$, et qu'on ait une covariance généralisée de Z de la forme

$$K^Z(x,y) = K^Y(x-y) + f(x) \; \delta_{xy}$$

avec $f(x) > 0 \quad \forall \; x$.

(Par exemple dans le cas poissonien, on a vu que $f(x)$ était égale à la fonction de moyenne pour Y, $m(x)$). Le krigeage est obtenu en minimisant $\mathrm{Var}(Y(x) - Y^*(x))$ tel que $(\delta_x - \sum_\alpha \lambda^\alpha \delta_{x_\alpha})$ est une combinaison linéaire autorisée, c'est-à-dire

$$\sum_\alpha \lambda^\alpha f_\alpha^\ell = f_x^\ell \qquad\qquad \ell = 0,\ldots,k$$

et la méthode de Lagrange donne les équations

$$\begin{aligned} \sum_{\alpha=1}^{N} \lambda^\alpha K^Z(x_\alpha, x_\beta) + \mu_\ell \, f_\beta^\ell \quad &= K^{Y,Z}(x - x_\beta) \qquad\qquad \beta = 1,\ldots,N \\ \sum_{\alpha=1}^{N} \lambda^\alpha \, f_\alpha^\ell \qquad\qquad\quad &= f_\alpha^\ell \qquad\qquad \ell = 0,\ldots,k \end{aligned}$$

NOTA. – Pour la poissonisation dans le cas non stationnaire ou même dans le cas localement stationnaire, la variance du bruit n'est pas stationnaire

$$C^Z(x,y) = C^Y(x-y) + m(x)\ \delta_{xy}$$

Pour une étude correcte, il faudra un estimateur de $m(x)$ localement (ce qu'on peut trouver par moindres carrés par exemple). Les équations du krigeage au point x dépendront alors du point x, et il faudra résoudre les équations linéaires pour chaque point à estimer.

Une hypothèse simplificatrice que nous appliquons au cas localement stationnaire consiste à prendre

$$C^Z(x,y) = C^Y(x-y) + m$$

où m est la moyenne spatiale de l'image.

Les équations du krigeage sont maintenant identiques pour chaque point x dans l'image (sauf pour un effet de bord) et il suffira donc de les résoudre une fois pour obtenir un système de pondérateurs $\{\lambda^a\}$ qui soit invariant par translation. L'hypothèse d'un bruit stationnaire peut être considérée comme raisonnable s'il n'y a pas de dérive importante sur l'image.

EXEMPLES

1. – Pour un premier exemple, considérons l'image gaussienne simulée avec un modèle sphérique d'une portée de 20 pixels et avec le palier présenté à la figure 2.1. La moyenne de cette image est de 100.0 et la variance de 97.362. Nous avons déjà vu au chapitre I que l'estimation des variogrammes sous-jacents et observés est faite correctement. L'image poissonisée est montrée à la figure 2.2 et l'image filtrée avec un voisinage de krigeage de 15×15 sur la figure 2.3. Le rapport signal/bruit initial est de 0.9288 et après filtrage, on trouve un rapport SSB estimé de 9.1365. Puisqu'il s'agit d'une simulation, on peut calculer le vrai rapport SSB après filtrage. Il est de 8.7166, et on trouve un bon accord entre l'estimateur de SSB et la valeur réelle.

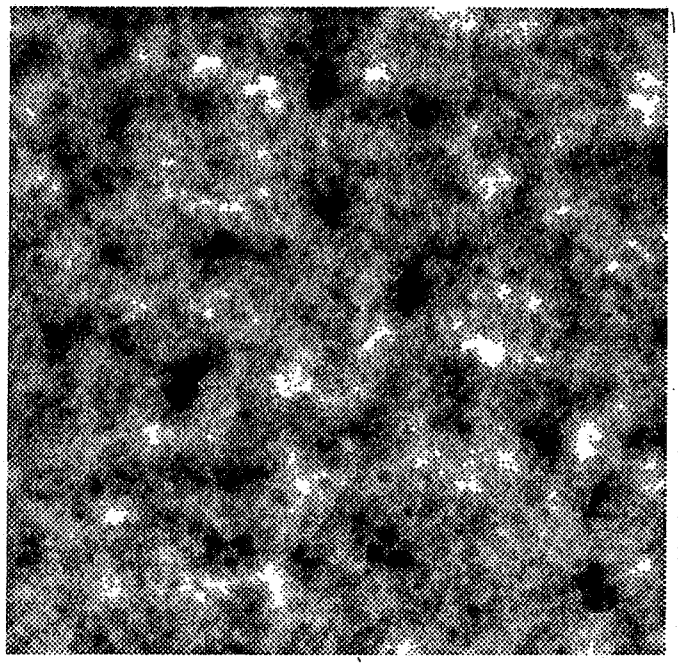


Figure 2.1 Image réelle

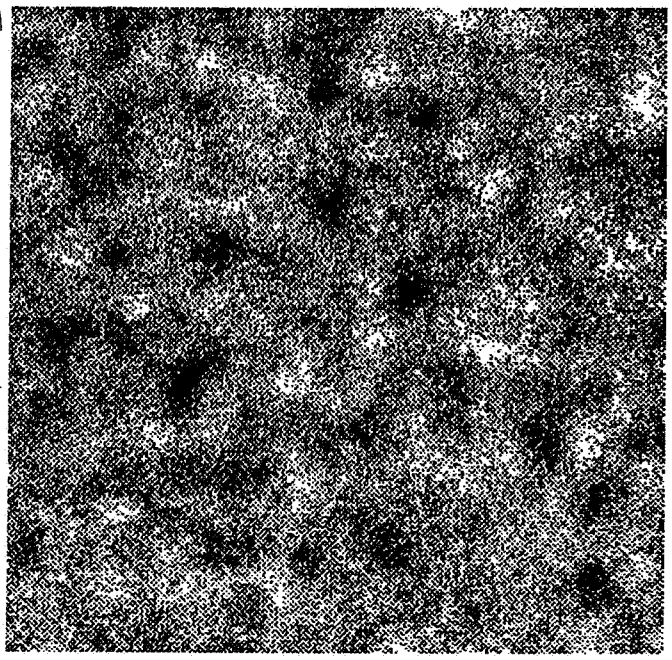


Figure 2.2 Image bruitée

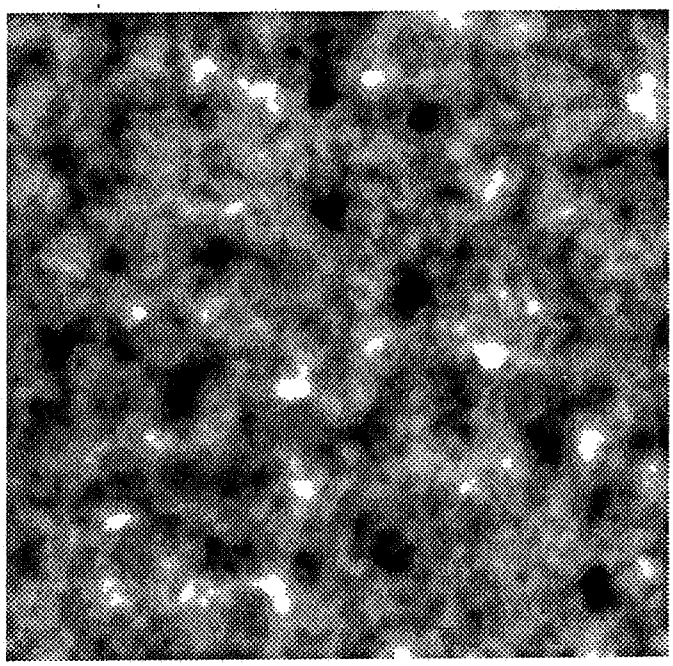


Figure 2.3 Filtrage linéaire

2. – On peut penser que le cas précédent a donné de bons résultats parce que les données simulées suivent une loi gaussienne. On reprend donc ici la même image, mais on applique une anamorphose avant de poissoniser. L'image sous-jacente est montrée à la figure 2.4 et son histogramme à la figure 2.5. La loi de $Y(x)$ (qui est dérivée d'une loi lognormale) à une moyenne de 13.763 et une variance de 52.705. L'image bruitée est montrée sur la figure 2.6, et le résultat d'un krigeage linéaire sur la figure 2.7. Le rapport signal/bruit initial est de 3.4138 et l'estimateur du rapport SSB après filtrage est de 15.412. La valeur réelle du rapport SSB après filtrage est de 13.521 et l'accord entre l'estimateur et la réalité est moins bon que dans le cas gaussien, mais reste satisfaisant.

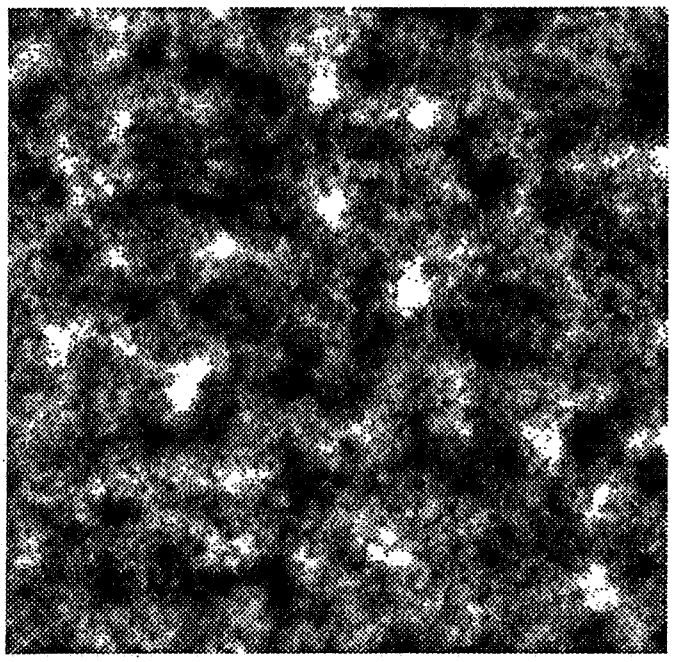


Figure 2.4 Image réelle

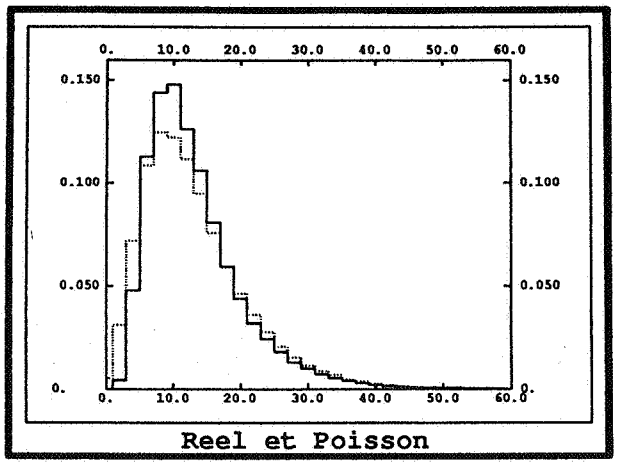


Figure 2.5 Histogramme de l'image sous-jacente (trait continu) et de sa poissonisée.

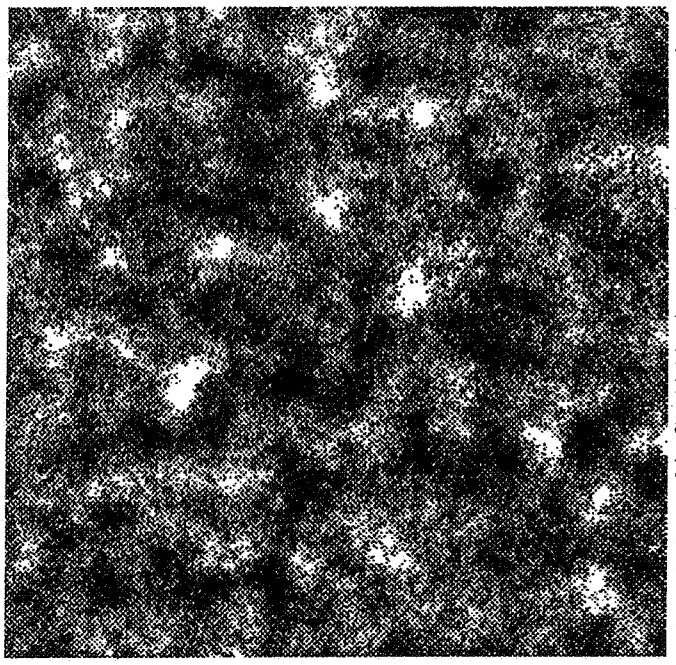


Figure 2.6 Image bruitée

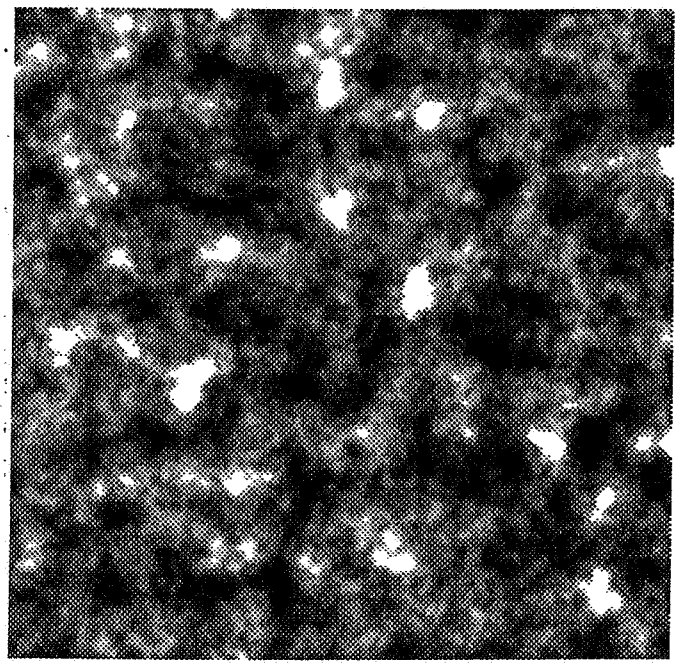


Figure 2.7 filtrage linéaire

3. – Sur la figure 2.8 on voit une image de microségrégation. Nous avons déjà vu cette image au chapitre I et le variogramme était modélisé comme la somme de deux modèles cubiques de portées 25 et 45 respectivement. Avec un filtre linéaire localement stationnaire de voisinage 15 × 15, on obtient la figure 2.9. On voit que le filtrage est très efficace à cause du fait que la portée du variogramme est assez longue et le comportement à l'origine du variogramme est quadratique. Sur un écran graphique couleurs on arrive même à discerner la solidification secondaire sur quelques dendrites (elle est à peine visible sur des images noir et blanc qui ne contiennent que 8 niveau de gris), ce qui n'était nullement évident sur l'image initiale. Le rapport signal/bruit initial est de 2.400 et l'estimation du rapport signal/bruit final est de 83.803.

4. – Sur la figure 2.10 on trouve une image du centre de la plaque. Ici le pas entre les pixels est de 25 μm , donc plus de 8 fois supérieur à celui de l'image de microségrégation de l'exemple 3. L'image filtrée est montrée sur la figure 2.11. On voit bien sur l'image qu'il existe un nombre important de dendrites qu'on peut observer à l'échelle de 4 à 10 pixels. Par conséquent le variogramme (figure 2.12) croît beaucoup plus rapidement que pour l'image de microstructure et on voit à peine le comportement quadratique à l'origine. Le rapport signal/bruit initial est de 5.522 et l'estimateur du rapport SSB après filtrage est de 11.82. En comparant ces résultats avec l'exemple précédent, on voit que le krigeage a "compris" que la deuxième image est beaucoup plus difficile à filtrer à cause des petites structures, et comme on peut le prévoir, les résultats seront moins bons que pour l'exemple de microségrégation.

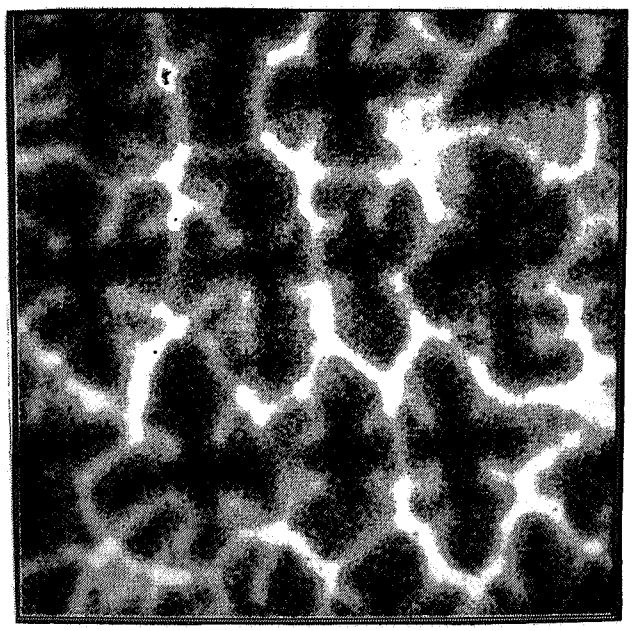


Figure 2.8 Microségrégration, 3µm par pixel, 512x512

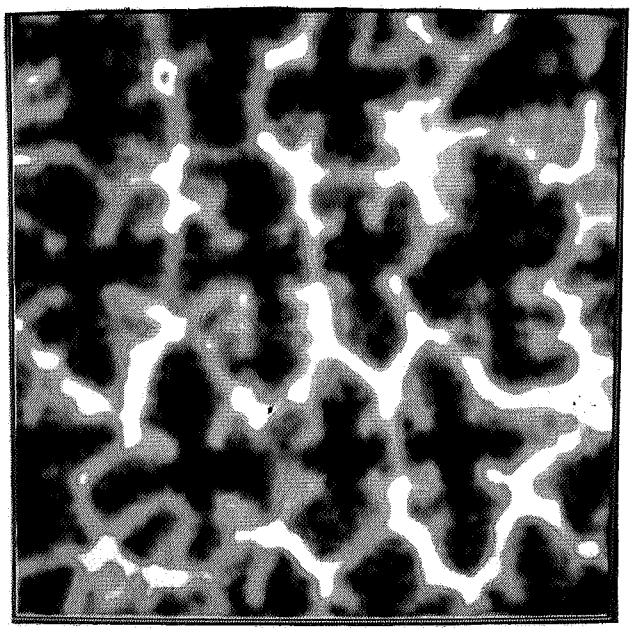


Figure 2.9 Filtrage linéaire

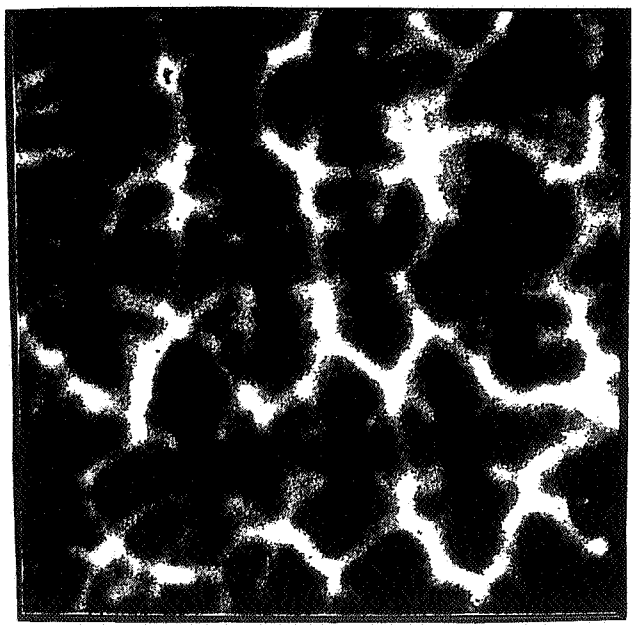


Figure 2.8 Microségrégration, 3µm par pixel, 512x512

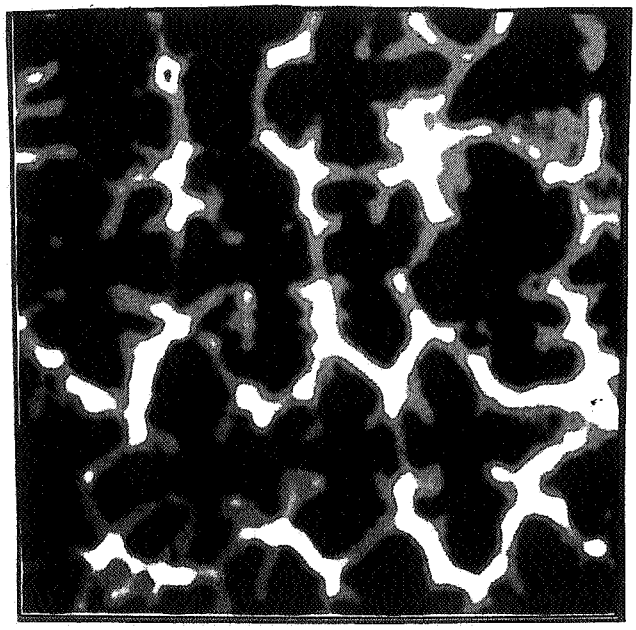


Figure 2.9 Filtrage linéaire

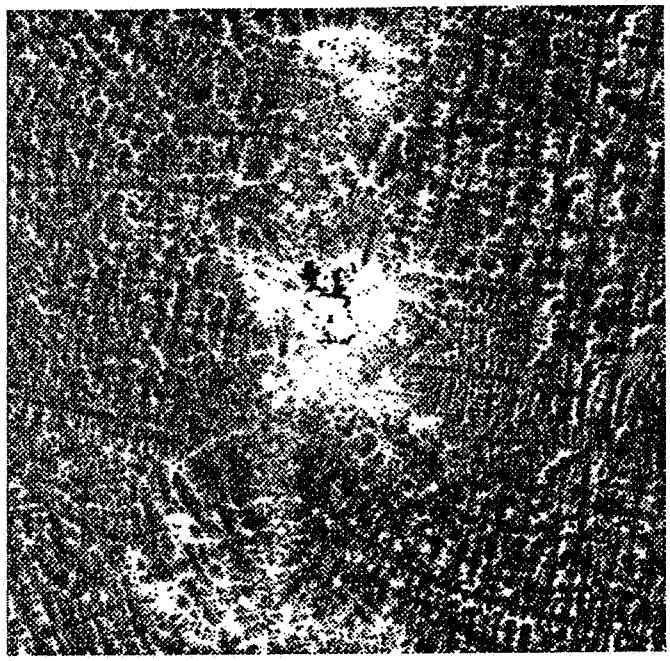


Figure 2.10 Image du centre de la plaque,
pas = 25µm, 512x512, élément = Mn

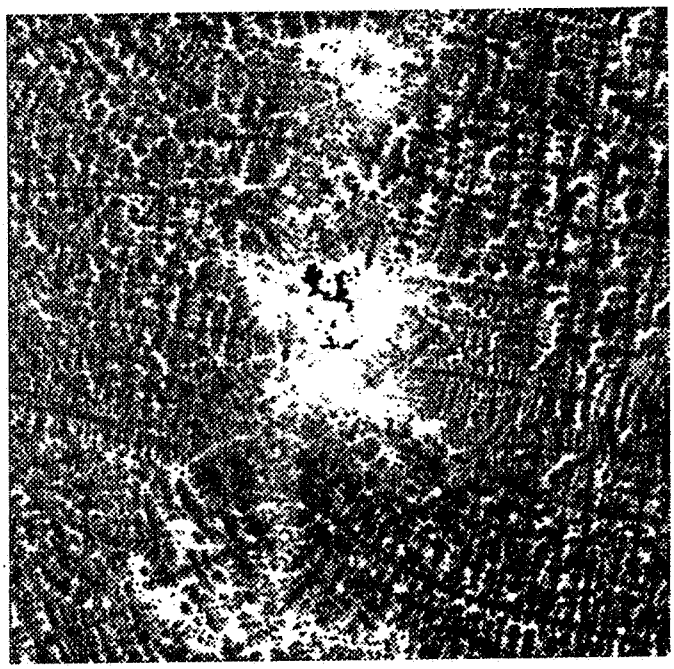


Figure 2.11 Filtrage linéaire

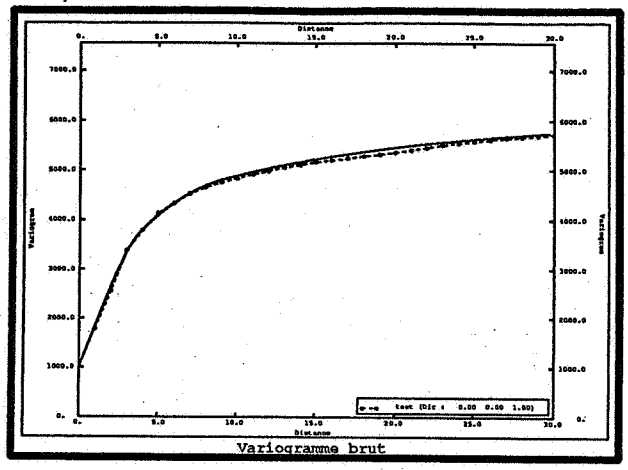


Figure 2.12 Variogramme du centre de la plaque (bruité) et son ajustement.

5. – On montre à la figure 2.38 une version filtrée du carbone. Pour un exemple de filtrage dans le cas non poissonien, on reprend l'exemple des "taux" du chapitre I. Sur la figure 2.13, on représente le taux sous-jacent. On montre la répartition de population à la figure 2.14 et le nombre d'occurrences, $Z(x)$, observé à la figure 2.15. En toute rigueur, pour filtrer le taux observé, il faudrait utiliser une covariance non stationnaire et par conséquent résoudre des équations de krigeage à chaque pixel. Pour une première approximation, nous avons filtré avec un bruit moyen et le résultat est présenté sur la figure 2.16.

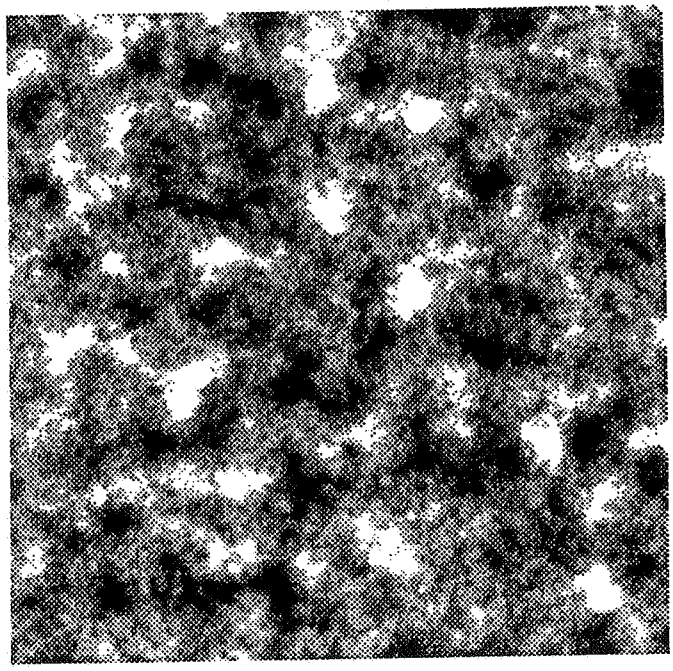


Figure 2.13 Taux réel

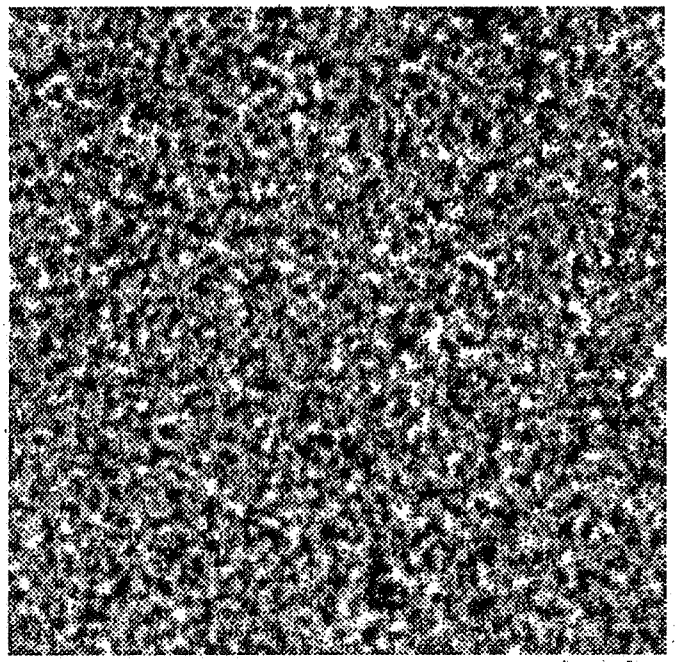


Figure 2.14 Population réelle

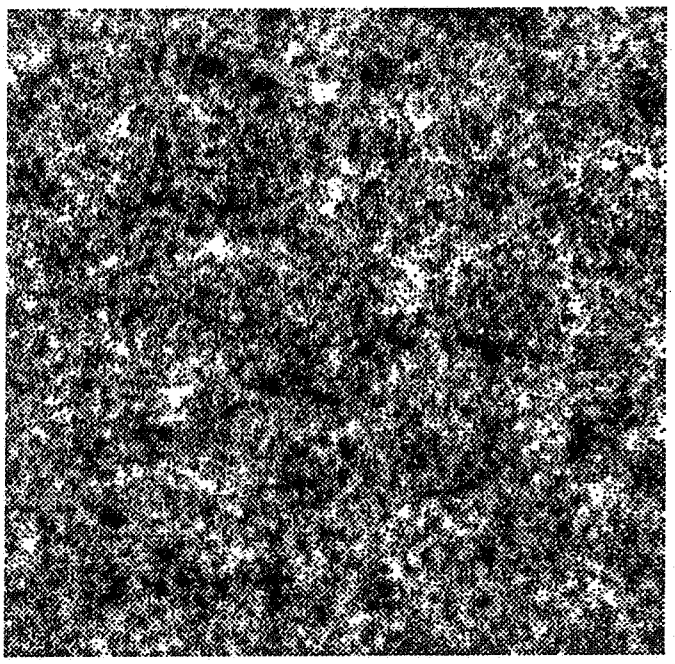


Figure 2.15 Nombre d'occurrences favorables

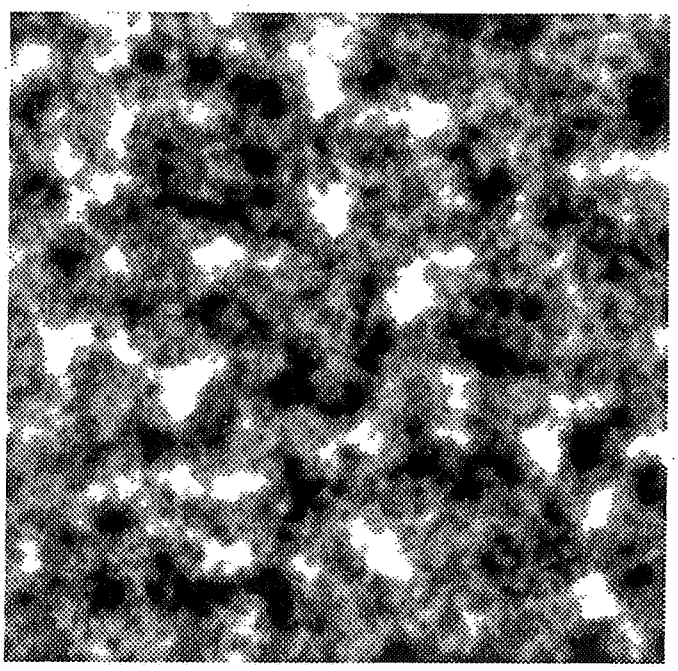


Figure 2.16 Filtrage linéaire

6. – Le filtrage linéaire peut être utilisé pour réduire le temps d'acquisition d'une image. Nous avons quatre acquisitions de l'élément P avec 12.5, 25, 50 et 100 *ms* par pixel. On applique un filtre linéaire à chaque image bruitée. Les images filtrées sont montrées à la figure 2.17, et les rapports signal sur bruit correspondants sont résumés au Tableau 2.1. On voit qu'en termes de rapport signal sur bruit l'image filtrée de 12.5 *ms* est aussi bonne que l'image brute de 50 ms. Ceci est vraisemblablement un peu trop optimiste, mais visuellement on peut constater que le filtrage de l'image à 25 *ms* donne de bons résultats. Il faut un temps d'acquisition supérieur à 100 *ms* pour obtenir une image d'assez bonne qualité sans filtrage.

Tableau 2.1.

Temps d'acquisition	SSB original	SSB après filtrage
15 *ms*	0.941	5.216
25 *ms*	1.883	6.929
50 *ms*	5.338	12.502
100 *ms*	10.676	18.613

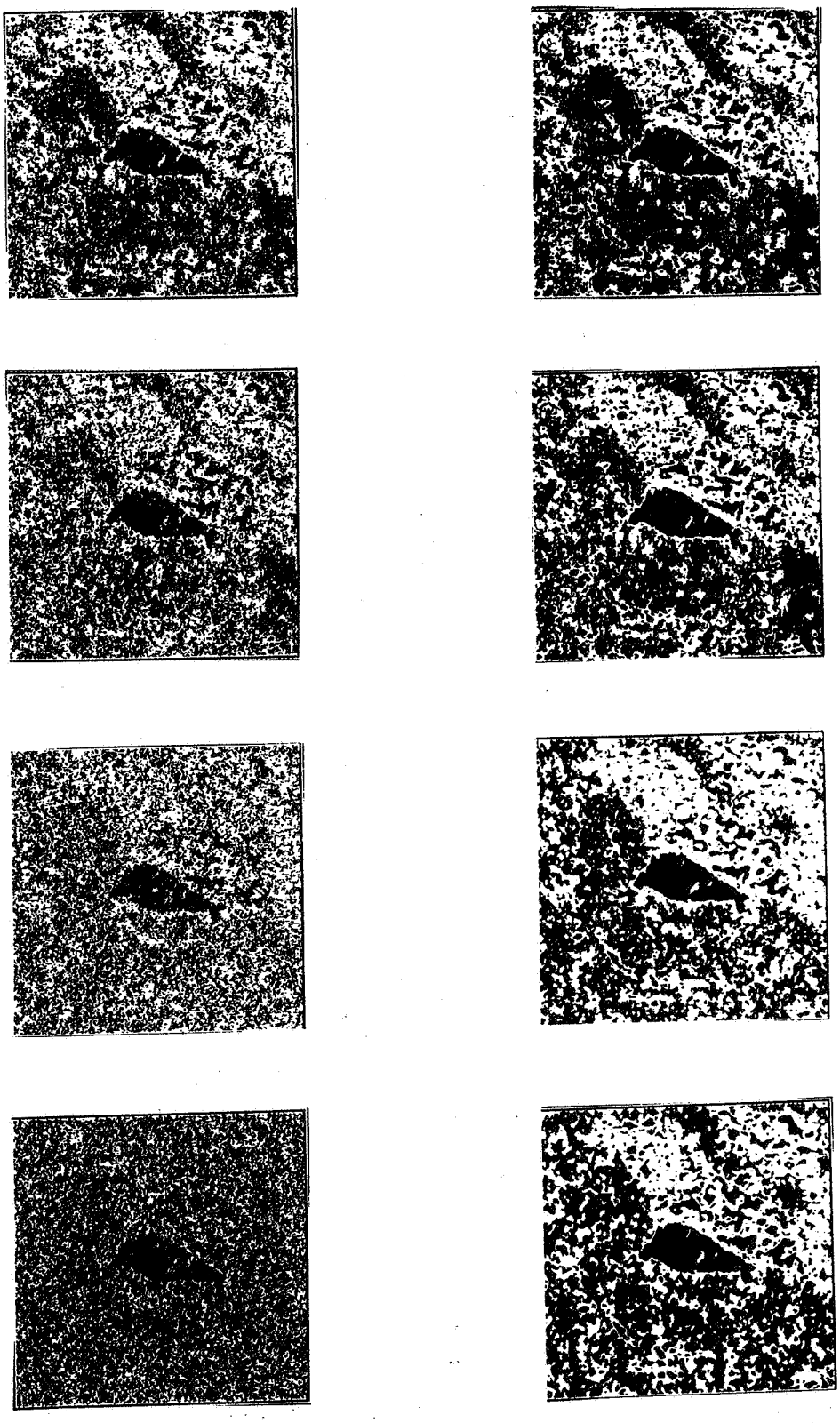
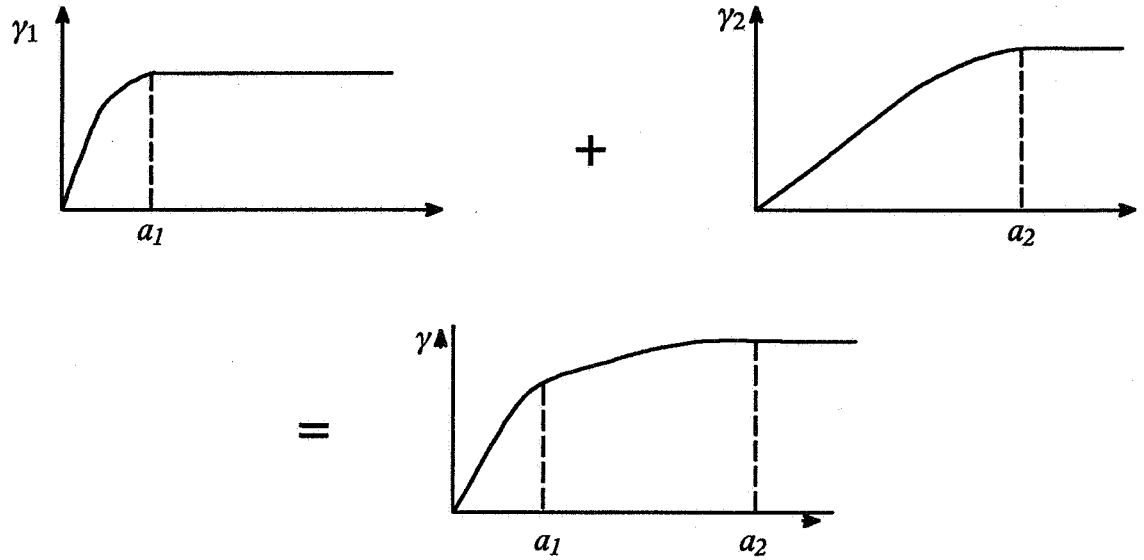


Figure 2.17

2.1.2. — *L'estimation des composantes spatiales*

Un phénomène modélisé par une fonction aléatoire peut avoir plusieurs structures à différentes échelles. Par exemple, sur les images de solidification, on voit des structures dendritiques de l'ordre de quelques dizaines de μm et des structures d'origine de type convection de l'ordre du cm et même plus grandes. Entre les deux, on trouve des structures de la taille des Vés ségrégés, etc. Le calcul d'un variogramme expérimental montre assez souvent l'existence et l'importance de ces structures.



Parallèlement, on trouve que l'ajustement d'un variogramme se fait souvent à l'aide de plusieurs modèles théoriques de structure simple $\gamma(h) = \Sigma\ \gamma_i(h)$. Dans le cas général, il n'y aura aucune raison pour que ces modèles élémentaires $\gamma_i(h)$ correspondent à une structure réelle dans l'image sous-jacente. Mais si l'on peut faire l'hypothèse que la fonction aléatoire sous-jacente $Y(x)$ admet la décomposition

$$Y(x) = \sum_i Y_i(x)$$

où $Y(x)$ et $\{Y_i(x)\}$ sont des fonctions aléatoires stationnaires d'ordre deux et les $Y_i(x)$ sont mutuellement orthogonales, $\mathrm{Cov}(Y_i(x), Y_j(y)) = 0$, et si on trouve la décomposition de la fonction de covariance (ou du variogramme) correspondante

$$C(h) = \sum_i C_i(h)$$

il peut être intéressant d'estimer les composants $Y_i(x)$ séparément à partir des données $\{Z(x_a)\}$.

Ainsi, si on prend $E[Y(x)] = E[Y_i(x)] = 0$, on peut considérer l'estimateur

$$Y_i^* = \sum_\alpha \lambda^\alpha Z_\alpha$$

et on démontre facilement que les équations du krigeage pour $\{\lambda^\alpha\}$ dans le cas stationaire sont

$$\sum \lambda^\alpha \text{ Cov } (Z(x_\alpha),\ Z(x_\beta)) = \text{Cov } (Z(x_\beta),\ Y_i(x))$$

et il est facile de démontrer que $\text{Cov } (Z(x_\beta),\ Y_i(x)) = C_i(x_\beta - x)$.

Dans le cas localement stationnaire, une difficulté réside dans le fait que les moyennes de $Y(x)$ et $\{Y_i(x)\}$ ne sont pas connues. Par conséquent, on ne sait pas décomposer la moyenne de $Y(x)$ en termes de ses composants $m_i = E[Y_i(x)]$. Dans un tel cas, le mieux qu'on puisse faire est d'estimer le résidu $Y_i'(x) = Y_i(x) - m_i$. On veut alors minimiser $\text{Var } [Y_i'(x) - Y_i'^*(x)]$ tel que $E[Y_i'(x) - Y_i'^*(x)] = 0$. La dernière condition s'écrit $\Sigma \lambda^\alpha = 0$ et les équations de krigeage s'écrivent

$$\sum_{\alpha=1}^{N} \lambda^\alpha C^Z(x_\alpha,\ x_\beta) + \mu = C^{Y_i}(x - x_\beta)$$

$$\sum_{\alpha=1}^{N} \lambda^\alpha = 0$$

Pour une étude plus rigoureuse et une généralisation aux FAI-k, voir [6].

EXEMPLES

1. – On commence avec une image simulée. Notre image sous-jacente est montrée à la figure 2.18. Il s'agit ici d'une somme de deux autres simulations avec des portées de 5 et 20 respectivement. Les images des deux structures sont montrées aux figures 2.19 et 2.20. Le variogramme sous-jacent $\gamma^Y(h)$ se décompose en termes de deux structures simples de la manière suivante

$$\gamma^Y(h) = 0.9\ \gamma^5(h) + 0.1\ \gamma^{20}(h)$$

L'image sous-jacente est randomisée par une loi de Poisson et l'image bruitée est montrée sur la figure 2.21. Le variogramme de l'image bruitée est montrée sur la figure 2.22. Le rapport signal/bruit initial est de 3.07.

On peut estimer les deux composantes et on voit les résultats sur les figures 2.23 et 2.24 respectivement. La composante de portée 5 est assez bien estimée, ce qui n'est pas très surprenant étant donné que la plupart de la variabilité est associée à cette échelle. Il est intéressant de voir qu'on arrive à obtenir un estimateur de l'échelle supérieure qui est acceptable même s'il n'y a qu'une contribution très faible à la variabilité de l'image initiale.

2. – On peut reprendre l'image du centre de la plaque déjà vue. Elle est présentée à nouveau sur la figure 2.25 et l'estimateur de la variabilité pour la première structure (4 pixels) est montré sur la figure 2.26. Ceci nous aide à voir plus clairement la structure dendritique.

Toujours à propos du problème de solidification, on voit sur la figure 2.27 une image prise dans la zone équiaxe. Le pas entre les pixels est de $25\mu m$. Cette image n'est évidemment pas homogène et on voit l'existence d'une inclusion avec des structures qui semblent assez différentes. Pourtant l'estimation de la première composante permet de voir très nettement la structure dendritique ou de petits amas de dendrites (fig. 2.28).

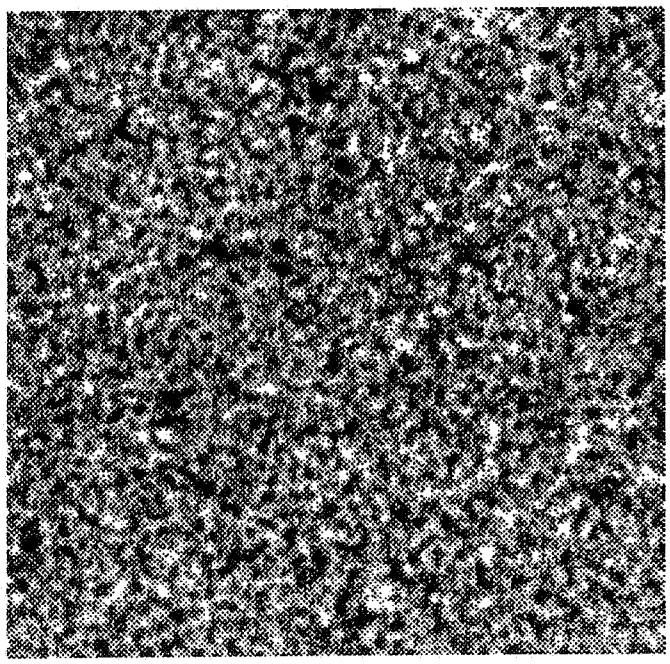


Figure 2.18 Image réelle

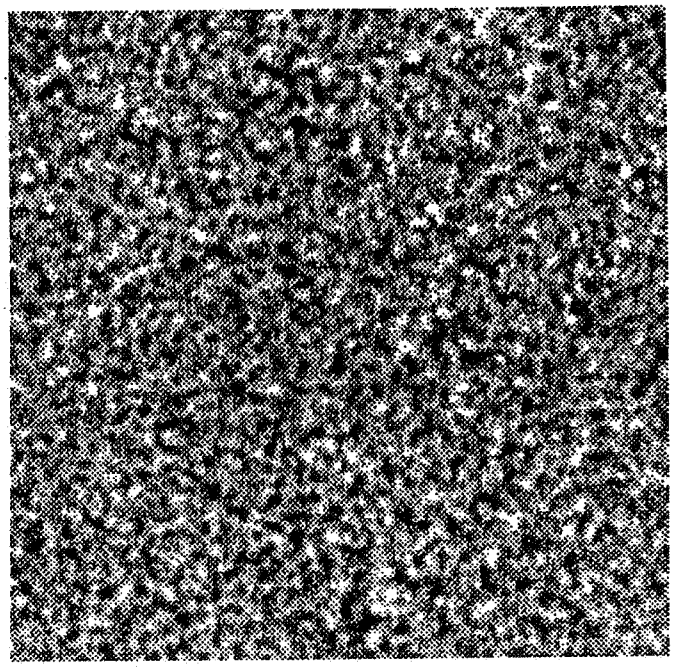


Figure 2.19 Composante de portée 5

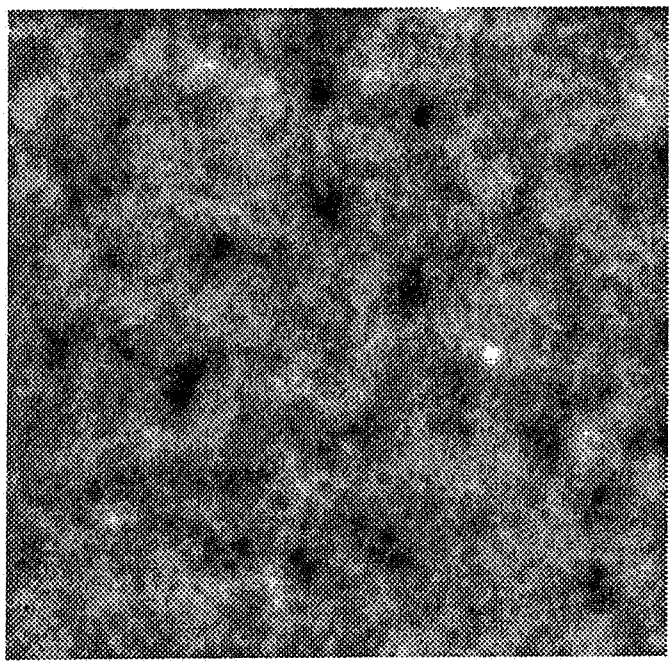


Figure 2.20 Composante de portée 20

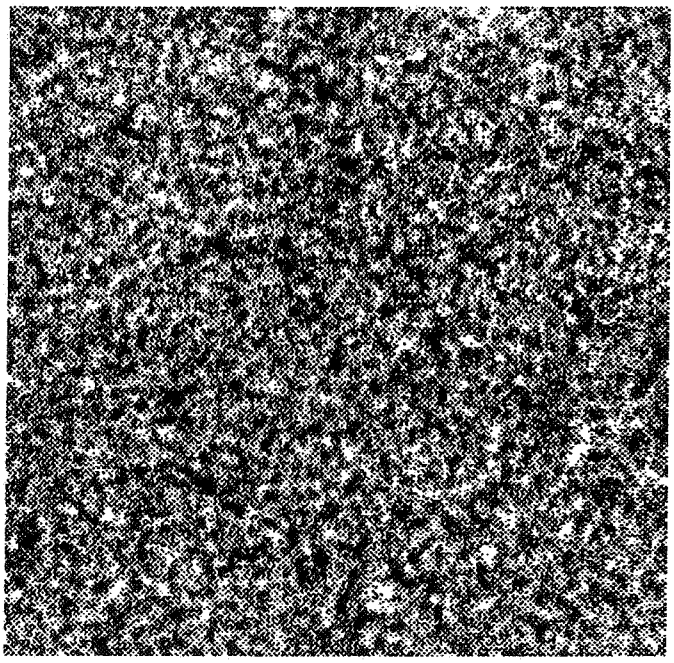
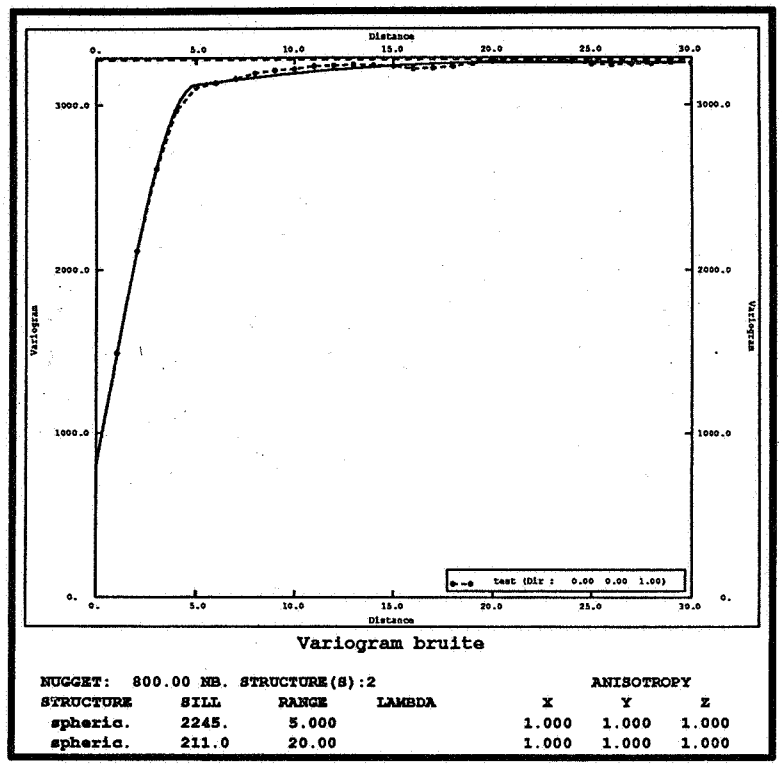


Figure 2.21 Image bruitée



Variogram bruite

NUGGET: 800.00 NB. STRUCTURE(S):2

STRUCTURE	SILL	RANGE	LAMBDA	ANISOTROPY X	Y	Z
spheric.	2245.	5.000		1.000	1.000	1.000
spheric.	211.0	20.00		1.000	1.000	1.000

Figure 2.22 Variogramme de l'image bruitée et son ajustement.

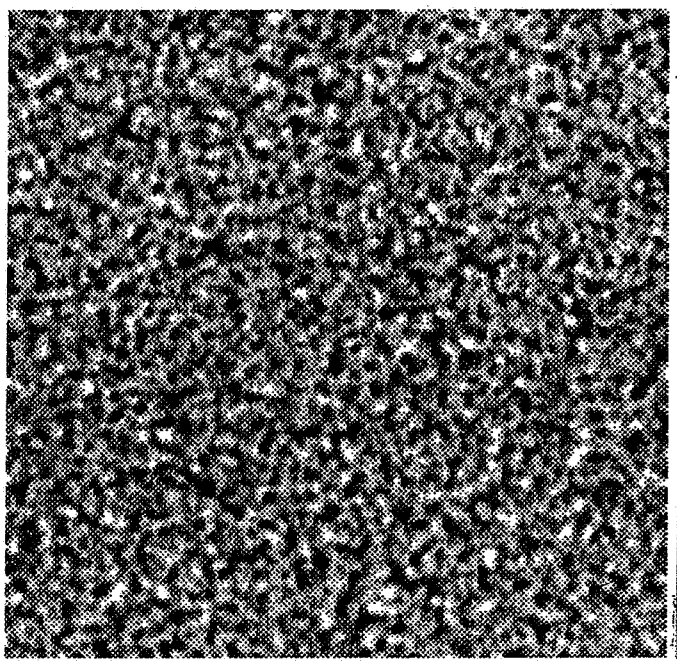


Figure 2.23 Estimation de la composante de portée 5

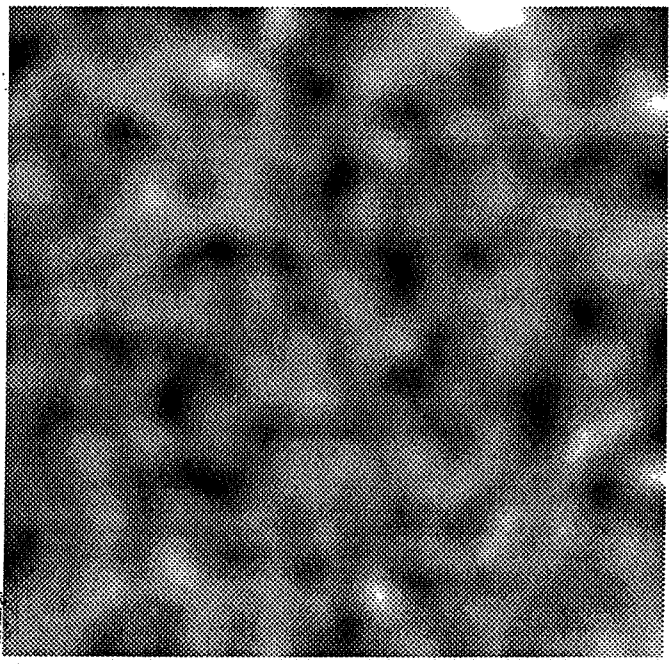


Figure 2.24 Estimation de la composante de portée 20

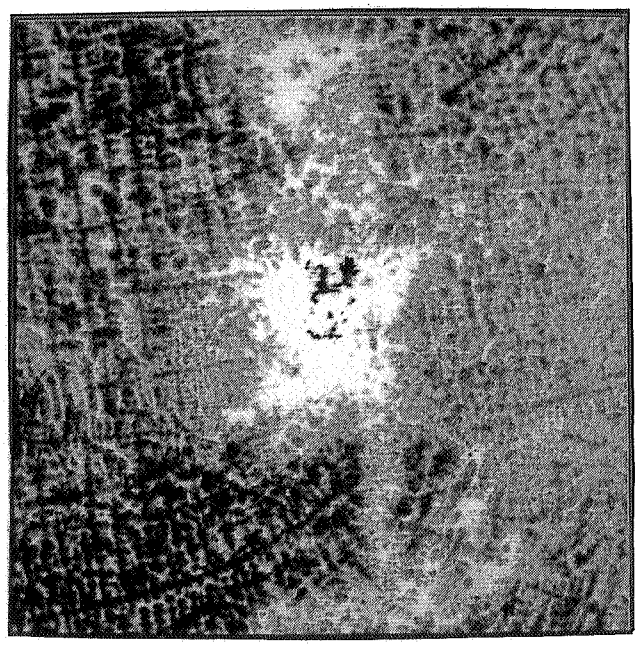


Figure 2.25 Centre de la plaque

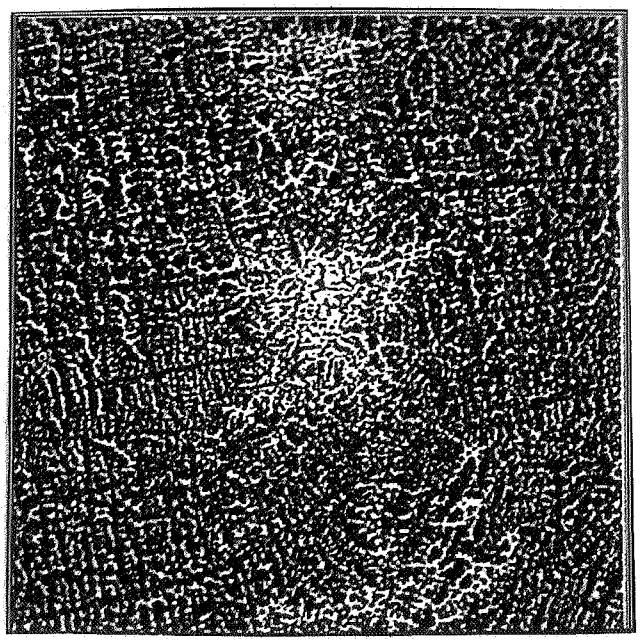


Figure 2.26 Estimation de la composante de portée 4

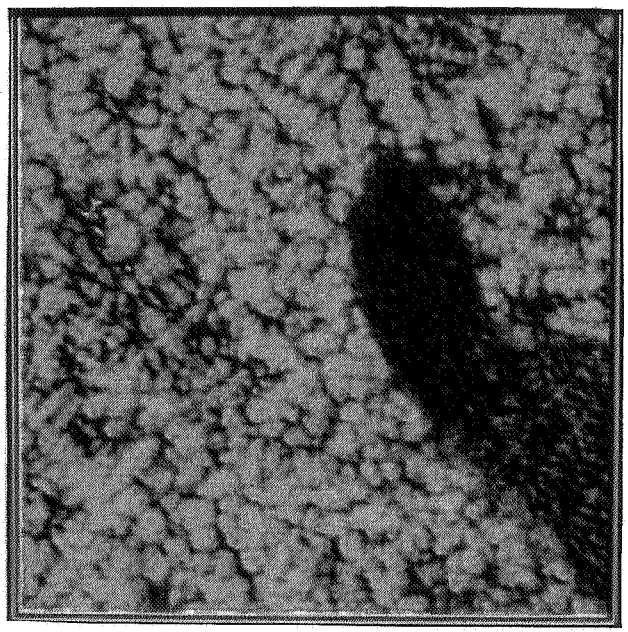


Figure 2.27 Zone equiaxe près du centre de la plaque, pas = 25µm, 256 x 256.

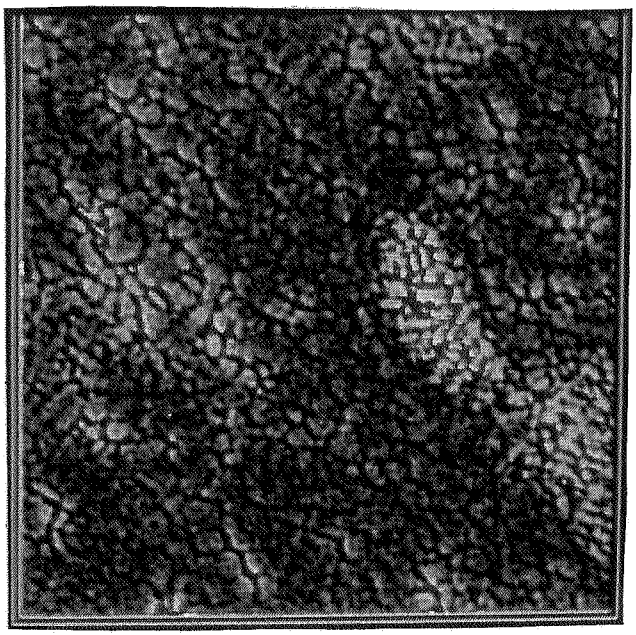


Figure 2.28 Estimation de la composante de portée 6

2.2. — *LE CAS MULTIVARIABLE*

Supposons maitenant qu'il y ait plusieurs éléments $\{Y_i(x)\}_{i=1}^n$ et l'on observe les données bruitées correspondantes $\{Z_i(x)\}_{i=1}^n$. Quand les éléments $Y_i(x)$ sont corrélés, on peut avoir intérêt à utiliser plusieurs éléments bruités $\{Z_i(x)\}$ pour estimer un élément sous-jacent. On veut alors considérer des estimateurs de la forme

$$Y_k^*(x) = \sum_{i=1}^{n} \sum_{\alpha=1}^{N} \lambda_i^\alpha \, Z_i(x_\alpha)$$

Evidemment, il faudra que

$$\text{Var}\left(Y_k(x) - Y_k^*(x)\right) \geq 0 \tag{2.4}$$

Mais cette condition impose des contraintes non triviales sur la modélisation des variogrammes et des variogrammes croisés. Une méthode souvent utilisée pour assurer la cohérence des modèles est le modèle linéaire [6], [7]. On admet qu'il existe des fonctions aléatoires orthogonales X_p^u telles que pour chaque élément i, on a des coefficients A_{iu}^p tels que

$$Y_i(x) = \sum_{u=1}^{n} \sum_{p=1}^{n} A_{iu}^p \, X_p^u$$

et tels que pour u fixe, toutes les fonctions aléatoires X_p^u ont la même covariance localement stationnaire $C^u(h)$. On interprète l'indice u comme l'indice d'échelle (effet de pépite, échelle de portée courte, échelle de portée longue, etc.). Dans ce cas les n fonctions aléatoires $\{X_p^u\}_{p=1}^n$ admettent une interprétation comme composantes principales associées à la variabilité à l'échelle u [6]. Puisque les X_p^u sont orthogonales, on trouve que

$$\text{Cov}\left[X_p^u(x+h),\ X_q^v(x)\right] = \delta^{uv}\ \delta_{pq}\ C^u(h)$$

et

$$\text{Cov}\left[Y_i(x+h),\ Y_j(x)\right] = \sum_u B_{ij}^u\ \ C^u(h)$$

avec

$$B_{ij}^u = \sum_p A_{iu}^p \, A_{ju}^p$$

Les matrices de variance/covariance B_{ij}^u sont donc positives, et il est facile de démontrer que (2.4) est respecté. Une méthode efficace et simple a été développée pour l'ajustement des modèles linéaires de corégionalisation [8], [9].

Ce modèle peut être généralisé aux FAI-k [6].

Le modèle pour les covariances de Z sera, après les calculs du chapitre I

$$\text{Cov}\left(Z_i(x),\ Z_j(y)\right) = \text{Cov}\left(Y_i(x),\ Y_j(y)\right) + f_i(x)\ \delta_{ij}\ \delta_{xy}$$

Par conséquent, on retrouve le même type de résultats que pour le cas univariable. En particulier, on peut filtrer $Y_k(x)$ avec l'estimateur $Y_k^*(x) = \sum_{i=1}^{n} \sum_{\alpha=1}^{N} \lambda_i\ Z_i(x_\alpha)$. Les équations de krigeage pour le cas localement stationnaire sont

$$\sum_{j=1}^{n} \sum_{\alpha=1}^{N} \lambda_j^\alpha\ \text{Cov}\ [Z_j(x_\alpha),\ Z_i(x_\beta)] + \mu_i \quad = \text{Cov}\ [Z_i(x_\beta),\ Y_k(x)]$$

$$\sum_{\alpha=1}^{N} \lambda_j^\alpha \quad = \delta_k^j$$

et de la même manière, on peut estimer les composantes principales $X_p^u(x)$ par

$$X_p^u(x) = \sum_{i=1}^{n} \sum_{\alpha=1}^{N} \lambda_i^\alpha\ Z_i(x_\alpha)$$

et on trouve les équations

$$\sum_{j=1}^{n} \sum_{\alpha=1}^{N} \lambda_j^\alpha\ \text{Cov}\ [Z_j(x_\alpha),\ Z_i(x_\beta)] + \mu_i \quad = A_{iu}^p\ \text{Cov}\ [Z_i(x_\beta), X_p^u(x)]$$

$$\sum_{\alpha=1}^{N} \lambda_j^\alpha \quad = 0$$

EXEMPLE

L'exemple suivant a déjà été présenté dans [10] et [11]. Ici nous avons trois éléments Co, Ti et W. Les résultats d'un cokrigeage comparés à ceux d'un krigeage

sont présentés au Tableau 2.2. Les images initiales et les images filtrées par cokrigeage sont présentées sur les figures 2.29 et 2.30 respectivement pour le cobalt et sur les figures 2.31 et 2.32 respectivement pour le tungstène.

Tableau 2.2.

Rapport signal/bruit			
	Cobalt	Titane	Tungstène
Initial	0.82	14.74	11.74
Univariable	5.92	36.56	24.82
Multivariable	7.11	43.05	28.01

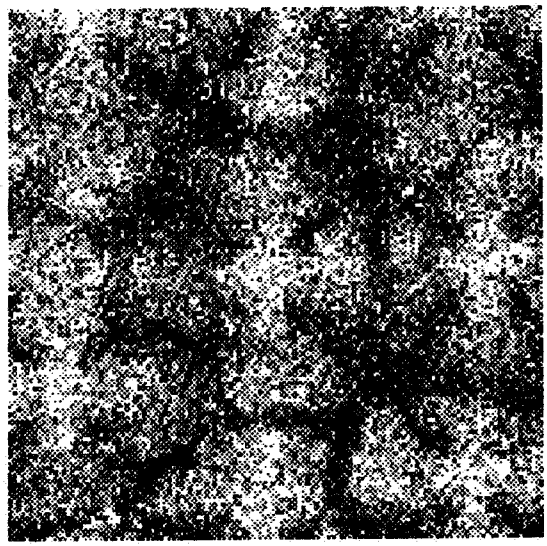


Figure 2.29 Cobalt

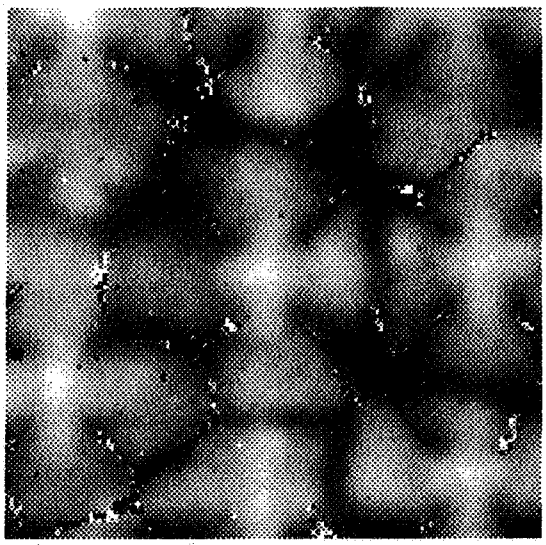


Figure 2.30 Cobalt filtré

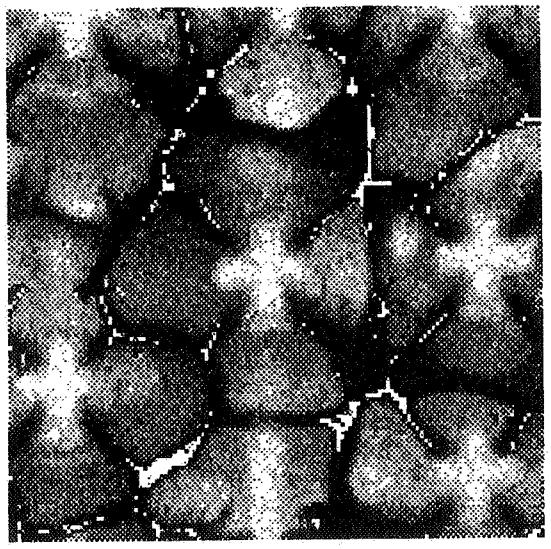


Figure 2.31 Tungstène

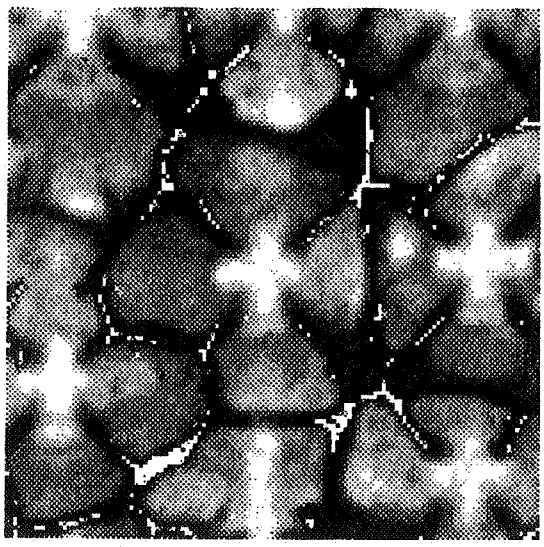


Figure 2.32 Tungstène filtré

Le variogramme étant modélisé avec un effet de pépite, un modèle sphérique de portée 4 et un modèle cubique de portée 17, on peut estimer les composantes principales pour les structures à 4 et 17 respectivement. Notons que pour chaque structure, il y aura trois composantes, mais qu'on ne montre ici que les deux premières. Sur les figures 2.33 et 2.34 on trouve les composantes pour la première structure et sur les figures 2.35 et 2.36 celles de la deuxième structure.

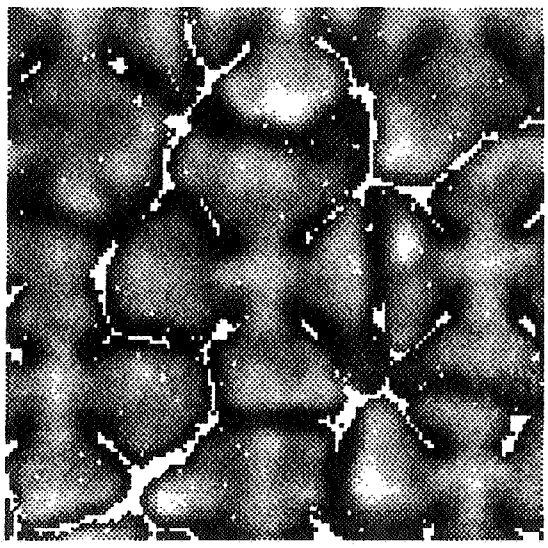


Figure 2.33 Première structure.
Première composante.

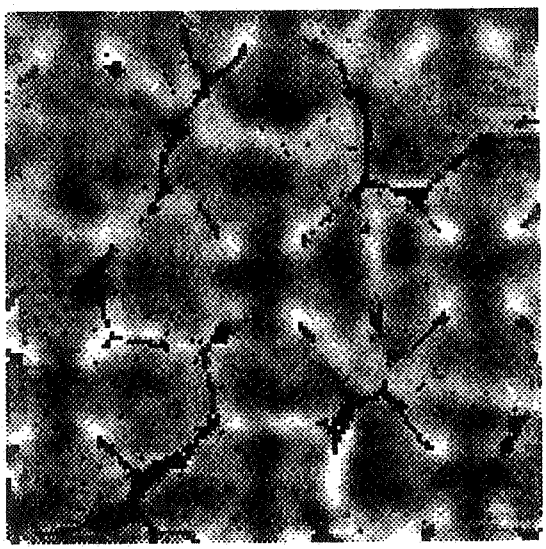


Figure 2.34 Première structure.
Deuxième composante.

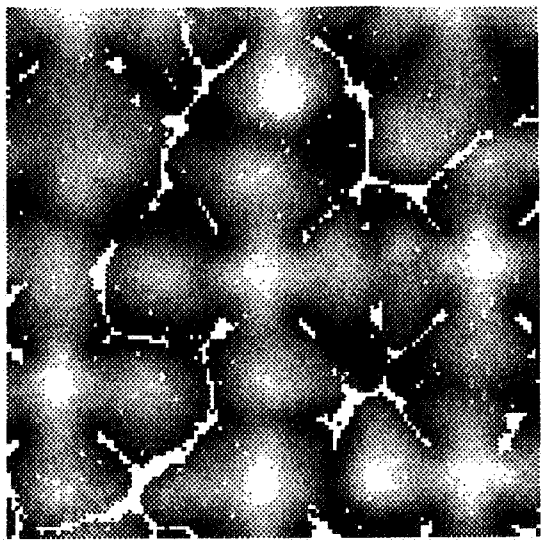


Figure 2.35 Deuxième structure. Première composante.

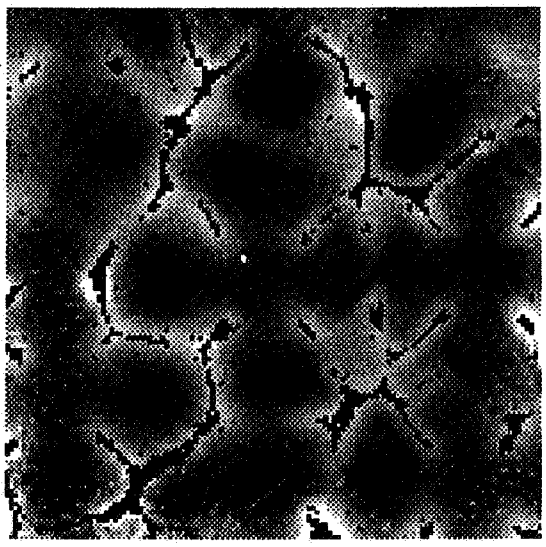


Figure 2.36 Deuxième structure. Deuxième composante.

La décomposition de la variabilité des trois éléments selon les composantes est présentée au Tableau 2.3

Tableau 2.3.

Pourcentage de la variance expliquée pour les CP				
Structure	Composante	Co	Ti	W
1	1	14.04	23.95	30.04
1	2	7.04	0.09	0.04
2	1	68.70	57.92	57.35
2	2	0.05	7.13	1.18

2.3. — ESTIMATION D'UN ÉLEMENT NON OBSERVÉ

Pour des raisons techniques liées à la microsonde, on ne peut obtenir des images du carbone qu'avec difficulté, et par conséquent on ne fait pas d'acquisition dans une étude typique. De même il est parfois intéressant d'avoir un estimateur du carbone.

Si l'on a fait plusieurs études d'apprentissage pour lesquelles le carbone est connu, et si l'on a pu faire l'hypothèse que ses acquisitions sont des résolutions indépendantes d'une fonction aléatoire, alors on peut calculer un modèle de variogramme moyen pour ces études, et appliquer ce modèle pour l'estimation du carbone dans d'autres études où celui-ci n'est pas connu.

Supposons que nous ayons les éléments $[Y_i(x)]$ $Y_1(x)$ étant le carbone, et que nous ayons trouvé des modèles de corégionalisation $\gamma_{ij}^{Y}(h)$ pour les éléments et $\gamma_{ij}^{Z}(h)$ pour les données observées, alors on peut considérer l'estimateur $Y_1^*(x) = \sum_{i=2}^{n} \sum_{a=1}^{N} \lambda_i^a \, Z_i(x_a)$.

Puisque la moyenne de Y_1 n'est pas connue exactement, il vaut mieux, comme dans l'estimation des composantes, n'estimer que le résidu $Y_1^*(x) - m_{Y_1}$. Par conséquent

$$\sum_{a=1}^{n} \lambda_i^a = 0$$

Il est facile de trouver les équations du krigeage

$$\sum_{j=2}^{n} \sum_{a=1}^{N} \lambda_j^a \, \text{Cov} \, [Z_j(x_a), \; Z_i(x_\beta)] + \mu_i \; = \text{Cov} \, [Z_i(x_\beta), \; Y_1(x)]$$

$$\sum_{a=1}^{N} \lambda_j^a \qquad\qquad\qquad\qquad\qquad\qquad = 0$$

EXEMPLES

1. – On applique ici la méthode au problème d'estimation du carbone. On n'a utilisé qu'une étude d'apprentissage, et bien que cela ne soit pas vraiment suffisant, on a néanmoins obtenu des résultats encourageants. On voit sur la figure 2.37 l'élément du carbone de l'étude d'apprentissage et sur la figure 2.38 celui que nous allons estimer.

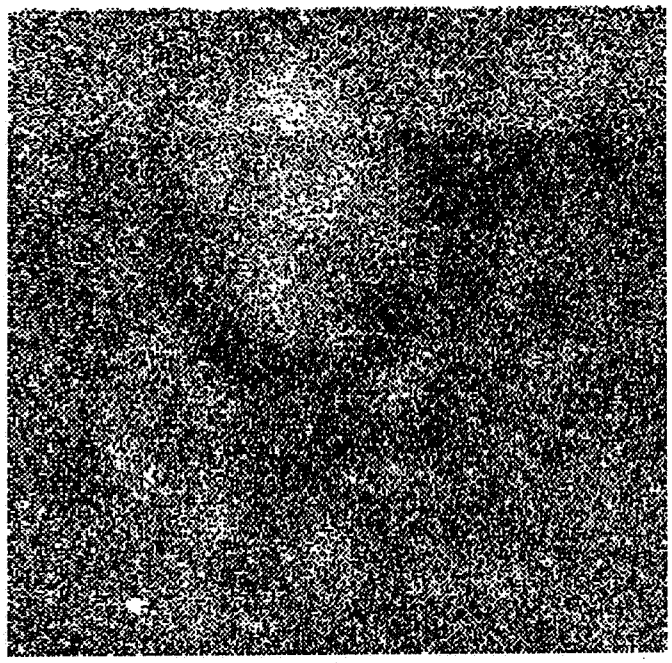


Figure 2.37 Carbone, étude d'apprentissage.

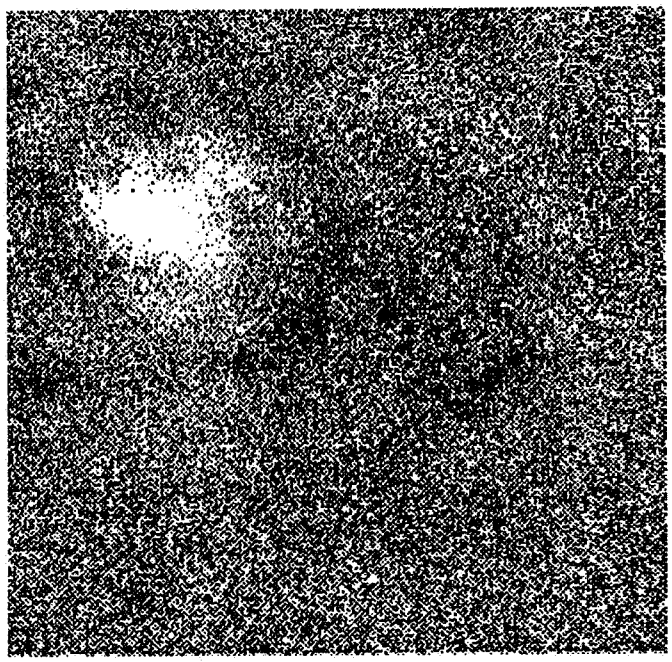


Figure 2.38 Carbone, à estimer.

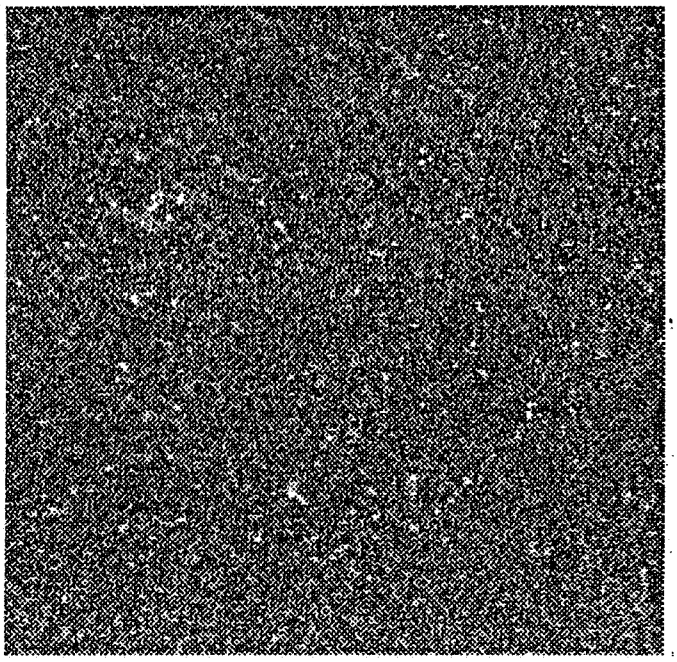


Figure 2.39 Premier élément, *Nb*

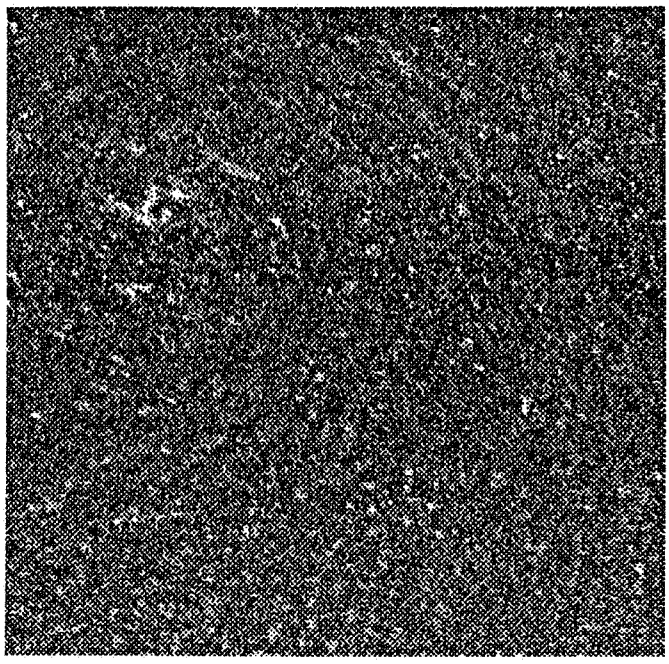


Figure 2.40 Deuxième élément, *P*

Nous disposons de deux autres éléments, Nb et P, qui sont les seules données utilisées pour l'estimation du carbone et qui sont montrés sur les figures 2.39 et 2.40 respectivement. Le résultat de l'estimation apparaît à la figure 2.41 et on peut le comparer avec l'image réelle (figure 2.38) et l'image réelle filtrée (figure 2.42). Il est à noter que l'estimation n'est pas très lisse. Ceci est dû au choix du voisinage de krigeage. En effet la variabilité du carbone est dominée par une structure à 65 pixels, et il nous faudra donc un très grand voisinage pour le filtrage. Au lieu d'utiliser un voisinage de 130 × 130 nous avons fait un sous-échantillonnage des points dans le voisinage et par conséquent la variabilité n'est pas suffisamment réduite. Une méthode permettant de choisir un voisinage de krigeage "optimal" proposé dans [12] peut être intéressante dans ce contexte.

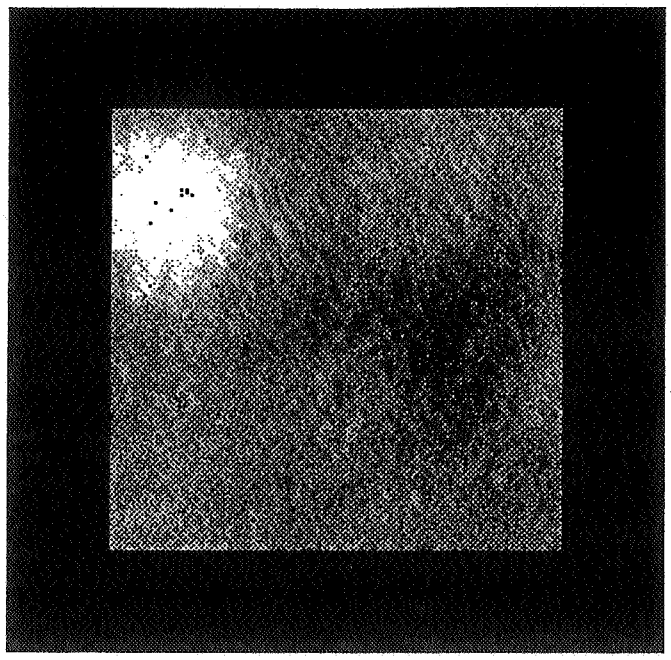


Figure 2.41 Carbone estimé

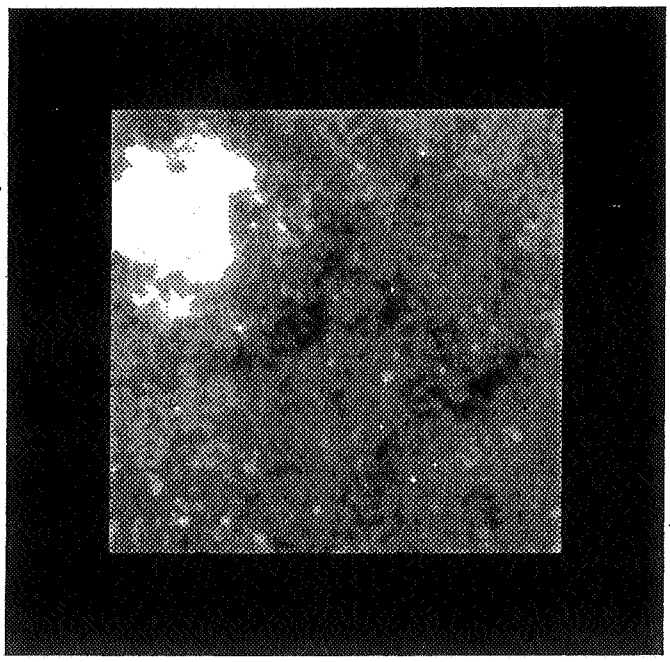


Figure 2.42 Carbone réel filtré

2. – Le relief d'une tôle brute destinée à fabriquer une automobile est représentée à la figure 2.43. A la figure 2.44, on voit la tôle finale après six couches de peintures. (Images de Peugeot S.A.). Il est intéressant de voir si le krigeage peut estimer, à partir de la tôle brute, la tôle telle qu'elle sera à la sortie du processus de peinture. On peut considérer la tôle peinte comme un élément inconnu et l'estimer à partir de la tôle brute. L'estimateur est donné à la figure 2.45. Bien qu'il ne corresponde pas exactement à la vraie tôle peinte, il donne une idée générale de la qualité de la peinture. Il se trouve que pour cette étude [13],[14], le krigeage était capable de prévoir approximativement les étapes intermédiaires entre tôle brute et tôle finale (nombre de couches égal à 1, 2, 3, 4, 5). On ne disposait pas des données pour pousser l'étude plus loin, mais on peut au moins envisager la possibilité que le krigeage peut être utilisé pour prévoir l'effet de la rugosité initiale de la tôle sur son aspect après peinture.

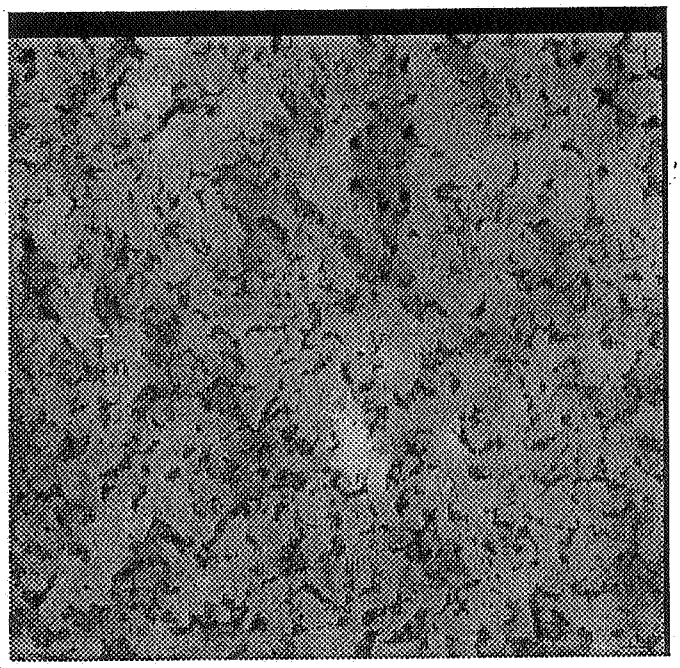


Figure 2.43 Tôle brute

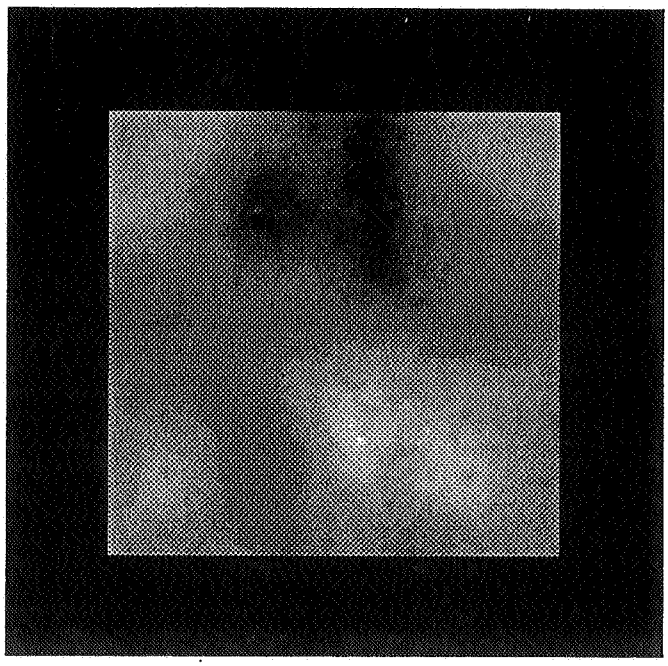


Figure 2.44 Tôle peinte

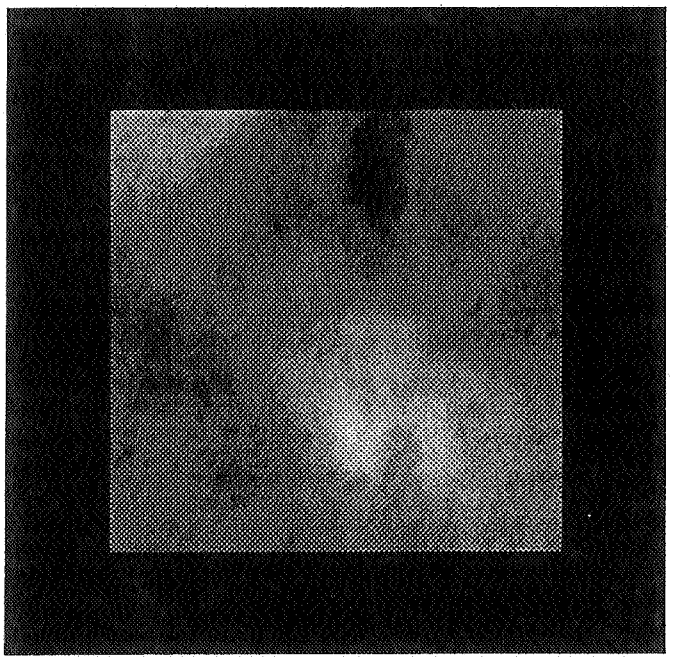


Figure 2.45 Estimation du tôle peinte

CHAPITRE III

ESTIMATION DE L'HISTOGRAMME SOUS-JACENT

3.0. — INTRODUCTION

Un problème important qui se pose au métallurgiste est de connaître l'histogramme sous-jacent d'un échantillon de matériau. Pour un support v, on veut connaître $F(x) = P[Y(v) \leq x]$ pour l'échantillon, où $Y(v) = 1/v \int_v Y(x)\, dx$ est la moyenne de $Y(x)$ dans v. En général, le support correspond à la taille du faisceau de la microsonde. Si l'on connaît l'histogramme $F(x)$, on peut calculer le pourcentage de valeurs qui dépassent un seuil donné et on peut construire des mesures globales du taux de ségrégation. Dans la pratique, on ne connaît pas $F(x)$ et il faut l'estimer. En effet, nous avons des données sur une grille régulière de taille d. Le support v est plus petit que d pour un bon nombre d'applications, et en particulier pour les applications traitées ici. Si les données étaient connues sans bruit, on aurait un estimateur $\hat{F}(x)$ de la loi $F(x)$. Malheureusement, au lieu de mesurer la concentration vraie en chaque point, on mesure une variable de Poisson dont le paramètre est la concentration vraie. Ceci déforme non seulement l'image mais aussi l'histogramme. Notre but est d'avoir un estimateur de $F(x)$ à partir de l'histogramme bruité qu'on appelle $g(n)$.

Dans ce chapitre, nous appliquons la méthode de régularisation pour l'estimation de l'histogramme sous-jacent. La régularisation a été développée pour résoudre des problèmes mal posés, dans le sens où elle donne des estimations stables [1]. On donne ici une idée intuitive des problèmes mal posés linéaires et de leur solution par la méthode de régularisation dans le contexte du problème de la poissonisation. Un traitement plus détaillé de la méthode est présenté aux chapitres IV et V. Dans certains cas, la méthode de régularisation est équivalente au krigeage. On traitera ici du cas stationnaire pour lequel il existe une démonstration très facile. Le cas général sera traité au chapitre V.

3.1. — POISSONISATION D'UNE LOI

Soit $F(x)$ une loi de probabilité associée à la mesure $\mu(x)$ qui représente la concentration vraie d'un échantillon. Appelons $G(n) = P(\mu)$ (on utilise aussi la notation $G = P(F)$) la poissonisation de F. Il s'agit d'une loi de probabilité discrète donnée par la loi de probabilité totale

$$G(n) = \int e^{-x} \frac{x^n}{n!} \, d\mu(x) \tag{3.1}$$

$G(n)$ est la loi qu'on observerait si l'on pouvait prendre une infinité d'échantillons selon la loi $F(x)$ sujette à un bruit poissonien.

Malheureusement on ne dispose que d'un nombre fini d'échantillons. Ce qui est disponible en pratique est une approximation $g(n)$ de $G(n)$. En général $g(n)$ n'est la poissonisation d'aucune loi $F(x)$ et il n'existe pas de méthode simple qui nous permette de retrouver $F(x)$ à partir de $g(n)$. Toutefois on a le résultat suivant.

THÉORÈME 3.1. — Si $G(n)$ est une poissonisation d'une loi $F(x)$, $G = P(F)$, alors F est unique.

Ce théorème, dont la démonstration est donnée au chapitre suivant, nous montre que si l'on peut associer notre échantillon $g(n)$ à une loi poissonisée $\hat{G}(n)$, alors cette loi $\hat{G}(n)$ peut être associée à une loi unique $\hat{F}(n)$. Une approche naturelle pour résoudre le problème sera donc de chercher la loi $\hat{G}(n)$ "la plus proche" de $g(n)$.

Si la loi sous-jacente est discrète (dans la méthode utilisée ici, on modélise F par une loi discrète)

$$F(dx) = \sum_{i=1}^{n} p_i \, \delta_{x_i}(dx) \tag{3.2}$$

où $p_i \geq 0$ et $\sum_{i=1}^{n} p_i = 1$.

L'équation (3.1) prend la forme

$$G(n) = \sum_{i=1}^{n} e^{-x_i} \frac{x_i^n}{n!} \, p_i \tag{3.3}$$

3.2. — *ESTIMATION PAR MOINDRES CARRÉS*

On propose une estimation de $F(x)$ de la forme

$$\hat{F}(dx) = \sum_{i=1}^{n} \hat{p}_i \, \delta_{x_i}(dx)$$

d'où

$$\hat{G}(n) = \sum_i e^{-x_i} \frac{x_i^n}{n!} \hat{p}_i$$

et on essaye maintenant de trouver la loi $\hat{G}(n)$ qui soit la plus proche de $g(n)$ au sens de $\ell^2(\omega)$. C'est-à-dire qu'on choisit $\hat{G}(n)$ telle que

$$d = \sum_0^\infty \left(g(n) - \hat{G}(n)\right)^2 \omega(n) \tag{3.4}$$

est un minimum. Les $\omega(n)$ sont des pondérateurs. Un bon choix est $\omega(n) = 1/g(n)$ si $g(n)$ est assez grand (des calculs pratiques montrent que ce choix de $\omega(n)$ semble assez bon quand $\sup_n g(n) > 300$). Pour un justificatif heuristique de ce choix de $\omega(n)$, voir § 3.4. Si $g(n)$ n'est pas suffisamment grand, il vaut mieux choisir $\omega(n) = 1$ (voir § 3.5).

Appliquons le développement de $\hat{G}(n)$ dans l'équation (3.4). Nous obtenons un estimateur de $\hat{p}_n$ comme solution de

$$\operatorname*{Min}_{\hat{p}_n} d = \sum_0^\infty \left(g(n) - \sum_i P_{in} \hat{p}_i \right)^2 \omega(n) \tag{3.5}$$

tel que $\hat{p}_i \geq 0$ et $\sum_{i=1}^{n} \hat{p}_i = 1$, où $P_{in} = P(i,n) = e^{-x_i} \frac{x_i^n}{n!}$ est la matrice de poissonisation.

On peut écrire l'équation (3.5) sous la forme suivante.

$$\operatorname*{Min}_{\hat{p}_n} = K - 2\sum_i k_i \hat{p}_i + \sum_{i,j} d_{ij} \hat{p}_i \hat{p}_j$$

$$\text{tel que } \hat{p}_i \geq 0 \text{ et } \sum p_i = 1$$

$$\text{où } K = \sum_n g(n)^2\omega(n)$$

$$k_i = \sum_n P_{in}\, g(n)\omega(n) \text{ et } d_{ij} = \sum_n P_{in}P_{jn}\omega_n$$

Ceci est un problème de programmation quadratique. Puisqu'il s'agit d'une projection sur l'ensemble convexe fermé donné par les contraintes, les équations (3.5) ont une solution unique.

3.2.1. — Problèmes mal posés

La solution donnée par cette méthode n'est malheureusement pas un bon estimateur de $F(x)$ dans la plupart des cas. Si nous avons la loi $F(x)$ donnée à la figure 3.1, et l'observation bruitée à la figure 3.2, on obtient l'estimateur par moindres carrés de la figure 3.3.

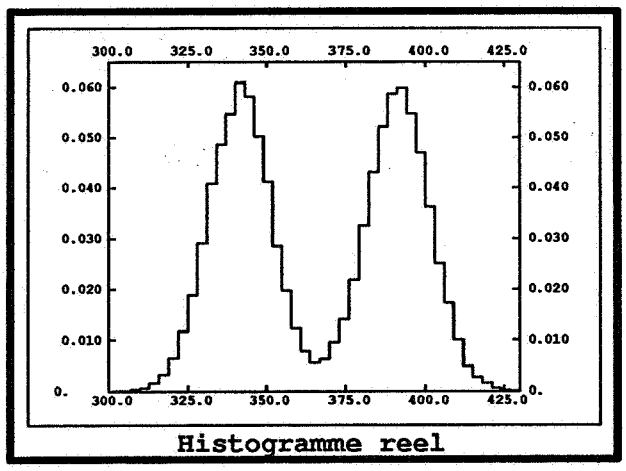


Figure 3.1 Histogramme réel

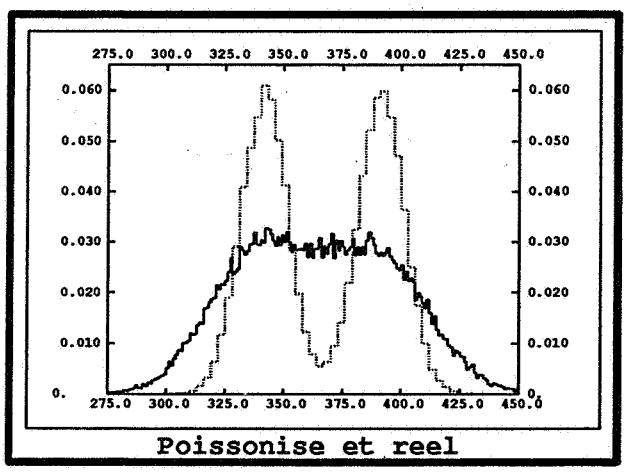


Figure 3.2 Poissonisé ... trait continu
Réel ... pointillé

La définition rigoureuse d'un problème mal posé sera présentée au chapitre suivant. On essaie ici de donner une définition intuitive. Notons que la poissonisation s'exprime en termes d'une équation intégrale (3.1).

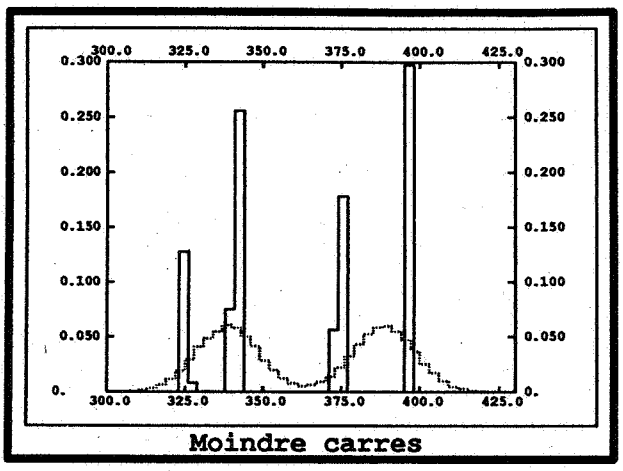


Figure 3.3 Estimateur ... trait continu
Réel ... pointillé

L'opération inverse ressemble donc à une différentiation. Mais on sait que deux fonctions différentiables qui sont très proches au sens de L^2 peuvent avoir des dérivées très différentes au sens de L^2. Dans les figures 3.4 et 3.5, on voit deux courbes qui sont presque identiques. Les dérivées d'ordre 3 sont présentées aux figures 3.6 et 3.7 sur la même échelle. On voit bien le même type de résultat pathologique qu'on avait observé pour l'estimateur de moindres carrés. Ceci nous donne la définition intuitive d'un problème mal posé. Un problème est mal posé si la solution existe mais n'est pas stable par rapport aux petites variations des données initiales.

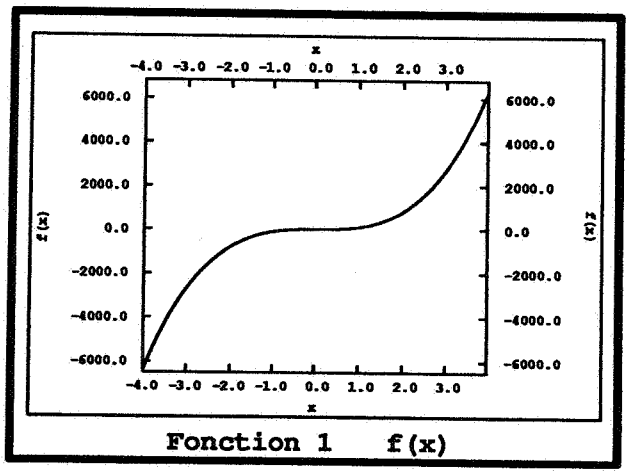


Figure 3.4 $f(x)$

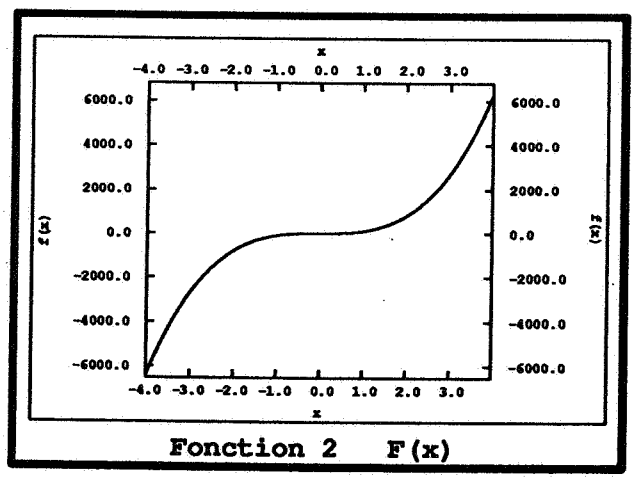


Figure 3.5 $F(x)$

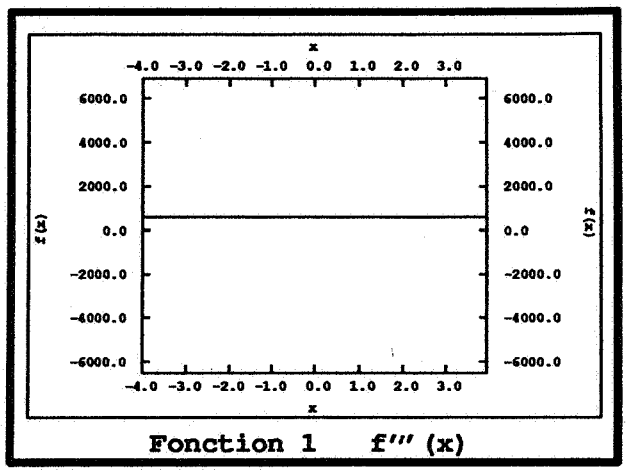


Figure 3.6 $f'''(x)$

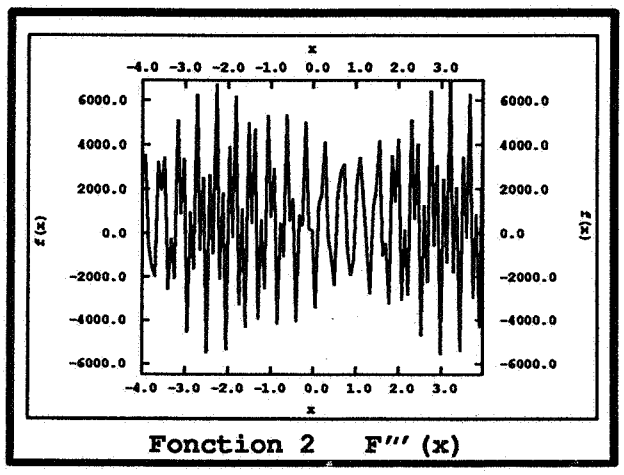


Figure 3.7 $F'''(x) = f(x) + \epsilon(x)$

3.2.2. — *Maximum de vraisemblance*

On démontrera au chapitre suivant l'existence d'un choix de pondérateurs $\omega(n)$, tel que la solution du système (3.5) satisfasse à $\hat{G}(n) = \omega(n)$. On appelle des pondérateurs ayant cette propriété des point fixes. De plus on démontre que le maximum de vraisemblance existe, est unique, et peut être calculé par le système (3.5) pour un point fixe $\omega(n)$. On peut conclure que la méthode du maximum de vraisemblance est instable pour le problème de poissonisation.

3.3. — *LES SOLUTIONS RÉGULARISÉES*

Tikhonov [1] a proposé une solution pour des problèmes mal posés de la façon suivante. Au lieu de minimiser d dans l'équation (3.5), nous minimisons

$$Q = d + \lambda\, \Omega\, (F) \tag{3.6}$$

sous les mêmes contraintes $\hat{p}_i \geq 0$ et $\sum_{i=1}^{n} \hat{p}_i = 1$, où Ω est une fonctionnelle positive et continue sur l'espace des lois initiales de probabilité. Dans les applications, on choisit souvent pour Ω soit l'entropie $\Omega(F) = -\sum \hat{p}_i \log \hat{p}_i$, soit une forme quadratique positive (par exemple un spline).

On ne considère ici que le deuxième cas, et donc on peut admettre que Ω admet une représentation matricielle $\Omega = (\Omega_{ij})$. On appelle le paramètre λ

paramètre de régularisation. Il détermine le degré de régularisation à effectuer. L'interprétation donnée par Tikhonov est la suivante. La minimisation des équations (3.6) équivaut à la minimisation de $\Omega(F)$ avec la contrainte supplémentaire

$$\sum (\hat{G}(n) - g(n))^2 \, \omega(n) = \epsilon \tag{3.7}$$

où ϵ est une fonction biunivoque de λ et représente la distance entre $G(n)$ et $g(n)$. Autrement dit, parmi toutes les lois possibles qui satisfont à (3.7), choisissons celle qui minimise $\Omega(F)$.

Pour pouvoir mettre en oeuvre la méthode de régularisation, il faut connaître $\Omega(F)$ et le paramètre de régularisation λ. En général on choisit $\Omega(F)$ a priori et on ajuste λ par la méthode de validation croisée. A la section 3.3.2, nous discuterons brièvement la méthode de validation croisée qui s'applique à tous les problèmes mal posés linéaires puis nous présenterons une méthode qui nécessite deux acquisitions mais qui semble meilleure que la validation croisée. Mais d'abord le lien entre le krigeage et la régularisation est noté dans le cas stationnaire.

3.3.1. — Le krigeage et la régularisation

Le problème de régularisation sans contrainte, soit :

$$\text{Minimiser} \quad Q = \sum_{n,m} \left(g(n) - \hat{G}(n)\right) B_{nm} \left(g(m) - \hat{G}(m)\right) + \sum_{i,j} \hat{p}_i \, \Omega_{ij} \, \hat{p}_j$$

où B_{nm} est une matrice symétrique définie positive (jusqu'à présent nous avons pris $B_{nm} = \delta_{nm}\omega_n$ mais dans un cas général on peut avoir besoin d'une matrice B_{nm} définie positive quelconque), et Ω_{ij} une matrice positive, est équivalent à un problème de krigeage. Dans cette section, on donne une démonstration dans un cas simple, quand Ω_{ij} est définie positive et quand les deux matrices Ω et B sont des matrices finies. Le cas général sera discuté au chapitre V.

Décrivons maintenant le problème de krigeage. Soit Y_j une variable aléatoire pour chaque j et

$$Z_n = \sum_i P(i,n) \, Y_i + \epsilon_n$$

est tel que

$$\mathrm{Cov}(Z_n, Z_m) = \sum_{i,j} P(i,n)\ \mathrm{Cov}(Y_i, Y_j)\ P(j,m) + \Sigma_{nm}$$

où Σ_{nm} est une matrice définie positive (Cette relation est vérifié si par exemple ϵ_n est indépendant de Y_i). Le problème du krigeage consiste à estimer Y_i à partir des données $\{Z_i\}$, $Y^{*i} = \sum \lambda_n Z_n$ tel que $\mathrm{Var}[Y_i - Y_i^*]$ est un minimum.

THÉORÈME 3.2. – Si Ω est strictement positive alors les problèmes de régularisation linéaire et de krigeage ont les mêmes solutions si l'on identifie p_i à Y_i et Z_j à G_j, et si $\mathrm{Cov}\ [Y_i, Y_j] = [\Omega^{-1}]_{ij}$ et $\mathrm{Cov}\ [\epsilon_n, \epsilon_m] = [B^{-1}]_{nm}$.

DÉMONSTRATION. – En effet, pour le problème de régularisation

$$Q = \sum_{n,m} \left(g(n) - \sum_i P(i,n)\ \hat{p}_i \right) B^{nm} \left[g(m) - \sum_j P(j,m)\ \hat{p}_j \right] + \sum_{i,j} \hat{p}_j\ \Omega^{ij}\ \hat{p}_j$$

et l'équation d'Euler $\dfrac{\partial Q}{\partial \hat{p}_i} = 0$ donne

$$\sum_j \left[\sum_{n,m} P(i,n)\ B^{nm}\ P(j,m) + \Omega^{ij} \right] \hat{p}_j = \sum_{n,m} P(i,n)\ B^{nm}\ g(m)$$

d'où $$\hat{p}_j = \sum_i \left[\sum_{n,m} P(i,n)\ B^{nm}\ P(j,m) + \Omega^{ij} \right]^{-1} \left(\sum_{n,m} P(i,n)\ B^{nm}\ g(m) \right)$$

L'inverse existe puisque Ω^{ij} est positive définie, et B étant définie positive, on a également P^tBP positive. Donc

$$\hat{p} = \left[P^tBP + \Omega \right]^{-1} P^tBg$$

et pour le problème de krigeage avec $K_{ij} = \mathrm{Cov}\ [Y_i, Y_j]$ et $\Sigma_{nm} = \mathrm{Cov}\ [\epsilon_n, \epsilon_m]$, les équations de krigeage pour l'estimateur $Y_k^* = \sum_n \lambda_n^k\ Z_n$ sont

$$\sum_n \left[\sum_{i,j} P_{in}\ K_{ij}\ P_{jm} + \Sigma_{nm} \right] \lambda_n^k = \sum_j P_{jm}\ K_{kj}$$

où dans la notation matricielle

$$\lambda^T[P\ K\ P^t + \Sigma] = K\ P^t$$

et

$$Y^* = \lambda^T Z \ \Rightarrow\ Y^* = K\ P^t[P\ K\ P^t + \Sigma]^{-1}\ Z$$

Mais la relation suivante est vraie pour toutes les matrices P, K et Σ quand K et Σ ont des inverses.

$$K\ P^t[P\ K\ P^t + \Sigma]^{-1} = [P^t\ \Sigma^{-1}\ P + K^{-1}]^{-1}\ P^t\ \Sigma^{-1}$$

Ceci est vrai parce que si l'on multiplie à gauche par $(P^t\ \Sigma^{-1}\ P + K^{-1})$ et à droite par $(P\ K\ P^t + \Sigma)$ alors l'égalité est vraie

$$\Leftrightarrow \qquad [P^t\ \Sigma^{-1}\ P + K^{-1}]\ K\ P^t = P^t\ \Sigma^{-1}\ [P\ K\ P^t + \Sigma]$$

$$\Leftrightarrow \qquad P^t\ \Sigma^{-1}\ P\ K\ P^t + P^t = P^t\ \Sigma^{-1}\ P\ K\ P^t + P^t$$

donc sous les conditions du théorème quand $Y_j = p_j$, $Z_n = g(n)$ et si $K^{-1} = \Omega$ et $\Sigma^{-1} = B$, les deux problèmes ont la même solution. ■

On peut également considérer le cas où Ω est seulement positive au lieu d'être strictement positive. La solution n'existe pas toujours dans ce cas, mais il reste équivalent à un problème de krigeage, et l'existence de la solution du problème spline est assurée quand le krigeage a une solution. Ce cas est traité au chapitre V et on se contente ici de donner quelques résultats qui seront démontrés au chapitre V.

Admettons que f^ℓ soit une base de l'espace linéaire des vecteurs p, tel que $\Omega(p) = 0$. Soit

$$Z_n = \sum_i P_{in}\ Y_i + \epsilon_i$$

Il se trouve qu'une inverse généralisée de Ω, c'est-à-dire une matrice K donnée par $\Omega K \Omega = \Omega$, est conditionnellement définie positive par rapport aux fonctions de dérive f^ℓ, et peut donc être utilisée comme covariance généralisée. On veut estimer Y_x par $Y_x = \Sigma\ \lambda_n\ Z_n$ tel que $(Y_x - \Sigma\ \lambda_n\ Z_n)$ est une combinaison linéaire autorisée, donc les λ_n doivent satisfaire à $\Sigma_{i,n}\ \lambda_n\ P_{in}\ f_x^\ell = f_x^\ell$. Si $\Sigma = B^{-1} = \text{Cov}\ (\epsilon_n, \epsilon_m)$, alors les équations de krigeage qui déterminent λ_n sont

$$\sum_n \lambda_n \Big(\sum_{i,j} P_{in}\ K_{ij}\ P_{jm} + \Sigma_{nm}\Big) + \sum_{i,\ell} \mu_\ell\ P_{im}\ f_i^\ell = \sum_i P_{im}\ K_{ix}$$

$$\sum_{i,n} \lambda_n P_{in} f_i^\ell = f_x^\ell$$

Avec K l'inverse généralisée de Ω, ces équations ont une solution si et seulement si la condition suivante est vraie

$$\sum_{i,\ell} b_\ell P_{in} f_i^\ell = 0 \quad \forall n \Rightarrow b_\ell = 0 \quad \forall \ell$$

Le problème de régularisation a une solution avec la même condition et admet les mêmes solutions.

Parfois dans la littérature sur la régularisation, on utilise une estimation *a priori* qu'on veut améliorer. Par exemple dans le problème de poissonisation, on peut avoir un estimateur préliminaire de la loi sous-jacente qu'on peut utiliser pour améliorer la solution. Cette loi *a priori* peut être choisie en fonction de connaissances physiques du problème, ou dans notre cas, parce que l'histogramme est l'histogramme d'une image, on peut considérer comme loi *a priori* l'histogramme obtenu par un filtrage d'image bruitée.

Supposons que la loi *a priori* soit $\{q_i\}$. Alors le nouveau problème de régularisation est :

$$\underset{\hat{p}_i}{\text{Min}}\ Q = \sum_{n;m} \left(g(n) - \hat{G}(n)\right) B_{nm} \left(g(m) - \hat{G}(m)\right) + (\hat{p}_i - q_i)\ \Omega_{ij}\ (\hat{p}_j - q_j)$$

On peut donner une interprétation de cette procédure en termes de krigeage. Il s'agit d'un krigeage en présence d'une dérive qui est fixée. Ceci est proche d'une technique appelée "technique de la dérive externe" et qui est souvent utilisée en géostatistique pétrolière [2]. En particulier, elle se réduit à la dérive externe si l'on minimise

$$\underset{\hat{p}_i,a}{\text{Min}}\ Q = \sum_{n;m} \left(g(n) - \hat{G}(n)\right) B_{nm} \left(g(m) - \hat{G}(m)\right) + (\hat{p}_i - aq_i)\ \Omega_{ij}\ (\hat{p}_j - aq_j)$$

c'est-à-dire si l'on utilise la "forme" de $\{q_i\}$ comme dérive et non sa valeur exacte. Pour plus de détails, se reporter à la note IV après le théorème 5.3 au chapitre V.

3.3.2. — La validation croisée

Supposons la forme quadratique positive Ω déjà choisie. Nous cherchons λ par validation croisée. Choisissons une grille des valeurs de λ, et fixons un de

ces λ. Puis pour chaque n on résoud les équations (3.6) sans utiliser la valeur de $g(n)$. On peut alors estimer la valeur $g(n)$ par $\hat{g}_\lambda(n)$. Si l'on applique cette procédure pour chaque n et chaque λ, on obtient la fonction de validation croisée

$$\mathrm{CV}(\lambda) = \sum_n (g(n) - \hat{g}_\lambda(n))^2$$

Cette procédure est évidemment beaucoup trop coûteuse à mettre en pratique, et on est amené à considérer l'approximation suivante. Supposons qu'on oublie pour l'instant les contraintes $p_i \geq 0$ et $\Sigma\, p_i = 1$ et supposons de plus que la matrice de poissonisation $P(i,n)$ soit tronquée à M termes tels que $\sum_{n=1}^{M} P(i,n) \geq 1 - \epsilon \quad \forall\, i$ et pour ϵ petit. Nous avons le résultat suivant.

THÉORÈME 3.3. – Si $\hat{p}_i$ minimise Q,

où
$$Q = \sum_n \omega(n)\, (g(n) - \hat{G}_\lambda(n))^2 + \sum_{i,j} \hat{p}_i\, \Omega_{ij}\, \hat{p}_j$$

et où
$$\hat{G}_\lambda(n) = \sum_i P(i,n)\, \hat{p}_i = \sum_i e^{-x_i} \frac{x_i^n}{n!}\, \hat{p}_i$$

et si $\hat{f}_\lambda^k(i)$ minimise Q_k, où

$$Q_k = \sum_{n \neq k} \omega(n) \left(g(n) - \hat{G}_\lambda^k(n)\right)^2 + \sum_{i,j} \hat{f}_\lambda^k(i)\, \Omega_{ij}\, \hat{f}_\lambda^k(j)$$

et où
$$\hat{G}_\lambda^k(n) = \sum_i P(i,n)\, \hat{f}_n^k(i)$$

alors la fonction de validation croisée

$$\mathrm{CV}(\lambda) = \sum_k \left(g(k) - \hat{G}_\lambda^k(k)\right)^2$$

prend la forme

$$\mathrm{CV}(\lambda) = \sum_n \left[\frac{g(n) - \hat{G}_\lambda(n)}{1 - H_{nn}^\lambda}\right]^2$$

où H^λ est la matrice définie par

$$H^\lambda = P[P^t B P + \lambda\, \Omega]^{-1}\, P^t B$$

et B est la matrice diagonale des pondérateurs $B_{nm} = \delta_{nm}\,\omega(n)$.

NOTA. – On appelle H^λ la matrice chapeau et on peut trouver une discussion de ses applications aux problèmes mal posés dans [3].

DÉMONSTRATION. – Q a un minimum quand (équation d'Euler)

$$\hat{p}_j = \sum_i \left[\sum_n \omega(n)\, P_{in}\, P_{jn} + \lambda\, \Omega^{ij} \right]^{-1} \left(\sum_m P_{im}\, \omega(m)\, g(m) \right) = \left[P^t B P + \lambda\, \Omega \right]^{-1} P^t B g$$

Alors

$$\hat{G}_\lambda = P\,\hat{p} = P\left[P^t B P + \lambda\, \Omega \right]^{-1} P^t\, B\, g = H^\lambda\, g$$

De la même manière, Q_k a un minimum quand

$$\hat{f}^k_\lambda(j) = \sum_i \left[\sum_{n \neq k} \omega(n)\, P_{in}\, P_{jn} + \lambda\, \Omega^{ij} \right]^{-1} \left(\sum_{m \neq k} P_{im}\, \omega(m)\, g(m) \right)$$

Appelons $A_{ij} = \left[\sum_n \omega(n)\, P_{in}\, P_{jn} + \lambda \Omega^{ij} \right]$ et $C_{ij} = \left[A^{-1} \right]_{ij}$.

Il est clair que

$$\sum_{n \neq k} \omega(n)\, P_{in}\, P_{jn} + \lambda \Omega^{ij} = A_{ij} - \omega(k)\, P_{ik}\, P_{jk} = A - \omega(k)\, q_k\, q_k^t$$

où $q_k = (P_{ik})_{i=1}^M$.

Pour des matrice A, a et b la formule suivante est vraie si tous les inverses existent, (formule de Gauss [4]).

$$(A + a^T b)^{-1} = A^{-1} - A^{-1}\, a^T (I + b\, A^{-1}\, a^T)^{-1}\, b\, A^{-1}$$

La formule de Gauss appliquée ici donne

$$\hat{f}^j_\lambda = \sum_i \left[C_{ij} + \frac{\sum_{i',j'} \omega(k)\, C_{ii'}\, P_{i'k}\, P_{j'k}\, C_{j'j}}{1 - \sum_{i',j'} \omega(k)\, P_{i'k}\, C_{i'j'}\, P_{j'k}} \right] \left(\sum_m P_{im}\, \omega(m)\, g(m) - P_{ik}\, \omega(k)\, g(k) \right)$$

Puisque $H^\lambda = PCP^tB$ et puisque $Cq_k q_k^t C$ est symétrique, il suit que $\hat{G}_\lambda^k(n) = P_{jk}\,\hat{f}_\lambda^j$ est donné par

$$\hat{G}_\lambda^k(n) = \sum_m H_{nm}^\lambda\, g(m) - H_{nk}^\lambda\, g(k) + \sum_m \frac{H_{nk}^\lambda\, H_{km}^\lambda\, g(m)}{1 - H_{kk}^\lambda} - \frac{H_{nk}^\lambda\, H_{kk}^\lambda\, g(k)}{1 - H_{kk}^\lambda}$$

$$= \hat{G}_\lambda(n) - H_{nk}^\lambda\, g(k) + \frac{H_{nk}^\lambda}{1 - H_{kk}^\lambda}\left[\hat{G}_\lambda(k) - H_{kk}^\lambda\, g(k)\right]$$

Maintenant estimons $g(h)$ par $\hat{G}_\lambda^k(k)$ par la validation croisée. Après des calculs élémentaires, on trouve que

$$\hat{G}_\lambda^k(k) = g(k) - \frac{g(k) - \hat{G}_\lambda(k)}{1 - H_{kk}^\lambda}$$

d'où

$$\sum_k (g(k) - G_\lambda^k(k))^2 = \sum_k \left[\frac{g(k) - \hat{G}_\lambda(k)}{1 - H_{kk}^\lambda}\right]^2$$

et le résultat est démontré. ■

Le calcul de la matrice chapeau H^λ est souvent instable et par conséquent $CV(\lambda)$ est une fonction difficile à minimiser pour λ. Une autre fonction, la validation croisée $GCV(\lambda)$, définie par

$$GCV(\lambda) = \frac{\sum (g(n) - G_\lambda(n))^2}{1 - \frac{1}{M}\mathrm{tr}(H^\lambda)}$$

où $\mathrm{tr}(H^\lambda) = \sum_n H_{nn}^\lambda$ est la trace de H^λ, a été proposée dans la littérature pour cette raison [5].

Dans certains cas, $GCV(\lambda)$ est beaucoup plus rapide à calculer parce qu'on peut connaître $\mathrm{tr}(H^\lambda)$ sans calculer H^λ [5,6]. Malheureusement, ces méthodes ne s'appliquent pas au problème de poissonisation et jusqu'à présent, il faut calculer H^λ avant de connaître $\mathrm{tr}(H^\lambda)$. En particulier la fonction GCV, tout en restant plus stable que la fonction CV, n'est pas suffisamment stable pour qu'on puisse appliquer un algorithme de minimisation numérique.

3.3.3. — *Deux acquisitions*

Dans le cas où nous avons deux acquisitions — ce qui est toujours possible si notre histogramme vient d'une image où l'on peut obtenir deux histogrammes en prenant un point sur deux de l'image pour chaque histogramme — on peut considérer une deuxième procédure qui est plus stable que la validation croisée. De plus cette méthode nous permet de travailler avec des contraintes d'inégalité, or cela n'était pas possible avec la validation croisée. Avec la première acquisition, et pour une valeur de λ fixe , on résoud les équations de régularisation pour obtenir $\hat{p}_1$. On peut donc calculer

$$DA_1(\lambda) = \sum\left(g_2(n) - \hat{G}_1(n)\right)^2$$

où $g_2(n)$ est la $n^{ième}$ valeur de la deuxième acquisition et $\hat{G}_1(n) = P(\hat{p}_1)$. De la même manière, on obtient

$$DA_2(\lambda) = \sum\left(g_1(n) - \hat{G}_2(n)\right)^2$$

En pratique, les fonctions $DA_1(\lambda)$ et $DA_2(\lambda)$ sont des fonctions très continues et présentent un minimum unique à $\lambda \geq 0$. Dans ce cas, on peut utiliser une méthode standard de type gradient pour trouver ce minimum. Supposons que λ_1 et λ_2 soient des minima de $DA_1(\lambda)$ et $DA_2(\lambda)$. Alors on choisit $\lambda = (\lambda_1 + \lambda_2)/2$ comme paramètre de régularisation, et on applique la régularisation avec ce choix de paramètre à l'acquisition pour laquelle $DA_i(\lambda)$ est la plus grande dans le voisinage de 0. On note avec quelque regret que ceci nous oblige à travailler uniquement sur la moitié des données disponibles.

Il est étrange de constater qu'il se trouve que cette restriction n'est pas très importante dans un bon nombre de cas. La variabilité de la qualité des estimateurs pour des acquisitions différentes est très élevée, et donc si nous avons choisi celle avec la plus petite variance d'erreur (ce qui est généralement le cas avec notre choix d'acquisition), il peut bien arriver que l'erreur soit du même ordre de grandeur que celle qu'on obtient avec les deux acquisitions. Ceci semble être le cas au moins pour le problème de poissonisation. (cf des exemples pratiques §3.4)

3.3.4. – *La norme d'erreur*

Nous avons suggéré un estimateur de la forme

$$Q = (g - \hat{G})^T B(g - \hat{G}) + \hat{p}^T \Omega \hat{p}$$

où $B_{nm} = \omega(n)\ \delta_{nm}$. Il se trouve que $\omega(n) = 1/(g(n))$ est un assez bon choix de $\omega(n)$ quand $g(n)$ est assez grand. Nous donnons maintenant un justificatif heuristique de ce choix en utilisant le théorème 3.3. D'abord si $\{p_i\}$ est la loi sous-jacente et $G(n) = P(p)$ est le poissonisé, on doit considérer comment $g(n)$, l'observation, varie autour de $G(n)$. Supposons que $G(n)$ soit tronqué à M termes tel que $\sum_{n=1}^{M} G(n) \geq 1 - \epsilon$ pour ϵ petit, et considérons maintenant $g(n)$ comme un échantillon de N points tirés selon la loi discrète et finie $G(n)$. Soit $(X(1),\ldots,X(M))$ le vecteur du nombre des événements prenant $1,\ldots,M$ respectivement. Nous avons donc $g(n) = X(n)/N$. Posons

$$Z_{n\alpha} = \begin{cases} 1 & \text{si le } \alpha^{\text{ème}} \text{ tirage est égal à } n \\ 0 & \text{sinon} \end{cases}$$

Donc $X(n) = \sum_{\alpha} Z_{n\alpha}$. Pour $\alpha \neq \beta$, $Z_{n\alpha}$ et $Z_{n\beta}$ sont indépendants et on a

$$\text{Cov}\ [X(n),\ X(m)] = \sum_{\alpha} \text{Cov}\ [Z_{n\alpha}\ Z_{m\alpha}]$$

Mais $\text{E}[Z_{m\alpha}] = G(m)$ et $\text{E}[Z_{n\alpha}\ Z_{m\alpha}] = 0$ pour $n \neq m$, donc si $n \neq m$, $\text{Cov}\ [X(n),\ X(m)] = -N\ G(n)\ G(m)$ et si $n = m$, le problème est binomial avec pour paramètre $G(n)$, donc

$$\text{Var}\ [X(n)] = N\ G(n)\ (1 - G(n))$$

Finalement on a

$$\text{Cov}\ [g(n),\ g(m)] = \frac{1}{N}\ (G(n)\ \delta_{nm} - G(n)\ G(m))$$

Dans la terminologie du krigeage, nous avons

$$Z(n) = \sum_{i} P(i,n)\ Y_i + \epsilon_n = G(n) + \epsilon_n$$

et il est vrai que

$$g(n) \sim [Z(n) | (G(1),\ldots,G(M)] \sim Z(n) | G(n)$$

Pour pouvoir appliquer le théorème 3.3, il nous faut une expression pour Cov[$Z(n),Z(m)$]. Mais

$$\begin{aligned}
&\mathrm{Cov}[Z(n), Z(m)] = \\
&= \mathrm{E}\big[\mathrm{Cov}\,[Z(n), Z(m) | G(n)\ G(m)]\big] + \mathrm{Cov}\,[\mathrm{E}\,(Z(n)|G(n)), \mathrm{E}\,(Z(m)|G(m)] \\
&= \mathrm{E}\,[\mathrm{Cov}\,(g(n), g(m))] + \mathrm{Cov}\,[\mathrm{E}(g(n)), \mathrm{E}(g(m))] \\
&= \frac{1}{N}\,\mathrm{E}\,[G(n)\ \delta_{nm} - G(n)\ G(m)] + \mathrm{Cov}(G(n),\ G(m)) \\
&= \Sigma_{nm} + (P\ K\ P^t)_{nm}
\end{aligned}$$

et nous avons une décomposition de la matrice de covariance de Z dans la forme requise par le théorème 3.3. La matrice de covariance du "bruit"

$$\Sigma_{nm} = \frac{1}{N}\,\mathrm{E}\,[G(n)\ \delta_{nm} - G(n)\ G(m)]$$

est positive, mais elle n'est pas strictement positive si l'on utilise toutes les valeurs M. Pour pouvoir utiliser cette matrice, nous sommes obligés d'utiliser l'information donnée par $(M - 1)$ termes seulement. Pour M suffisamment grand, ceci ne constitue pas une objection importante. Néanmoins nous n'utilisons que l'approximation de Σ_{nm} qui a été faite en négligeant les termes d'ordre deux. Donc

$$\Sigma_{nm} \sim \frac{1}{N}\,\mathrm{E}\,[G(n)\ \delta_{nm}].$$

En pratique, si $g(n)$ est assez grand (sup $g(n) > 300$ par exemple), on peut remplacer $G(n)$ par $g(n)$ et la matrice de covariance du bruit devient

$$\Sigma_{nm} = \frac{g(n)}{N}\ \delta_{nm}$$

d'où, par le théorème 3.3, on trouve $B_{nm} = \dfrac{N}{g(n)}\ \delta_{nm}$. Comme le facteur constant N est automatiquement pris en compte dans la régularisation, on peut prendre $B_{nm} = \dfrac{1}{g(n)}\ \delta_{nm}$ et on obtient la norme de χ^2 utilisée dans les exemples.

3.3.5. – Estimation de K

Dans le problème de poissonisation, on veut estimer un nombre assez faible de points $\hat{p}_i$. On peut donc estimer tous les points à l'aide d'un seul système

d'équations et la méthode la plus simple est d'utiliser l'approche de la régularisation au lieu de formuler le problème en termes de krigeage pour lequel les matrices seront beaucoup plus grandes. Pour d'autres types de problèmes mal posés tels que la déconvolution, le nombre de points à estimer peut être beaucoup trop grand pour être résolu par un seul système d'équations. Dans un tel cas, il est souvent préférable de résoudre le problème par une formulation du krigeage et d'utiliser un voisinage glissant.

Dans les applications de la méthode de régularisation, le choix de la norme Ω se fait *a priori*, et $\Omega(f) = \int |f''(x)|\, dx$ est choisi assez souvent comme norme de régularisation. Il est bien connu que le choix de cette norme est équivalent au krigeage à l'aide de la covariance généralisée $K(h) = |h|^3$ dans une dimension et $K(h) = |h|^2 \log |h|$ dans deux dimensions. En géostatistique non stationnaire, la covariance généralisée n'est pas choisie *a priori* mais peut être ajustée aux données quand on travaille dans un voisinage glissant. La méthode est déjà très bien établie et il n'y aura pas lieu d'en donner une description complète ici (voir [7] pour une bonne introduction). On propose ici une extension possible de cette méthode aux problèmes mal posés.

Supposons que

$$Z_n = \sum_i A_{in}\, Y_i + \epsilon_n$$

A_{in} étant la matrice de distorsion (pour le problème de poissonisation $A_{in} = P_{in} = e^{-x_i} \frac{x_i^n}{n!}$). Nous admettons que A_{in} n'est pas singulière, mais elle peut être très instable. Ceci est le cas de la poissonisation et plus généralement des problèmes mal posés quand l'espace initial est compact (voir chapitre IV). Pour des pondérateurs $\{\lambda_n\}$, nous avons

$$\sum_n \lambda_n Z_n = \sum_{n,i} \lambda_n A_{in}\, Y_i + \sum_n \lambda_n\, \epsilon_n$$

$$\text{Posons } b_i = \sum_n \lambda_n A_{in} \quad \Rightarrow \quad \sum_n \lambda_n Z_n = \sum_i b_i\, Y_i + \sum_n \lambda_n\, \epsilon_n$$

et si maintenant on admet que Y est une FAI-k et que la matrice de variance/covariance de ϵ_n est connue (on admet généralement que ϵ est stationnaire), alors $\mathrm{Var}(\Sigma\, \lambda_n Z_n)$ existe si $\{b_i\}$ est une CLA, donc

$$\Leftrightarrow \quad \sum_i b_i f_i^\ell = 0 \quad \Leftrightarrow \quad \sum_{i,n} \lambda_n A_{in} f_i^\ell = 0 \qquad (3.8)$$

Choisissons maintenant une famille de mesures $\{\lambda_n\}$ telle que (3.8) est réalisée. Ceci peut se faire en utilisant la même méthode qu'en géostatistique non stationnaire. C'est à dire choissons un point Z_m et estimons le par la méthode des moindres carrés. Donc $Z_m^* = \sum_{n \neq m} \mu_n Z_n$ où les $\{\mu_n\}$ sont des pondérateurs associés à l'estimateur des moindres carrés avec les dérives de la forme $\sum_i A_{in} f_i^\ell$, $\ell = 0,\ldots,k$. Il suit que $\lambda = (\delta_m - \sum \mu_n \delta_n)$ est une mesure telle que $b_\lambda = A\lambda$ est une combinaison linéaire autorisée pour la fonction aléatoire Y.

Si nous avons deux acquisitions $Z^1 = AY + \epsilon_1$ et $Z^2 = AY + \epsilon_2$ et si nous avons une famille de mesures $\{\lambda\}$ telle que chaque $\lambda = (\lambda_1,\ldots,\lambda_N)$ satisfait à (3.8), il est vrai que

$$\text{Cov}\left[\sum \lambda_n Z_n^1, \sum \lambda_m Z_m^2\right] = \text{Var}\left(\sum b_i Y_i\right) = \sum b_i b_j K_{ij}^Y \qquad (3.9)$$

où K_{ij}^Y est la covariance généralisée pour Y.

Prenons K_{ij}^Y de la forme $K_{ij}^Y = \sum_{p=1}^{4} b_p K_{ij}^p$ où les K_{ij}^p sont des modèles simples δ, $|h|$, $|h|^3$ et $|h|^2 \log |h|$. Puisque nous avons maintenant une famille de mesures $\{\lambda\}$, on peut calculer $d_\lambda = (\Sigma \lambda_n Z_n^1)(\Sigma \lambda_m Z_n^2)$ pour chaque mesure λ et d_λ est donc un estimateur de Cov $(\Sigma \lambda_n Z_n^1)(\Sigma \lambda_m Z_n^2)$. Il est possible alors d'utiliser l'équation (3.9) dans une régression pour estimer $[b_p]^4$. Pour plus de détail sur ce point et les tests qu'on doit effectuer, voir [7].

Cette méthode a l'inconvénient d'utiliser des modèles non stationnaires. Si un modèle stationaire est mieux adapté aux données sous-jacentes $\{Y_x\}$ l'ajustement obtenu par la méthode décrite peut ne pas être bon. Il semble difficile d'ajuster les modèles stationnaires usuels, en effet pour une matrice de distorsion A_{in} quelconque, $\sum_i A_{in} y_i$ ne sera pas stationnaire et donc les données observées n'auront pas de variogramme stationnaire. Dans un cas important, quand A_{in} est une convolution $Z = AY = \int K(x-y)\, Y(y)$, on démontre facilement que $\sum_i A_{in} Y_i$ est stationnaire et si C_{ij}^Y est la matrice de covariance de Y, $A^T C^Y A$ est la covariance stationnaire de $\sum_i A_{in}\, Y_i$. On peut donc calculer le variogramme des

données observées et choisir un variogramme pour Y tel que le variogramme transformé $A^T\gamma A$ ajuste bien les valeurs expérimentales. C'est la démarche qui est suivie dans [8].

3.4. — EXEMPLES ET CAS D'ÉTUDE

1. — Le premier exemple est un cas où l'estimateur de moindres carrés fonctionne très bien sans qu'on utilise de régularisation. La loi qu'on veut estimer est démontrée à la figure 3.8. A la figure 3.9, le poissonisé est représenté en trait continu et la loi à estimer en pointillé. Il n'est pas difficile de voir pourquoi l'estimateur de moindres carrés fonctionne bien : il n'y a presque pas de bruit, sauf un petit peu près du mode de la loi. La loi consiste en 60 000 points observés Il s'agit de la superposition de six lois gaussiennes avec des moyennes et des variances très différentes.

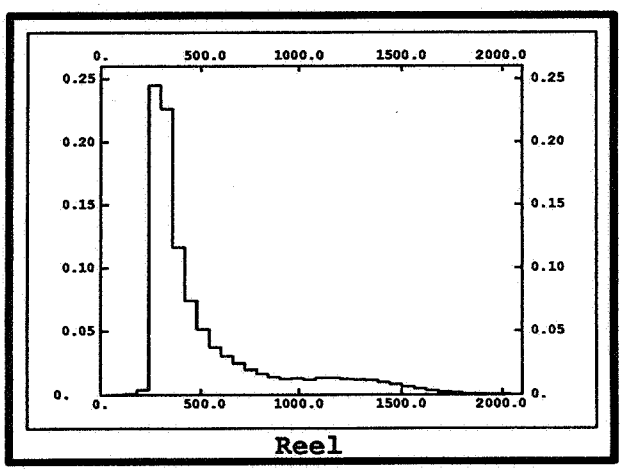


Figure 3.8 Loi sous-jacente

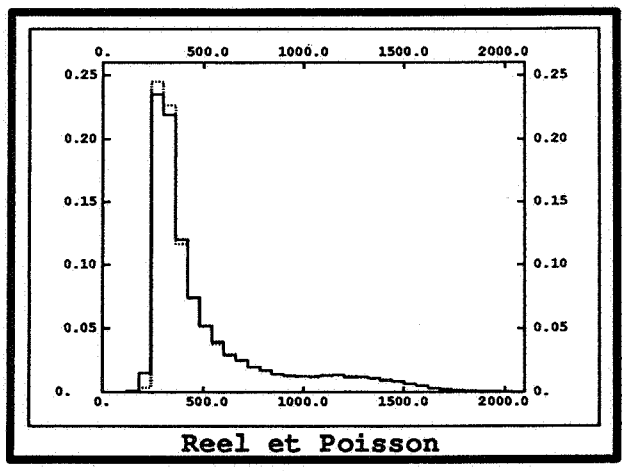


Figure 3.9 Poissonisé ... trait continu
Sous-jacente ... pointillé

Nous avons utilisé deux types d'estimateurs (tous les deux sans régularisation). Le premier est avec les pondérateurs $\omega(n) = 1$, c'est donc l'estimateur ℓ^2 classique. L'autre est avec $\omega(n) = 1/(g(n))$ qu'on appelle l'estimateur χ^2. L'estimateur χ^2 est représenté en trait continu sur la figure 3.10 et la loi à estimer est en pointillé. Il en est de même pour l'estimateur ℓ^2 à la figure 3.11. L'erreur d'estimation pour χ^2 est de 0.418×10^{-4} et, pour ℓ^2, elle

est de 2.387×10^{-4}. On voit que l'estimateur χ^2 est meilleur que ℓ^2 dans ce cas. Il semble que ceci soit souvent vrai lorsqu'on a un grand nombre de données et on a $\sup_n g(n) > 300$, mais dans la plupart des cas la différence entre χ^2 et ℓ^2 n'est pas très grande et le choix à faire entre les deux semble moins important que le choix du paramètre de régularisation ou même de la norme Ω.

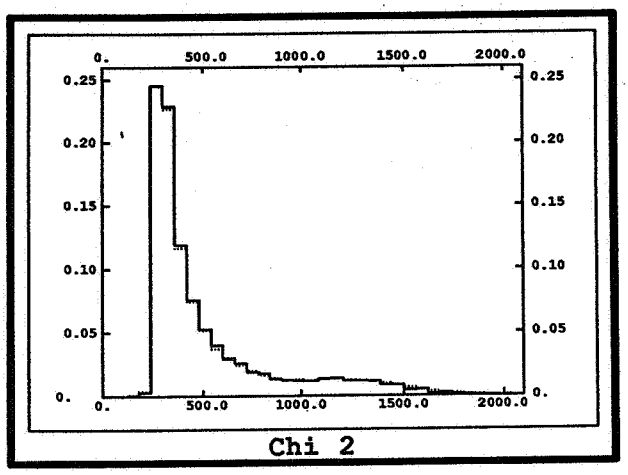


Figure 3.10 χ^2 ... trait continu
Sous-jacente ... pointillé

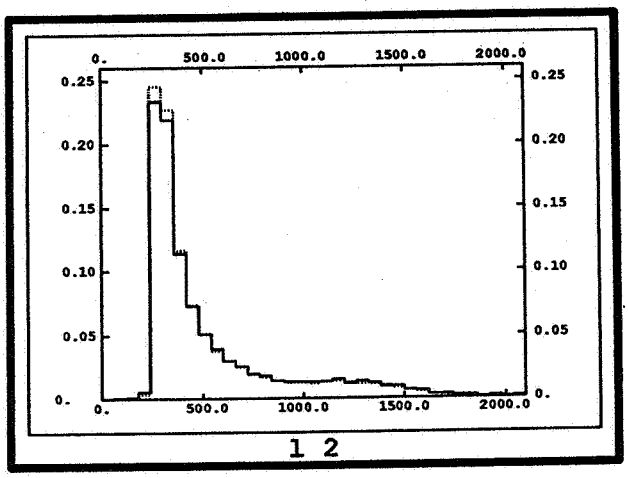


Figure 3.11 l^2 ... trait continu
Sous-jacente ... pointillé

2. — On reprend ici l'exemple ébauché à la figure 3.1, et on présente les lois réelles et poissonisées à nouveau à la figure 3.12.

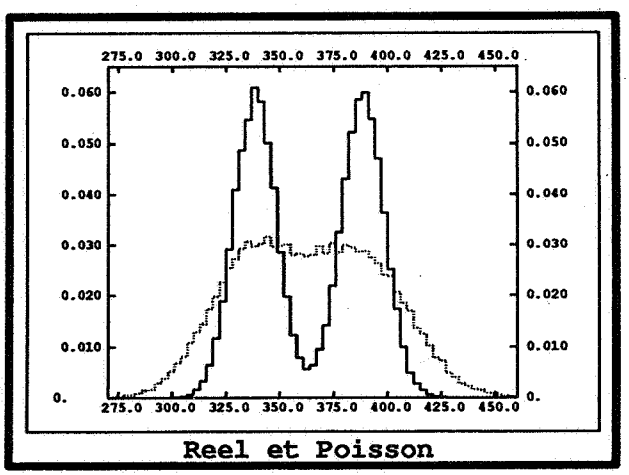


Figure 3.12 Réel ... trait continu
Poisson ... pointillé

Pour ceci, nous avons deux acquisitions de 60 000 points chacune. Si on regroupe les acquisitions, on a un histogramme de 120 000 points et on peut essayer d'estimer le paramètre de régularisation par validation croisée. On prend comme norme de régularisation une cubique (inverse d'une matrice de covariance du modèle cubique) d'une portée de 70. La fonction CV(λ) est montrée à la figure 3.13. Comme on peut le constater, elle est très irrégulière à cause des instabilités dans la matrice chapeau, et on ne peut pas la minimiser à l'aide de la méthode du gradient. On remarque quand même qu'il y a un minimum vers 0.001. Le résultat de la régularisation est donné à la figure 3.14 en trait continu. La variance d'erreur d'estimation est de 1.195×10^{-3}.

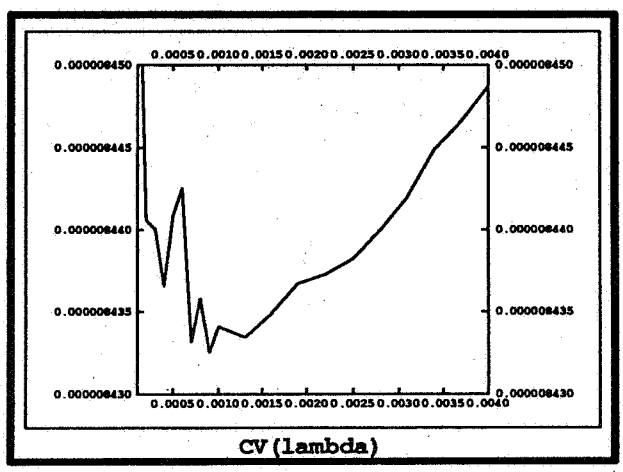


Figure 3.13 La fonction $CV(\lambda)$

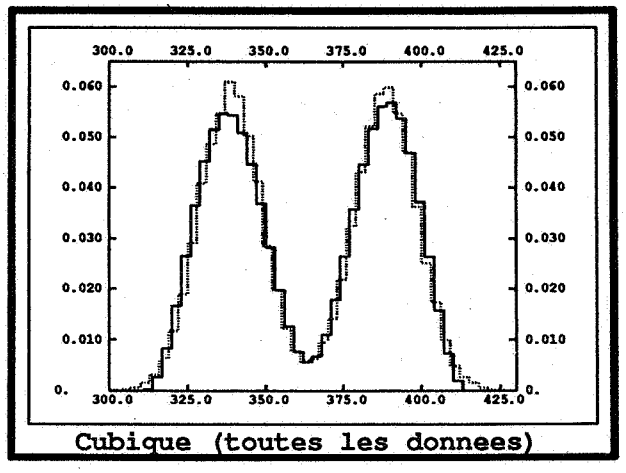


Figure 3.14 Estimateur (trait continu) obtenu avec $\lambda = \min_{\mu} CV(\mu)$

Si l'on utilise les deux acquisitions, il est possible de calculer les fonctions $DA_1(\lambda)$ et $DA_2(\lambda)$, dont une est montrée sur la figure 3.15. Il est très facile de minimiser cette fonction avec une procédure numérique. En choisissant l'acquisition pour laquelle $DA(\lambda)$ est la plus grande, la régularisation donne la figure 3.16 et la variance d'erreur est de 0.9747×10^{-3}. Dans ce cas on a un meilleur estimateur même en n'ayant utilisé que la moitié des données!

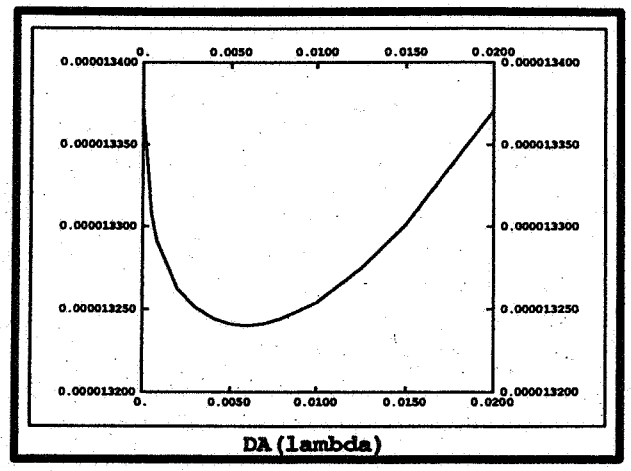


Figure 3.15 La fonction $DA_1(\lambda)$

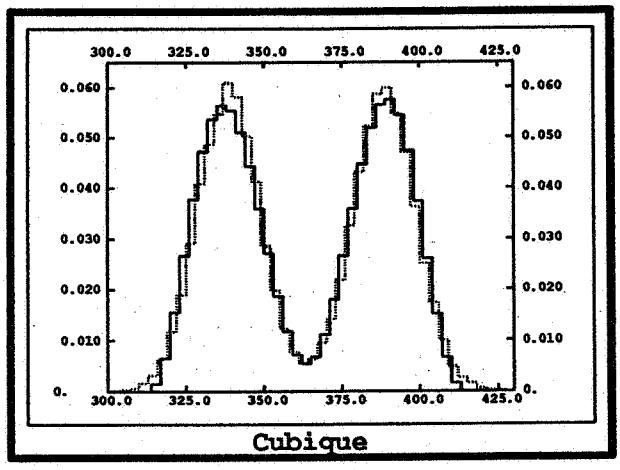


Figure 3.16 Estimateur (trait continu) avec $\lambda = \frac{1}{2}(\min_{\mu} DA_1(\mu) + DA_2(\mu))$ et un modèle cubique

Finalement, pour mettre en évidence l'influence du choix de la norme de régularisation, on a fait un deuxième essai avec un effet de pépite comme norme. Evidemment ceci est moins bien adapté à l'histogramme sous-jacent qui a une allure assez lisse..Le résultat de la régularisation avec ce choix est présenté à la figure 3.17 et l'erreur d'estimation est de 1.695×10^{-3}. L'estimation est moins bonne que ce qu'on avait obtenu avec le modèle cubique.

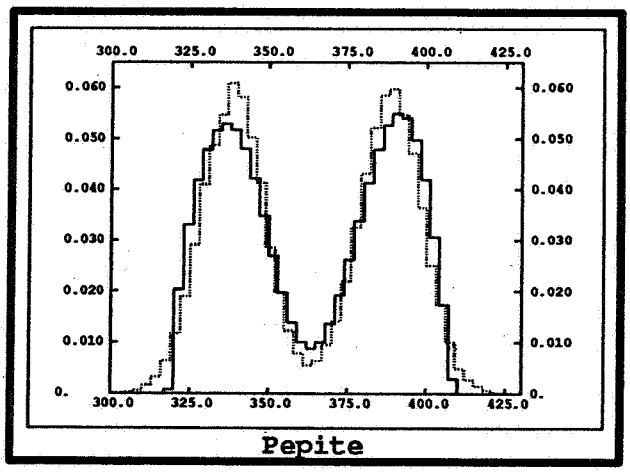


Figure 3.17 Estimateur avec un effet de pépite

3. — Si l'histogramme sous-jacent admet un mode quasi-atomique et des queues longues, et si on a un bruit poissonien important, il peut arriver que la méthode de régularisation ne soit pas très bien adaptée, en particulier pour un choix de norme quadratique. Ceci s'interprète facilement à l'aide de la formulation en terme de fonction aléatoire. Si le mode est très élevé, ni l'histogramme lui-même, ni les incréments d'ordre 1 ne ressemblent à une réalisation d'une fonction aléatoire stationnaire. Parce que le bruit est élevé, le filtrage est obligé de lisser de façon importante et en particulier il risque de lisser le mode. Un exemple de loi avec une queue longue est donné à la figure 3.18 avec sa poissonisée en pointillé. Ici il y a 30 000 points. On choisit un modèle cubique pour la régularisation et le résultat est donné à la figure 3.19. La variance d'erreur est de 0.845×10^{-2}. Nous ne savons pas actuellement comment traiter l'estimation de ce type d'histogrammes que l'on considère comme un mélange de lois continues et de lois discrètes.

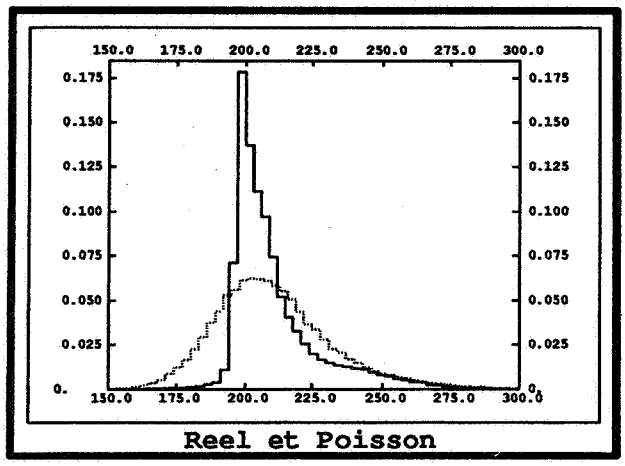


Figure 3.18 Réel ... trait continu
Bruité ... pointillé

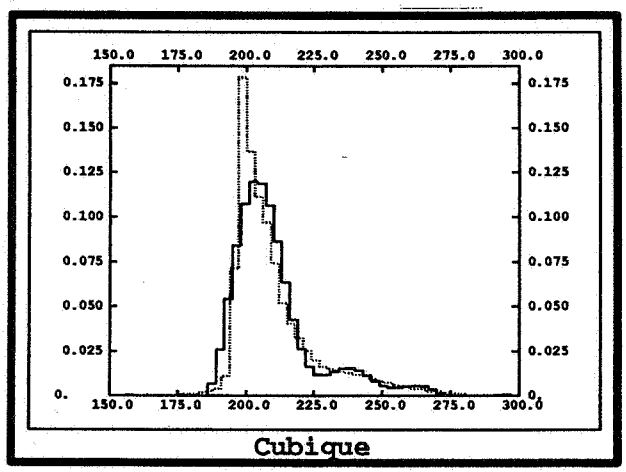


Figure 3.19 Estimateur ... trait continu
Réel ... pointillé

4. — Il est intéressant de comparer la méthode directe d'estimation de l'histogramme sous-jacent avec l'estimateur obtenu en filtrant une image par krigeage linéaire. La figure 3.20 est une image simulée et la figure 3.21 est la même image observée avec un bruit poissonien. Le rapport signal/bruit est de 0.128. Un filtrage linéaire avec un voisinage de 15×15 a été effectué et le résultat est montré à la figure 3.22. Le filtrage n'est pas particulièrement efficace du fait que le bruit est très élevé et que la portée du variogramme n'est que de 8 pixels. Le rapport signal/bruit après filtrage est de 2.245.

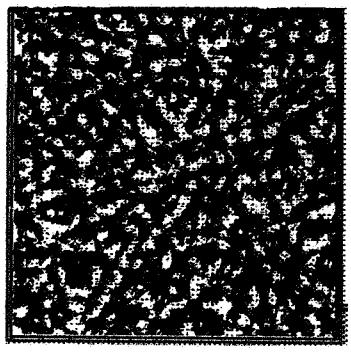


Figure 3.20 Image sous-jacente

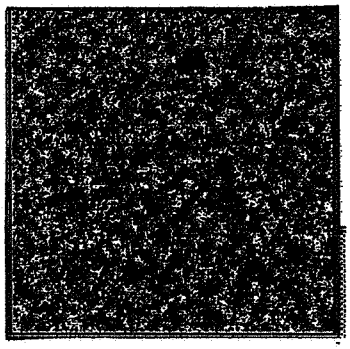


Figure 3.21 Image poissonisée

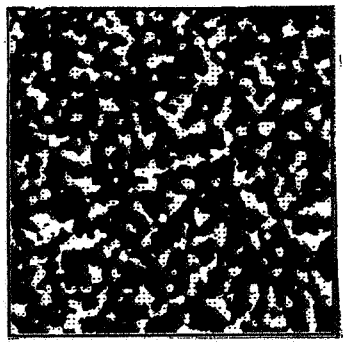


Figure 3.22 Image filtrée

Les histogrammes réel et poissonisé sont montrés à la figure 3.23 (le réel en trait continu). Sur la figure 3.24, on trouve l'estimateur par krigeage, c'est-à-dire l'histogramme de l'image krigée. Il est représenté par un trait continu et l'histogramme réel est en pointillés. On voit que l'histogramme obtenu par krigeage a une variance qui est trop petite. Les queues de l'histogramme sont plus courtes et le mode est trop élevé. Ceci est caractéristique d'un krigeage fait sur un champ plus ou moins homogène quand le voisinage de krigeage est grand. (Dans un cas non homogène, les queues seront toujours trop courtes, mais il est possible que localement le filtrage ne soit pas suffisamment fort et le mode sera trop petit, comme nous en verrons un exemple plus loin).

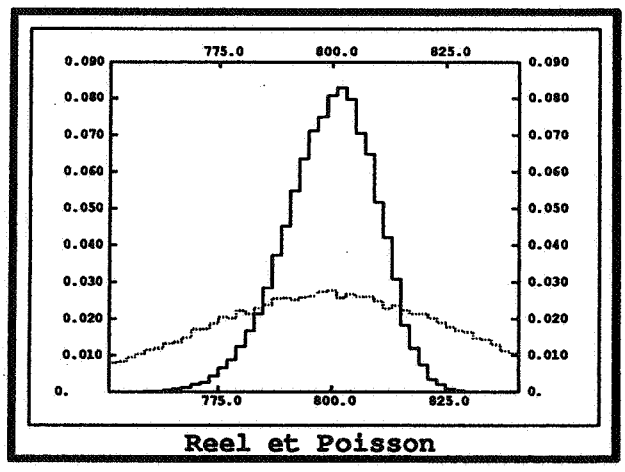


Figure 3.23 Histogramme de l'image sous-jacente (trait continu) et sa poissonisée.

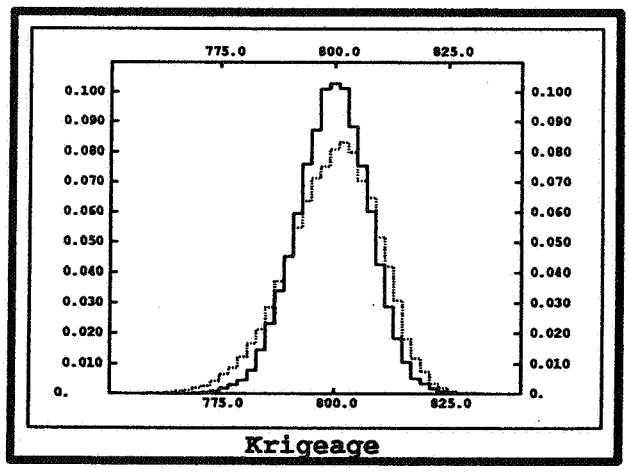


Figure 3.24 Histogramme de l'image filtrée (trait continu) et l'histogramme sous-jacente

En pratique, on veut savoir si l'histogramme obtenu par krigeage de l'image correspond bien à la réalité, et malheureusement nous n'avons pas l'histogramme réel. Un test partiel consiste à poissoniser l'image filtrée (ou directement, son histogramme) et de comparer le résultat avec la poissonisée réelle. Puisque le problème est mal posé, on ne s'attend pas avoir un écart important entre les deux poissonisées, mais on peut rechercher des erreurs systématiques qui nous indiquent de quelle manière l'histogramme krigé est erroné. A la figure 3.25, on voit en trait continu la poissonisée de l'image filtrée. Sur les trois valeurs centrales, elle est plus élevée que la poissonisée réelle, ce qui nous indique que le mode de l'histogramme de l'image krigée est effectivement plus grand que l'histogramme réel. Il est difficile de quantifier cette méthode. Tout ce qu'on peut dire est que l'écart entre les lois sous-jacentes doit être plus grand que l'écart entre les poissonisées.

L'histogramme de cette image étant bien symétrique, il est très facile à estimer par la méthode de régularisation. Le paramètre a été ajusté par la méthode de double acquisition, $\lambda = 1.25$ pour un modèle "spline" et le résultat est montré à la figure 3.26.

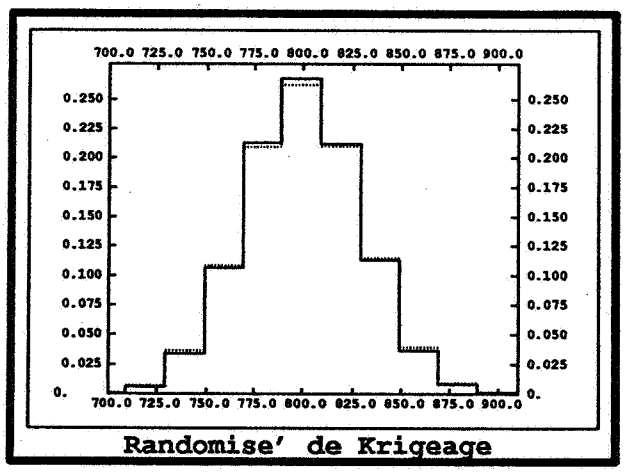


Figure 3.25 Poissonisée de l'image filtrée (trait continu) et poissonisée réelle.

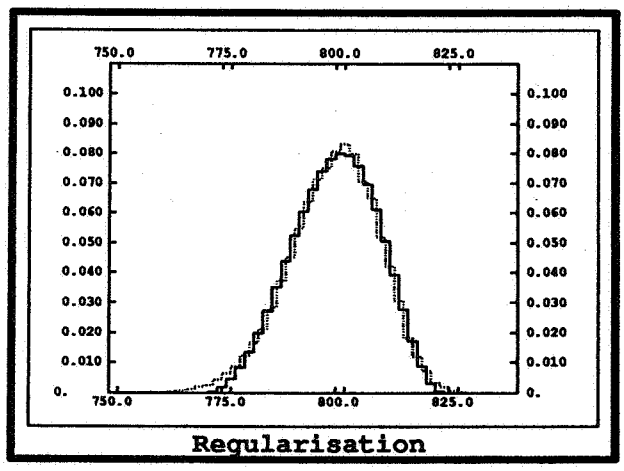


Figure 3.26 Estimateur par régularisation (trait continu)

5. — Sur les figures 3.27 et 3.28 on voit une deuxième simulation et sa poissonisée. Dans ce cas, il y a nettement moins de bruit (rapport signal/bruit égal à 0.9288), et de plus la portée du variogramme de l'image étant de 20 pixels, il

est beaucoup plus grande que dans le dernier exemple. Le filtrage, qui est montré à la figure 3.29, est beaucoup plus efficace (rapport signal/bruit égal à 9.137).

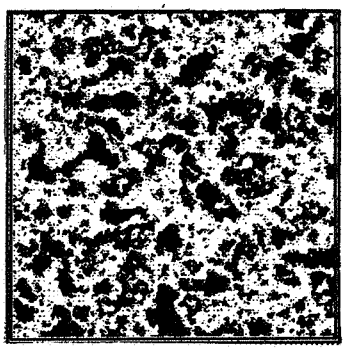


Figure 3.27 Image sous-jacente

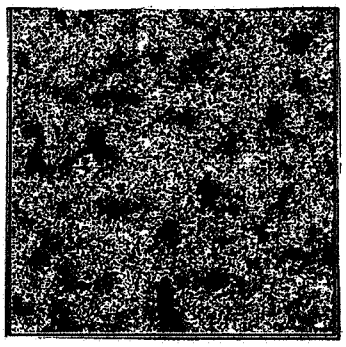


Figure 3.28 Image poissonisée

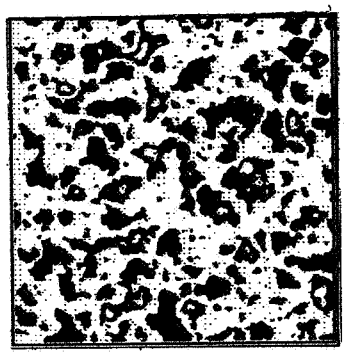


Figure 3.29 Image filtrée

Les histogrammes réel (trait continu) et poissonisé sont montrés à la figure 3.30, et l'histogramme obtenu par krigeage de l'image est sur la figure 3.31. On note que les queues de l'histogramme obtenu par krigeage sont beaucoup trop longues. Ceci est partiellement dû au fait qu'on a utilisé un voisinage de 15 × 15, ce qui est petit quand la portée est de 20 pixels. On voit que l'estimation est meilleure que dans le dernier exemple. Ceci est naturel parce qu'il y avait moins de bruit au départ.

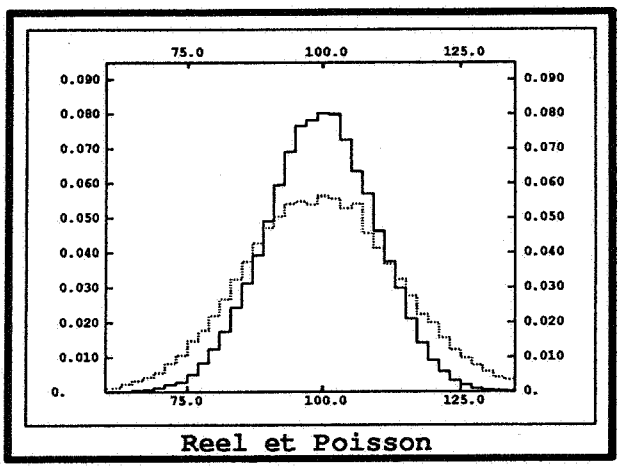


Figure 3.30 Histogramme de l'image sous-jacente (trait continu) et de sa poissonisée.

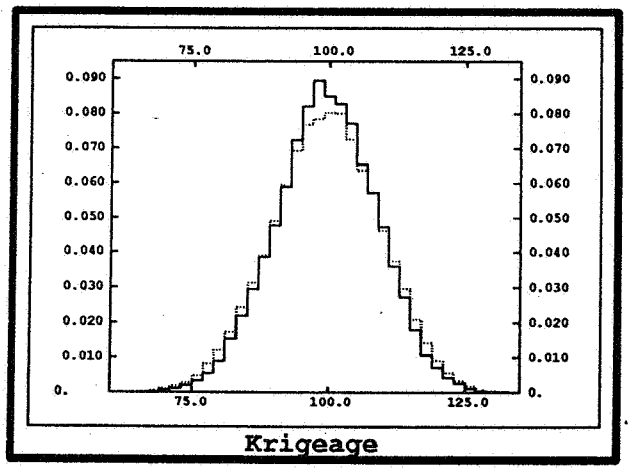


Figure 3.31 Histogramme de l'image filtrée (trait continu) et histogramme de l'image sous-jacente.

Le résultat d'une régularisation appliquée à l'histogramme brut est montré à la figure 3.32. Il est plus proche de l'histogramme réel que l'estimateur obtenu par krigeage de l'image.

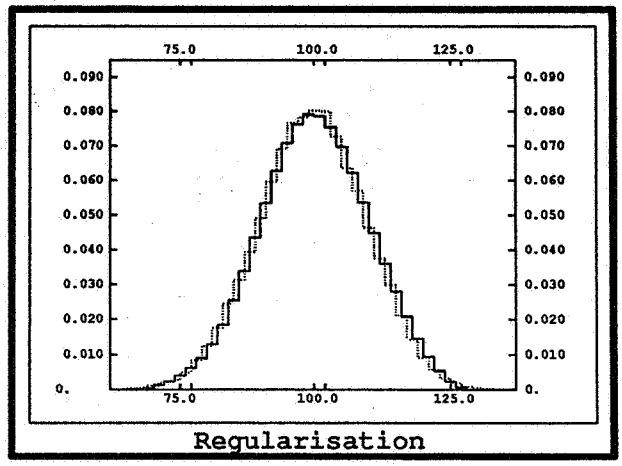


Figure 3.32 Estimateur par régularisation (trait continu)

6. — On continue avec l'image de l'exemple précédent, mais ici on applique une anamorphose à l'image initiale pour obtenir une queue plus longue. L'image initiale est montrée à la figure 3.33, sa poissonisée et son filtrage sont sur les figures 3.34 et 3.35 respectivement. Les rapports signal/bruit intial et final sont de 3.414 et 15.41.

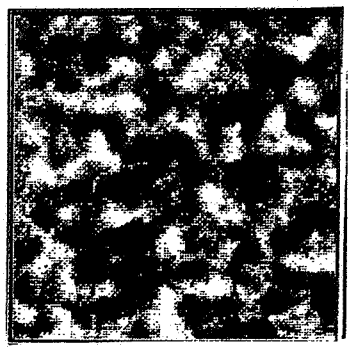


Figure 3.33 Image sous-jacente

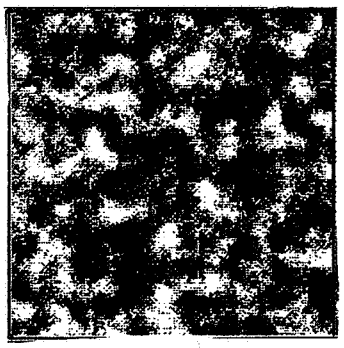


Figure 3.34 Image poissonisée

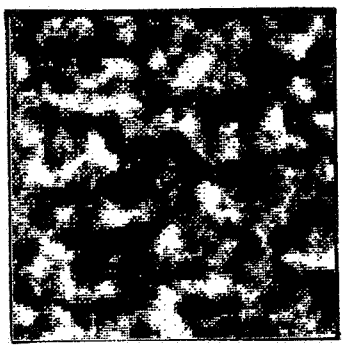


Figure 3.35 Image filtrée

Sur les figures 3.36-3.38 sont représentés les histogrammes réels et poissonisés, krigeage et régularisation respectivement. Ici le krigeage donne une

estimation remarquablement efficace de la loi sous-jacente. On note quand même qu'il surestime les valeurs au début de la queue. Cette tendance sera plus marquée si la queue de l'histogramme est plus longue. Néanmoins, comme conclusion pratique, on peut dire que si le rapport signal/bruit final dépasse 12, le krigeage de l'image donne un assez bon estimateur de l'histogramme sous-jacent.

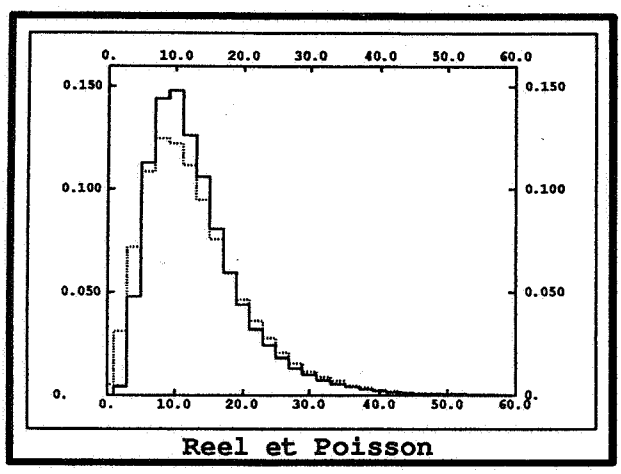


Figure 3.36 Histogramme de l'image sous-jacente (trait continu) et de sa poissonisée.

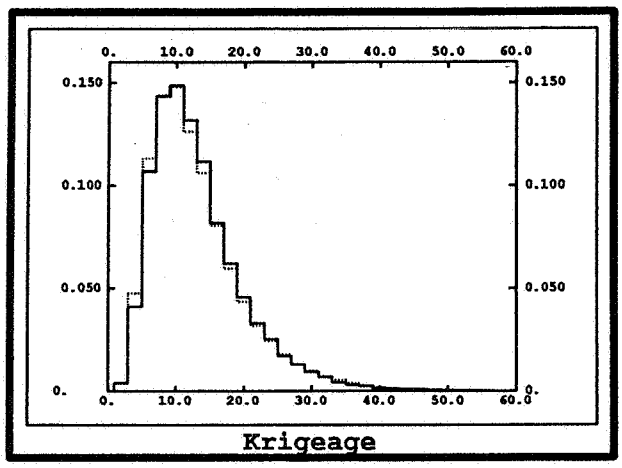


Figure 3.37 Histogramme de l'image filtrée (trait continu) et histogramme de l'image sous-jacente.

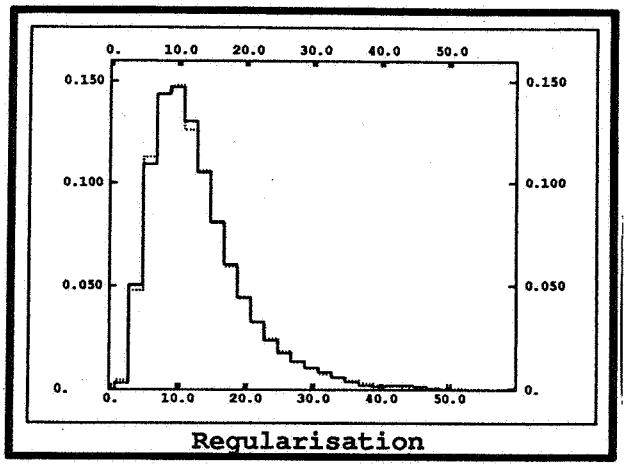


Figure 3.38 Estimateur par régularisation (trait continu)

En ce qui concerne la régularisation, on voit que, bien que la queue soit assez longue, la régularisation a donné un très bon résultat (même légèrement meilleur que le krigeage). Il faut donc que le mode de l'histogramme soit vraiment très élevé

pour que l'estimation par régularisation pose problème. Malheureusement cela peut très bien se produire en pratique comme nous allons le voir dans les exemples suivants.

7. — Nous étudions deux éléments *Mn* et *Nb* pris sur le même échantillon. Il s'agit d'une étude de macroségrégation (taille réelle de l'image : 6.4 mm × 6.4 mm) dans les *V* ségrégés. Pour le premier élément (fig. 3.39), nous avons un rapport signal/bruit de 7.769 et pour l'image filtrée (fig. 3.40), ce rapport est de 14.637.

Figure 3.39 Image brute 25μm par pixel.

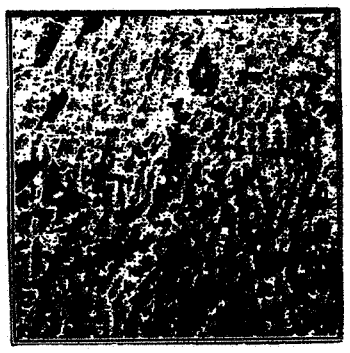


Figure 3.40 Image filtrée

Les histogrammes des estimateurs (en trait continu) et de l'image bruitée sont montrés aux figures 3.41 et 3.42. Avec un rapport signal/bruit aussi élevé, on pourrait penser que le krigeage donnera un très bon estimateur de l'histogramme sous-jacent. Mais la queue de l'histogramme est extrêmement longue (beaucoup plus longue qu'on ne peut le voir sur les figures), et il faut rappeler que dans un cas pareil, le rapport signal/bruit peut être artificiellement élevé. Il est donc possible que le krigeage n'ait pas lissé suffisamment pour obtenir un bon estimateur de l'histogramme. Regardons maintenant l'autre élément pour mieux comprendre.

Sur la figure 3.43, on voit l'image observée de l'élément *Nb* et on voit son filtrage sur la figure 3.44. Les rapports signal/bruit initial et final sont de 2.355 et 3.997. L'histogramme obtenu est montré à la figure 3.45, et celui obtenu par régularisation sur la figure 3.46.

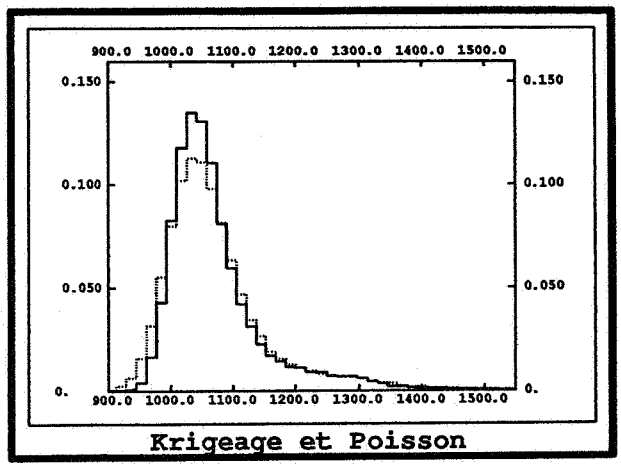


Figure 3.41 Histogramme de l'image filtrée (trait continu) et histogramme de l'image brute

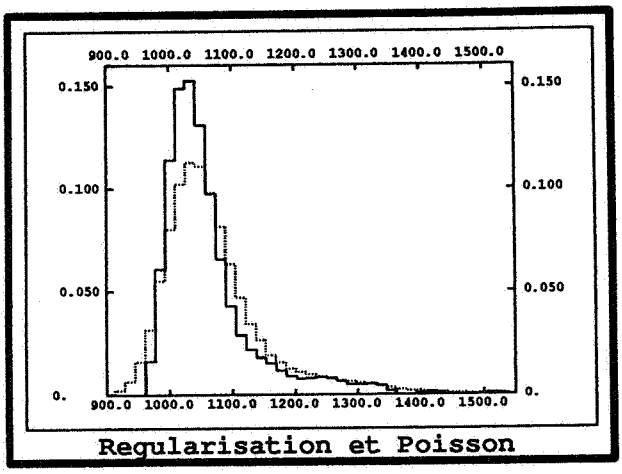


Figure 3.42 Estimateur par régularisation (trait continu) et histogramme de l'image brute.

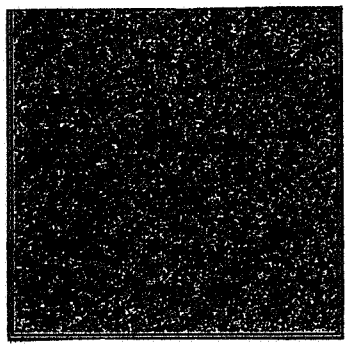


Figure 3.43 Image brute 25µm par pixel.

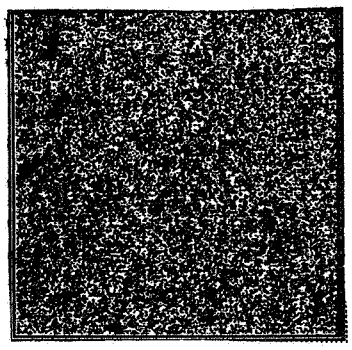


Figure 3.44 Image filtrée

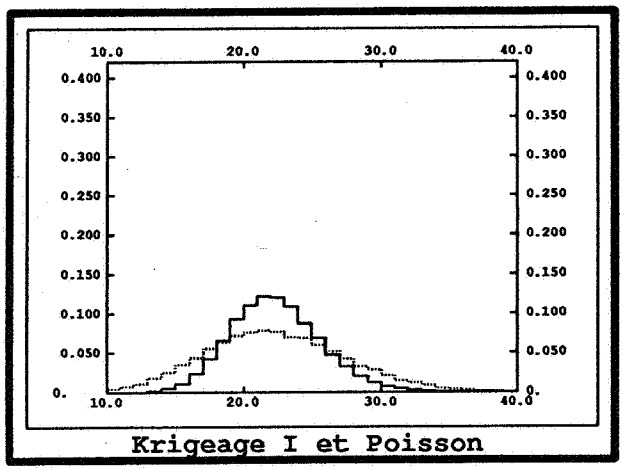


Figure 3.45 Histogramme de l'image filtrée de fig. 3.44

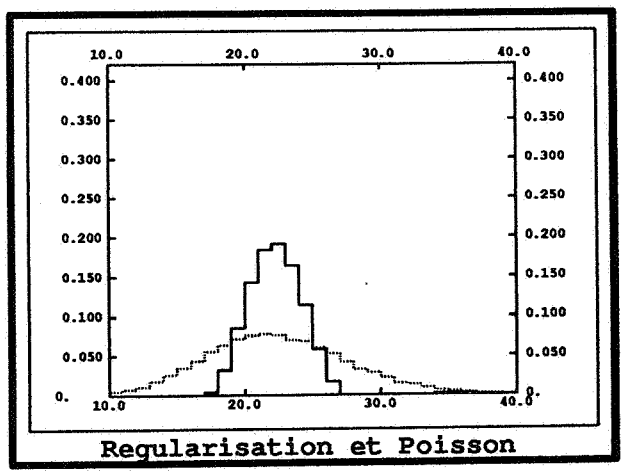


Figure 3.46 Estimateur par régularisation

Appliquons maintenant un masque sur les valeurs les plus elevées de l'image observée (valeurs de très forte concentration dans la zone liquide). En masquant 0.5% des données et en filtrant, on obtient maintenant la figure 3.47. L'histogramme de l'image filtrée est montré à la figure 3.48. Les rapports signal/bruit sur l'image masquée sont passés à 0.097 et 1.89 pour l'initial et le filtré respectivement. Entre le premier et le deuxième krigeage, avec le premier qui ne lisse pas suffisamment et le second qui lisse trop, il semble que l'histogramme proposé par la régularisation soit un bon compromis. Finalement notons que la régularisation donne deux petites "bosses" avec un poids total de 0.7% en dehors du centre de l'histogramme (dont on voit le premier vers 35 à 38). La régularisation suggère donc qu'il a 0.7% de valeurs très élevées, ce qui correspond bien au nombre de valeurs très élevées sur l'histogramme poissonisé ($\simeq$ .65%) .

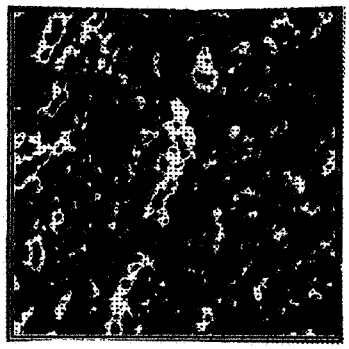


Figure 3.47 Image filtrée

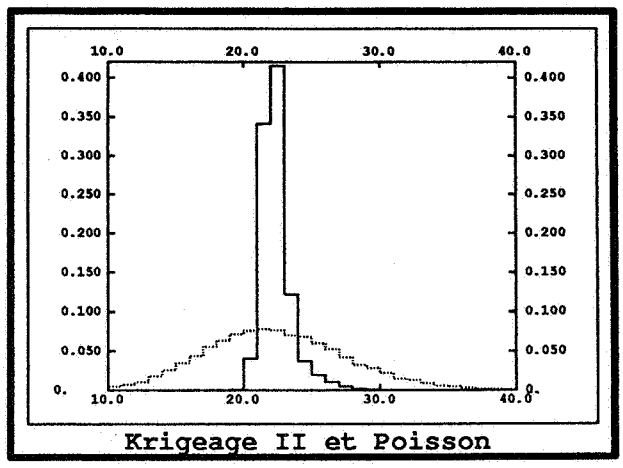


Figure 3.48 Histogramme de l'image filtrée de fig. 3.47

On peut utiliser l'estimateur de l'histogramme pour obtenir des mesures de taux de ségrégation. La mesure la plus simple est définie par

$$\mathrm{TS}_1 = \frac{\sigma_Y}{m_Y}$$

où σ_Y^2 et m_Y^2 sont la variance et la moyenne sous-jacentes respectivement. Pour obtenir cette mesure il n'est pas nécessaire d'avoir un estimateur de la loi sous-jacente. En effet, $\sigma_Z^2 = \sigma_Y^2 + m_Y$ et $m_Z = m_Y$ donc

$$TS_1 = \frac{\sqrt{\sigma_Z^2 - m_Z}}{m_Z}$$

et on peut estimer le taux de ségrégation à partir des données initiales. Cette mesure dépend uniquement de la moyenne et la variance et donne donc les mêmes résultats pour toute loi ayant les deux premiers moments identiques. De plus cette mesure est instable s'il existe un petit nombre de fortes valeurs qui perturbe l'estimation de la variance. Par exemple pour l'image de *Nb* l'estimateur de TS est de 0.373 quand on n'applique pas de masque mais avec le masque il est de 0.094.

La connaissance de l'histogramme sous–jacent nous permet de définir d'autres mesures du taux de ségrégation qui doivent être plus robustes. Par exemple si q_x est le quantile à $x\%$, c'est à dire $P[Y \leq q_x] = x$, alors un choix possible pour la définition de taux de ségrégation est

$$TS_2 = \frac{q_{.95} - q_{.05}}{m_Y}$$

Si l'on applique cette définition, on trouve la même valeur du taux de ségregation pour *Nb* avec ou sans masque, 0.275. On ne peut pas comparer différentes mesures de taux de ségrégation, mais elles sont définies de telle sorte qu'il est possible de comparer des éléments différents et même des études différentes (puisque nous avons divisé par la moyenne, nous avons obtenu une mesure sans dimension). Pour le premier élément, *Mn*, nous obtenons une mesure du taux de ségrégation avec les quantiles de 0.253, et donc pour cette mesure, les deux éléments semblent avoir une ségrégation du même ordre de grandeur, avec peut–être *Nb* un peu plus ségrégé. Pour la première méthode de mesure de la ségréation, *Mn* a un taux de 0.078. Le problème ici est qu'on ne sait pas si l'on doit le comparer avec le taux mesuré avec ou sans masque.

Le deuxième mesure du taux de ségrégation donne une valeur plus stable et par conséquent plus facile à comparer avec le taux obtenu sur d'autres éléments et d'autres études.

8. — Nous allons maintenant considérer un autre exemple réel. Cette fois, les images proviennent d'une microségrégation (1.2 mn/m de réduction douce, sans brassage; zone basaltique, taille 256 × 256, 25 μm par pixel), et nous avons deux éléments *Si* et *P*. Sur la figure 3.49, on voit le premier élément et son filtré est montré sur la figure 3.50. Les rapports signal/bruit sont de 2.691 et 6.376 respectivement. Les histogrammes estimés sont montrés sur les figures 3.51 et

3.52. On voit encore ici que la régularisation a un mode plus élevé que celui du krigeage et que le krigeage a une queue beaucoup plus longue que la régularisation. De plus la régularisation a deux modes qui ne sont vraisemblablement que des artefacts dus à la non-stationnarité des incréments sur l'histogramme.

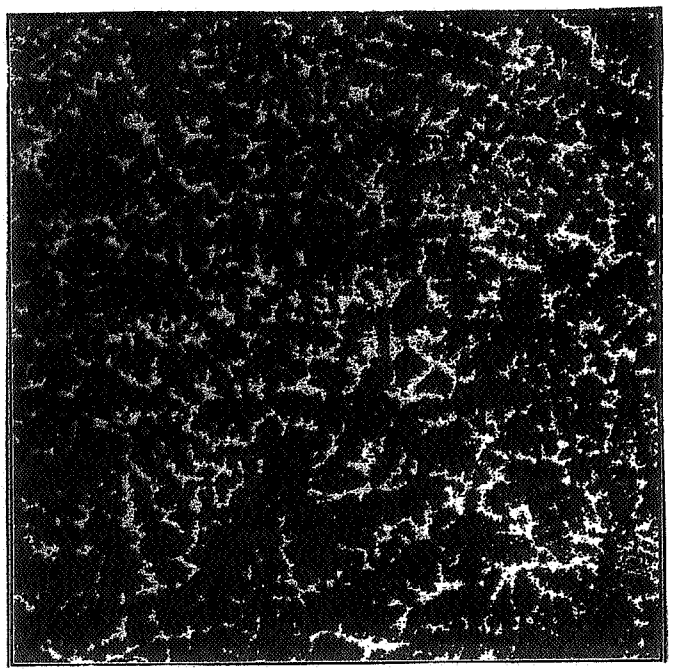


Figure 3.49 Image brute

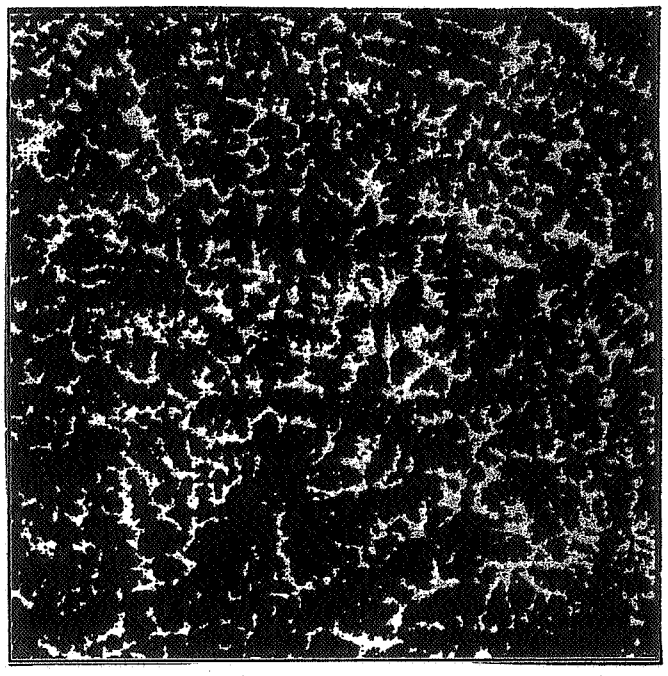


Figure 3.50 Image filtrée

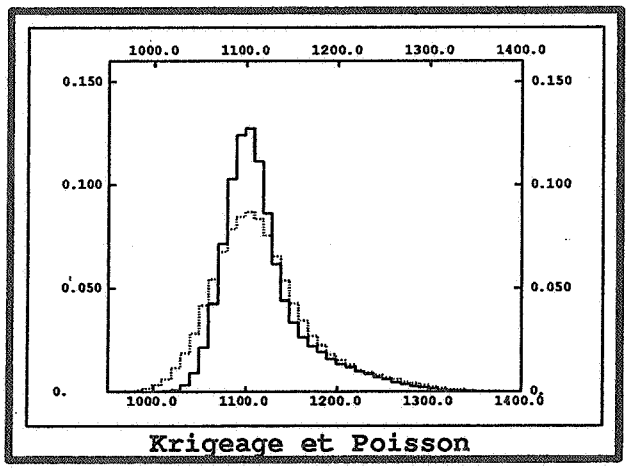


Figure 3.51 Histogramme de l'image filtrée (trait continu) et histogramme de l'image brute.

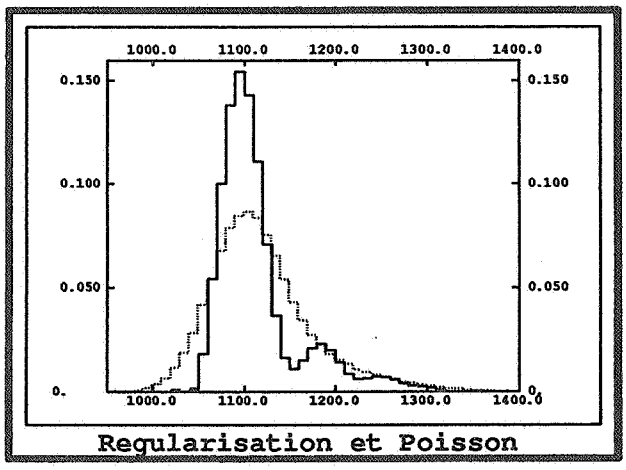


Figure 3.52 Estimateur par régularisation (trait continu) et histogramme de l'image brute.

Pour voir si le krigeage a bien restitué l'histogramme, on peut repoissoniser l'image filtrée. Le résultat est montré à la figure 3.53 (repoissonisée en trait continu). On voit bien que l'estimateur obtenu par krigeage de l'image a sous-estimé le mode et a exagéré la queue entre 1 150 et 1 250. L'histogramme vrai a vraisemblablement un mode de l'ordre de grandeur de celui prévu par régularisation, et avec une queue qui est plus petite que le krigeage, mais plus lisse que la régularisation. On peut essayer de combiner ces deux estimateurs en utilisant l'histogramme obtenu par krigeage de l'image comme *a priori* pour la régularisation (voir § 3.3). L'estimateur ainsi obtenu est présenté sur la figure 3.54. Il ressemble beaucoup au premier estimateur de régularisation mais les deux "bosses" sur la queue ont diminué. Cette procédure qui consiste à utiliser l'histogramme obtenu par krigeage de l'image comme loi *a priori* me semble assez douteuse parce qu'elle donne un biais vers l'histogramme de l'image krigée qui peut être assez mauvais comme estimateur.

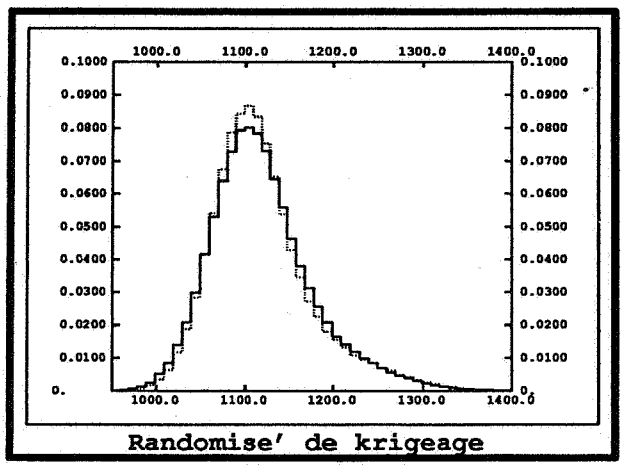


Figure 3.53 Poissonisée de l'image filtrée (trait continu)

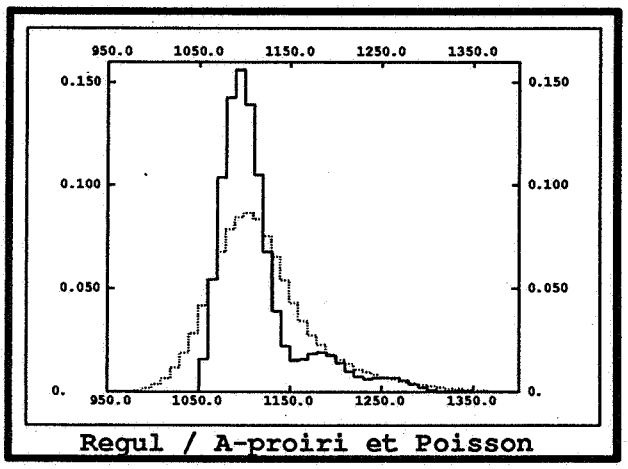


Figure 3.54 Estimateur par régularisation avec krigeage comme loi a-priori

Pour le deuxième élément, on voit les images observées et filtrées sur les figures 3.55 et 3.56. Les rapports signal/bruit sont de 0.171 et 5.636. Les histogrammes sont montrés sur les figures 3.57 et 3.58 et la repoissonisée de l'image filtrée sur la figure 3.59. Il est difficile d'analyser la repoissonisée, mais on peut peut-être conclure que la queue de l'image krigée est trop longue entre 425 et 450 et trop courte après. Il n'est pas possible de conclure que le mode de l'histogramme de l'image bruitée est trop élevé à partir de l'histogramme repoissonisé.

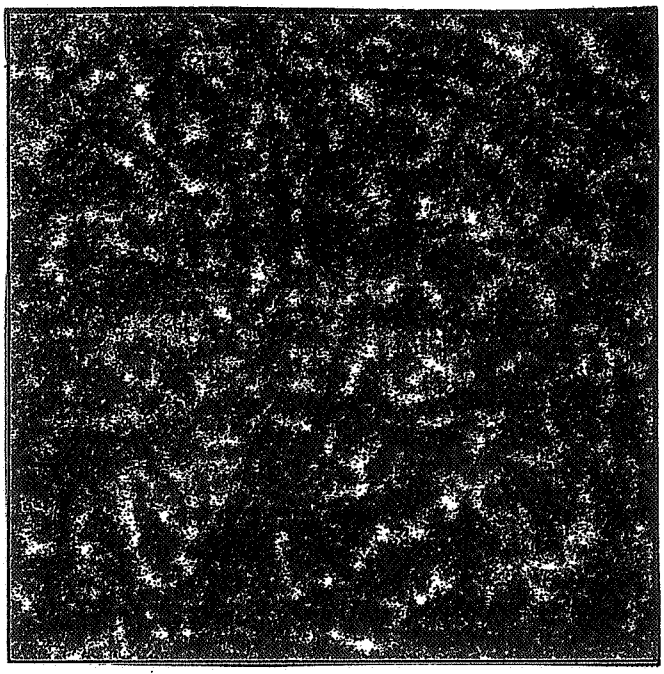


Figure 3.55 Image brute

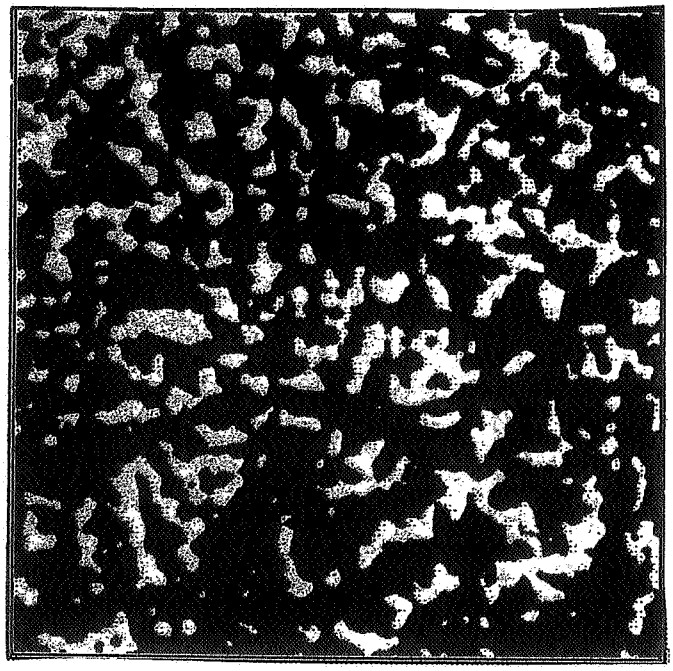


Figure 3.56 Image filtrée

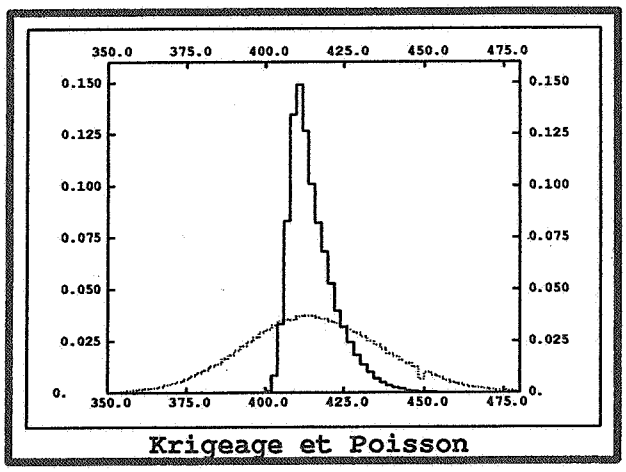


Figure 3.57 Histogramme de l'image filtrée (trait continu) et histogramme de l'image brute.

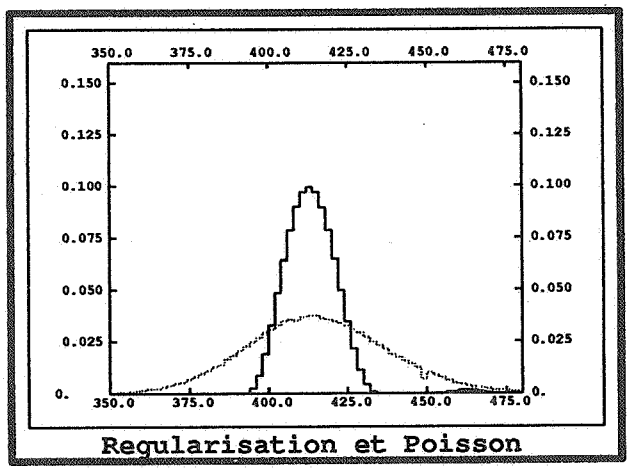


Figure 3.58 Estimateur par régularisation (trait continu) et histogramme de l'image brute.

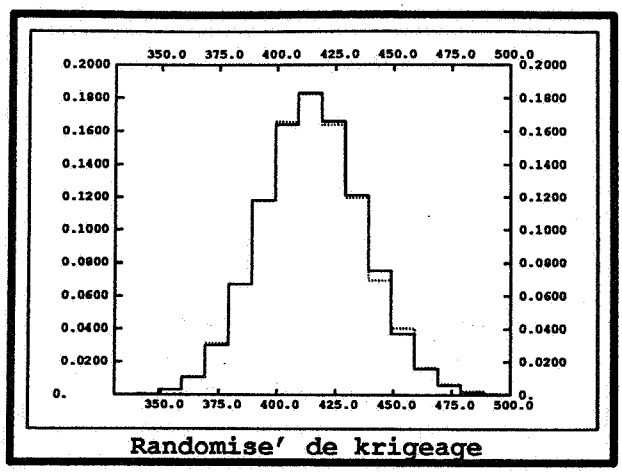


Figure 3.59 Poissonisé de l'image filtrée
(trait continu)

9. — Il est intéressant de savoir comment on doit répartir le temps d'acquisition par pixel et le nombre de pixels sur lesquels on mesure le signal de façon à avoir un temps d'acquisition total constant, ou ce qui revient au même, tel que le côut en temps d'utilisation de la microsonde soit fixe

$$\text{temps total} = \text{temps par pixel} \times \text{nombre de pixels}$$

Ceci est un problème difficile qui n'a pas encore été traité de manière générale. On considère ici un cas d'étude sur une simulation. On voit que, pour cet exemple, la régularisation peut donner un assez bon estimateur de la loi sous-jacente même lorsque le temps d'acquisition est faible. Une conclusion provisoire est qu'il vaut mieux mesurer un grand nombre de points, même avec un bruit élevé que d'avoir très peu de données mais avec un bruit faible. De plus, l'avantage d'avoir beaucoup de données est qu'on peut les représenter sous forme d'image et qu'on a également beaucoup plus de possibilités d'échantillonner de petites régions à forte concentration.

Sur la figure 3.60, on voit la loi réelle qu'on veut estimer. On échantillonne maintenant 5 000 points avec cette loi multipliée par un facteur de 10 et on poissonise chaque point. On voit l'histogramme poissonisé en trait continu sur la figure 3.61. La loi sans bruit des 5 000 points est en pointillés. L'estimateur par régularisation se trouve à la figure 3.62.

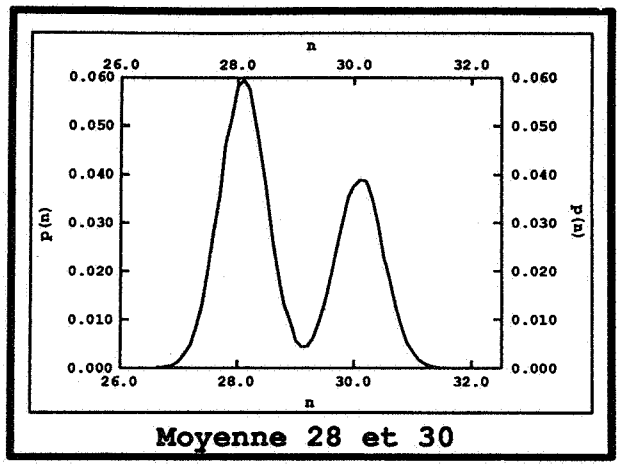


Figure 3.60 Loi réelle

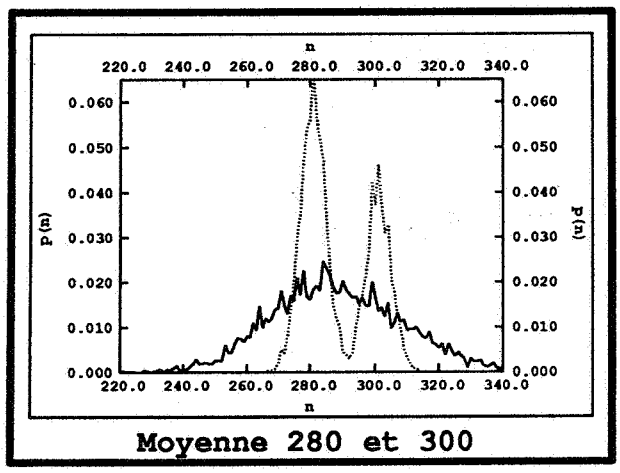


Figure 3.61 Sous-jacente à 5000 points (trait continu) et sa poissonisée.

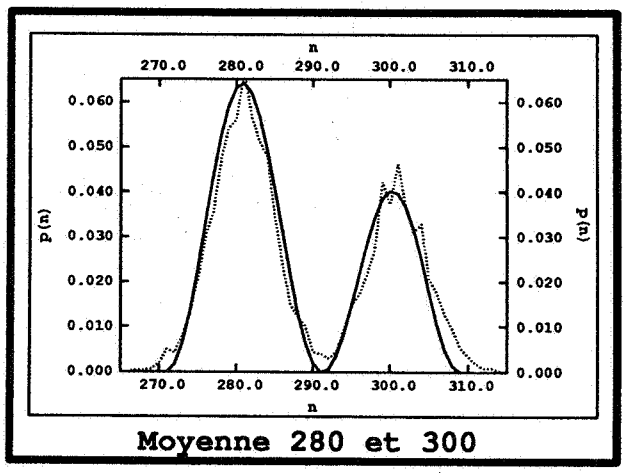


Figure 3.62 Estimateur par régularisation

On peut ensuite considérer un échantillon avec un temps d'acquisition dix fois plus grand (autrement dit on multiplie la loi sous-jacente par un facteur de dix). Pour garder un temps d'acquisition total constant, il faut donc échantillonner dix fois moins de données. Sur la figure 3.63, on voit en trait continu l'échantillonnage de 500 points avec un bruit poissonien, et en pointillés, les valeurs sans bruit.

On note que la seule loi qu'on puisse estimer, celle des 500 points sans bruit, n'est pas elle-même un très bon estimateur de la loi sous-jacente de la figure 3.60. La régularisation est donnée à la figure 3.64, et est moins satisfaisante que l'estimation avec 500 points même s'il y a moins de bruit.

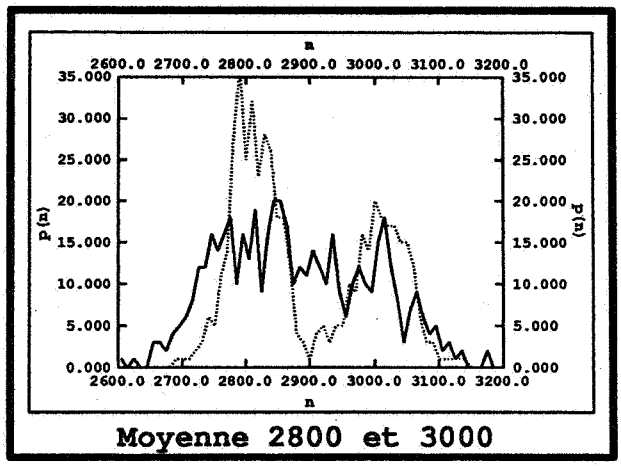


Figure 3.63 Sous-jacente à 500 points (trait continu) et sa poissonisée.

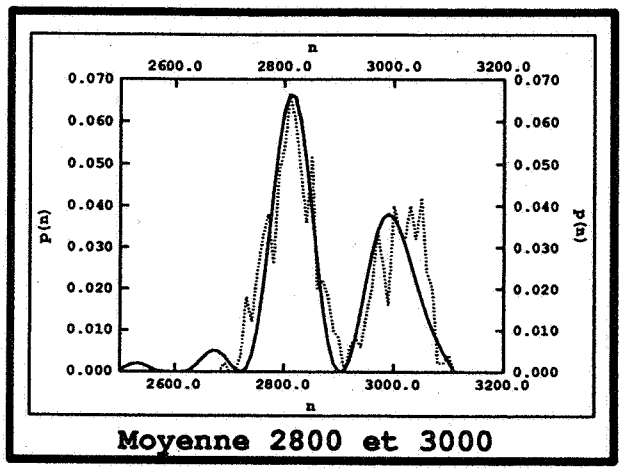


Figure 3.64 Estimateur par régularisation

Finalement, on n'échantillonne que 200 points mais avec deux fois et demi moins de bruit. La poissonisée de l'échantillonnage est représentée en ligne continue à la figure 3.65. Ici on voit qu'on obtient finalement quelque chose qui semble être un bon estimateur pour l'échantillon non bruité. Mais en fait l'échantillon non bruité est un très mauvais estimateur pour la loi sous-jacente de la figure 3.60. La régularisation est montrée à la figure 3.66.

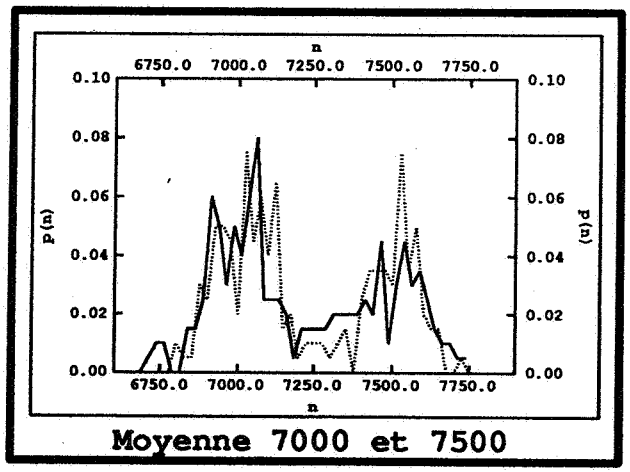


Figure 3.65 Sous-jacente à 200 points (trait continu) et sa poissonisée.

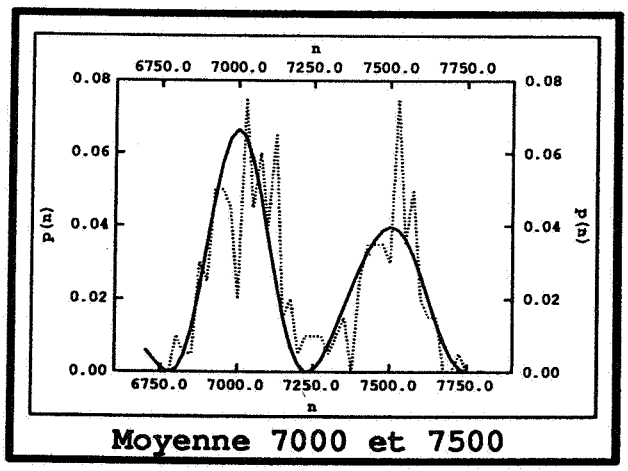


Figure 3.66 Estimateur par régularisation

Il n'est même pas vrai que la poissonisée soit un bon estimateur de la loi sans bruit parce que, si l'on échantillonne avec 50 000 points au lieu de 200, on obtient la figure 3.67 (ici la poissonisée est en pointillé) et on voit que l'estimation sans régularisation n'est pas très bonne. Avant d'obtenir une assez bonne estimation de la loi sous-jacente sans régularisation, il nous faudra encore 10 à 50 fois plus de temps d'acquisition par pixel. Sur la figure 3.68 on voit l'estimateur avec régularisation et avec l'échantillon de 50 000 points, comme on pouvait imaginer, dans une telle situation la régularisation donne des très bon résultats.

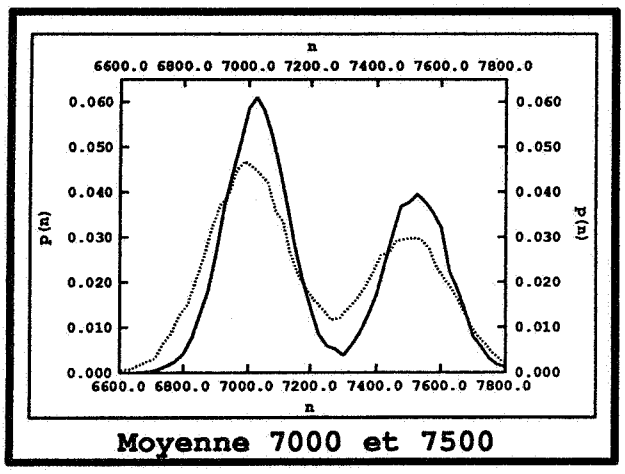


Figure 3.67 Sous-jacente à 50.000 points (trait continu) et sa poissonisée.

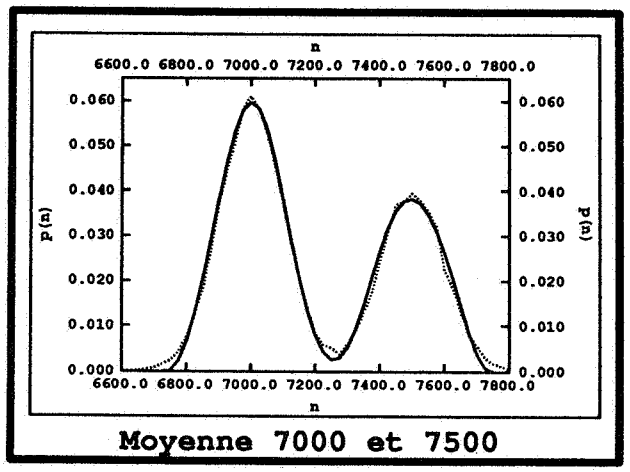


Figure 3.68 Estimateur par régularisation

En conclusion, si la qualité d'une image est mauvaise (rapport signal/bruit inférieur à 1), on peut estimer l'histogramme soit par régularisation, soit en prenant l'histogramme de l'image krigée. En général, la régularisation est plus fiable. Si l'on voulait obtenir une estimation directe de la loi en prenant moins de données, mais avec un temps d'acquisition plus élevé, on serait obligé de multiplier le temps par pixel par un facteur très grand (peut être de l'ordre d'un millier) et très difficile à connaître *a priori*.

CHAPITRE IV

LES PROBLÈMES MAL POSÉS

4.0. — INTRODUCTION

Dans le chapitre précédent, nous avons suggéré que la poissonisation était un exemple de problème mal posé. Dans ce chapitre, nous donnons une définition des problèmes mal posés et quelques exemples seront considérés. Nous examinerons ensuite plus en détail le cas poissonien parce qu'il constitue un problème assez typique, étant stable, au moins théoriquement pour un choix particulier de norme, mais instable pour la norme ℓ^2 dont on a besoin pour les projections utilisées au chapitre précédent. Une conséquence de l'instabilité des estimateurs ℓ^2 est l'instabilité des estimateurs de type maximum de vraisemblance, ce qui sera démontré au paragraphe 4.4.

DÉFINITION. — Soit F et G deux espaces métriques et $A : F \to G$ tel que :

1) A est injective
2) A est continue
3) A^{-1} n'est pas continue.

On dit alors que le problème inverse, à savoir : connaissant $g \in G$, trouvons $f \in F$ tel que $f = A^{-1}g$, est mal posé.

Une petite pertubation de g peut nous conduire à des valeurs très différentes de f à cause de la discontinuité de A^{-1}. Ceci est un problème réel, car dans la pratique on ne peut observer qu'une approximation de g, et même si on connaissait g avec exactitude, l'inversion numérique par ordinateur nous oblige à approximer g et A par la précision de la machine, ce qui suffit dans certains cas à provoquer de grandes variations sur l'estimateur de f.

Considérons maintenant quelques exemples de problèmes mal posés.

1. — Soit E et F des espaces de Banach et $A : E \to F$ une application linéaire et continue. Si ker $A = \{x;\ Ax = 0\}$ est tel que ker $A = \{0\}$, alors A est injective

et $A^{-1} : A(E) \to E$ existe où $A(E)$ est l'image de A. Si $A(E)$ est complet, c'est un espace de Banach et le théorème de l'application ouverte, ([1], voir § 5.2), implique que A^{-1} est continue. Par contre si $A(E)$ n'est pas complète, alors A^{-1} n'est pas continue et le problème inverse est mal posé.

2. — Si $A : C^1[a,b] \to C[a,b]$ est défini par $Af = f'$, alors le problème est mal posé parce que

$$\text{si } f(x) = g(x) + a \text{ Cos } \omega x \quad \Rightarrow f'(x) = g'(x) - a\omega \text{ Sin } \omega x$$

et pour a petit $\|f-g\| = \sup_x |f(x)-g(x)|$ est petit, mais pour ω suffisamment grand, $\|f'-g'\|$ peut être aussi grand qu'on veut. Notons que ce problème peut être stabilisé par un choix de norme de type Sobolev sur l'espace de départ.

3. — Soit

$$Af = \int K(x,y)\, f(y)\, dy$$

avec $K \in L^2(\mathbb{R}^n \times \mathbb{R}^n)$ et $f \in L^2(\mathbb{R}^n)$. Il est bien connu que A est complètement continue ([2]) et que ses valeurs propres n'ont de point limite qu'éventuellement à 0. Si 0 est un point limite mais n'est pas lui-même une valeur propre — ce qui sera le cas s'il n'existe pas de $f \neq 0$ tel que $Af = 0$, alors A a un inverse. Mais puisque 0 est dans le spectre, ce n'est pas un point régulier et par conséquent A^{-1} n'est pas borné.

4. — Comme cas particulier de l'exemple précédent, on considère la convolution dans L^2

$$g = Af = \int_{-\infty}^{\infty} K(x-y)\, f(y)\, dy\,.$$

Prenons la transformée de Fourier $\tilde{g} = \mathcal{F}(g)$ et on obtient $\tilde{g} = \tilde{K}\tilde{f}$, d'où

$$f = \mathcal{F}^{-1}(\tilde{f}) = \mathcal{F}^{-1}\left[\frac{\tilde{g}}{\tilde{K}}\right]$$

si l'inverse existe (i.e $\tilde{K} \neq 0$).

Supposons maintenant que l'on observe une perturbation de g, soit $g' = g + \epsilon$. Alors notre estimateur de f, f', sera

$$f' = \mathfrak{F}^{-1}\left[\frac{\tilde{g}'}{\tilde{K}}\right] = f + \mathfrak{F}^{-1}\left(\frac{\tilde{\epsilon}}{\tilde{K}}\right)$$

mais si ϵ est un bruit blanc et si $\tilde{K}(\omega) \to 0$ quand $\omega \to \infty$, l'estimateur n'existera pas. Dans un calcul numérique nous appliquons un seuillage du spectre et une "solution" existe, mais elle est instable par rapports aux petits changements de point de seuillage.

NOTA. — Le même problème persiste pour d'autres noyaux qui ont un théorème de multiplication pour une transformation, par exemple $\int_0^\infty \frac{1}{y} f(y)\, K(\frac{x}{y})\, dy$ et la transformée de Mellin.

Le lemme topologique élémentaire classique suivant donne une première approche des problèmes mal posés.

LEMME 4.1. — Soit E et F des espaces métriques et $A : E \to F$ continue et telle que A^{-1} existe. Soit $K \subset E$ compact, alors $A(K)$ est compact dans F et $A^{-1}|_{A(K)}$ est continue.

DÉMONSTRATION. — Si $\{A(f_k)\}$ est une suite dans $A(K)$ avec $Af_k \to g \in F$, alors puisque $\{f_k\} \subset K$, il existe une sous-suite $f_{k_n} \to f \in K$. A étant continue, $Af_{k_n} \to Af$. Donc Af_{k_n} est une sous-suite convergente et $A(K)$ est compact.

Supposons maintenant que A^{-1} ne soit pas continue; il existe une suite $g_n \to g$ dans $A(K)$ telle que $A^{-1}g_n \not\to A^{-1}g$. Posons $f_k = A^{-1}g_k$ et $f = A^{-1}g$. Dans la topologie induite sur K par le métrique de E, on peut trouver un ouvert V tel que $f \in V$, et puisque $f_n \not\to f$, il existe une sous-suite $\{f_{n_k}\} \subset K \backslash V$. Notons que $K \backslash V$ est complet. K est compact et dans la sous-suite $\{f_{n_k}\}$, il a lui-même une sous-suite qui converge. Appelons la $f_{k'} \to h \in K \backslash V$. Alors $h \notin V$. Mais puisque A est continue, $Af_{k'} \to Ah$. Donc $Ah = g$ et $h = A^{-1}g \in V$ ce qui mène à une contradiction. Donc A^{-1} est continue. ■

Ce lemme est très important parce que si l'on peut associer les données observées g à un élément de $A(K)$ de manière continue (par exemple par une

projection), l'application inverse $A^{-1}|_K$ étant continue, nous avons une application continue entre les données et K. Mais c'est précisément ce que nous avons fait avec l'estimateur de χ^2 (sans régularisation), et pourtant l'estimation est souvent très mauvaise (nous avons discrétisé le problème pour obtenir un ensemble compact des estimateurs possibles, donc l'application inverse est continue, et les matrices sont non singulières mais elles restent très instables parce que l'ensemble compact n'est pas suffisamment petit). Il semble donc évident que cette façon de procéder ne puisse être utile que lorsque K est un compact suffisamment petit.

Avant de passer au problème de poissonisation, nous aurons besoin de quelques définitions et théorèmes. Pour plus de détails et pour les démonstrations de ces théorèmes, voir [3].

4.1. — DÉFINITIONS ET THÉORÈMES

Soit $\mathcal{M}_+$ l'espace des mesures de probabilités positives (sur $\mathbb{R}^+$). Donc $\mu \in \mathcal{M}_+$ si et seulement si Supp $\mu \subset \mathbb{R}^+$ et $\int \mu = 1$. Si K est un compact dans $\mathbb{R}^+$, nous appelons $\mathcal{M}_{+K}$ l'espace des mesures avec support dans K.

DÉFINITION. — On dit que $\{\mu_n\} \subset \mathcal{M}_+$ converge faiblement vers μ et l'on écrit $\mu_n \xrightarrow{\omega} \mu$ si pour chaque intervalle de continuité borné, nous avons $\mu_n(I) \to \mu(I)$. Dans la littérature la convergence faible s'appelle aussi la convergence vague. En général, μ est tel que $\int \mu \leq 1$. Appelons $\overline{\mathcal{M}_+}$ la fermeture de $\mathcal{M}_+$ pour cette convergence. On dit que $\mu \in \mathcal{M}_+$ est une mesure de probabilité, et que μ est une mesure de probabilité dégénérée si $\mu \in \overline{\mathcal{M}_+} \setminus \mathcal{M}_+$. Si $\mu_n \xrightarrow{\omega} \mu$ et si μ est une mesure de probabilité, on parle d'une convergence complète et on écrit $\mu_n \xrightarrow{C} \mu$ (on l'appelle aussi la convergence étroite). On peut démontrer que ceci est le cas si et seulement si $\mu_n(I) \to \mu(I)$ pour chaque intervalle de continuité (borné ou non borné).

Une formulation équivalente est donnée par le théorème suivant.

THÉORÈME (HELLY-BRAY). – Si les $\{\mu_n\}$ sont des mesures de probabilité

a) $\mu_n \xrightarrow{\omega} \mu$ ssi $E\mu_n[f] \to E\mu[f]$ $\forall f \in C_0(-\infty, \infty)$

où $f \in C_0(-\infty, \infty)$ si f est continue et $f(x) \to 0$ quand $x \to \infty$ or $x \to -\infty$

b) $\mu_n \xrightarrow{C} \mu$ ssi $E\mu_n[f] \to E\mu[f]$ $\forall f \in C(-\infty, \infty)$.

où $f \in C(-\infty, \infty)$ si f est continue et bornée.

NOTA. – $\mathcal{M}_0$ avec la convergence complète est un espace métrique, [3], avec la métrique $d(\mu_1, \mu_2)$ définie par $d(\mu_1, \mu_2)$ est égal à l'infimum de h tel que

$$\mu_1(]-\infty, x-h]) - h \leq \mu_2(]-\infty, x]) \leq \mu_1(]-\infty, x+h]) + h$$

pour chaque x. Il s'agit ici de la métrique de Lévy.

DÉFINITION. – Une séquence de mesures μ_n est équicontinue (dans [3], Feller utilise le terme *stochastiquement borné*), si pour chaque $\epsilon > 0$, $\exists\, N, a$ tel que

$$\mu_n(-a, a) \geq 1 - \epsilon \qquad \forall\, n \geq N$$

et nous avons l'important théorème suivant.

THÉORÈME (HELLY). –

1) Chaque suite $\{\mu_n\} \subset \overline{\mathcal{M}_0}$ a une sous-suite $\{\mu_{n_k}\}$ qui converge faiblement dans $\overline{\mathcal{M}_0}$, $\mu_{n_k} \to \mu \in \overline{\mathcal{M}_0}$. Donc $\overline{\mathcal{M}_0}$ est compact pour la convergence faible.
2) Pour que la limite μ soit une mesure de probabilité, $\mu \in \mathcal{M}_0$, il est nécessaire et suffisant que $\{\mu_n\}$ soit équicontinue.
3) Pour que $\mu_n \xrightarrow{\omega} \mu$, il est nécessaire et suffisant que chaque sous-suite convergente converge vers μ.

Définissons maintenant l'idée d'une famille équicontinue.

DÉFINITION. – On dit qu'une famille $\mathcal{M}_{0S}$ des mesures de probabilité est équicontinue si une suite $\{\mu_n\} \subset \mathcal{M}_{0S}$ qui converge faiblement est équicontinue, et donc par le théorème de Helly, elle converge fortement également.

Notons que la définition s'applique aux familles dans $\mathcal{M}_b$. En effet, si μ est une mesure de probabilité dégénérée, alors μ ne peut pas appartenir à une famille équicontinue parce que la suite $\{\mu_n\}$ où $\mu_n = \mu \quad \forall\, n$ converge faiblement mais n'est pas équicontinue.

Notons aussi que si $\mathcal{M}_{bS}$ est fermé alors le théorème de Helly implique que $\mathcal{M}_{bS}$ est compact pour la convergence complète. Finalement pour chaque suite $\{\mu_n\} \subset \mathcal{M}_{bS}$, convergente ou non, il est vrai que $\{\mu_n\}$ est équicontinue parce que sinon, pour $k = 1, 2,\ldots$, il existe $\epsilon > 0$, $\epsilon < 1$ et une sous-suite $\{\mu_{n_k}\}$ telle que $\mu_{n_k}(-k,k) < 1-\epsilon \quad \forall\, k$. Par le théorème de Helly, $\{\mu_{n_k}\}$ a elle-même une sous-suite convergente, mais celle-ci ne peut pas être équicontinue, ce qui contredit l'équicontinuité de $\mathcal{M}_{bS}$.

Les résultats de compacticité sont regroupés dans le lemme suivant.

LEMME 4.2. —

1) $\overline{\mathcal{M}_b}$ est compact pour la convergence faible
2) $\mathcal{M}_b$ est compact pour la convergence complète
3) Si $\mathcal{M}_{bS}$ est une famille équicontinue, la convergence faible implique la convergence complète. De plus, si $\mathcal{M}_{bS}$ est fermé alors $\mathcal{M}_{bS}$ est compact pour les deux topologies.

Le lemme suivant nous donne une méthode pour générer des familles équicontinues convexes. L'addition et la multiplication scalaire des mesures sont définies de manière évidente.

LEMME 4.3. — La fermeture de l'enveloppe convexe d'une famille équicontinue est elle aussi équicontinue.

DÉMONSTRATION. — Soit B une famille équicontinue et

$$\mathfrak{B} = \left\{ \sum_{i=1}^{N} \lambda_i \mu_i \; ; \; \mu_i \in B,\; N \text{ fini},\; \sum \lambda_i = 1, \lambda_i \geq 0 \right\}$$

et $C(B) = \overline{\mathfrak{B}}$ la fermeture de l'enveloppe convexe de B.

Nous allons d'abord démontrer que $\mathfrak{B}$ est équicontinue, et ensuite un argument de type diagonal nous donnera le résultat pour $C(B)$.

Si $\mu_n \xrightarrow{\omega} \mu$ avec $\mu_n \in \mathfrak{B}\ \forall\, n$, donc $\mu_n = \sum_{i=1}^{N_n} \lambda_i^k\, \mu_i^k$, avec $\mu_i^k \in B$. Si pour $a \in \mathbb{R}$

$$\mu_i^k(-a,a) \geq 1-\epsilon \qquad \forall\, i = 1,\ldots,N_k$$

alors $\mu_k(-a,a) = \Sigma\, \lambda_i^k\, \mu_i^k\,(-a,a) \geq \Sigma\, \lambda_i^k\,(1-\epsilon) = 1-\epsilon$.

Puisque $\mu_k = \Sigma\, \lambda_i^k\, \mu_i^k$, on peut considérer la suite

$$\left\{\mu_1^1, \mu_2^1, \ldots, \mu_{N_1}^1,\ \mu_1^2, \ldots, \mu_i^k, \ldots\right\} \subset B$$

Or nous avons déjà noté que chaque suite en B est équicontinue et il existe donc K tel que

$$\mu_k(-a,a) = \Sigma\, \lambda_i^k(-a,a) \geq 1-\epsilon \qquad \forall\, k \geq K$$

La suite $\{\mu_k\} \subset \mathfrak{B}$ est alors équicontinue.

Maintenant $\mu_n \xrightarrow{\omega} \mu$ avec $\mu_n \in C(B)$. Pour chaque μ_n, il existe une suite $\{\mu_n^k\} \subset \mathfrak{B}$ telle que $\mu_n^k \xrightarrow{C} \mu_n$. Maintenant si $\mu_n \overset{C}{\not\to} \mu$, il existe un intervalle de continuité I tel que $E_{\mu_n}(I) \not\to E_\mu(I)$ et il existe une sous-suite $\bar{\mu}_k$ telle que $d(E_{\bar{\mu}_k}(I),\ E_\mu(I)) > \epsilon\ \forall\, k$. Pour chaque terme dans la sous-suite, on peut choisir $\ell \geq k$ tel que $d\left(E_{\bar{\mu}_k^\ell}(I),\ E_{\bar{\mu}_k}(I)\right) < \frac{\epsilon}{2}$ où $\{\bar{\mu}_k^\ell\}$ est une suite dans $\mathfrak{B}$ avec $\bar{\mu}_k^\ell \xrightarrow{C} \bar{\mu}_k$. On sait que $\bar{\mu}_k^\ell \xrightarrow{\omega} \mu$ puisque la suite diagonale converge $\bar{\mu}_k^k \xrightarrow{\omega} \mu$. De plus, puisque $\{\bar{\mu}_k^\ell\}$ est équicontinue, la convergence est complète et $\bar{\mu}_k^\ell \xrightarrow{C} \mu$.

En revanche, l'inégalité triangulaire donne

$$d\left(\mathrm{E}_\mu(I),\ \mathrm{E}_{\bar{\mu}_k}(I)\right) \leq d\left(\mathrm{E}_\mu(I),\ \mathrm{E}_{\bar{\mu}_k^\ell}(I)\right) + d\left(\mathrm{E}_{\bar{\mu}_k^\ell}(I),\ \mathrm{E}_{\bar{\mu}_k}(I)\right) \qquad \forall\, k$$

et donc

$$d\left(\mathrm{E}_\mu(I),\ \mathrm{E}_{\bar{\mu}_k^\ell}(I)\right) \geq \frac{\epsilon}{2} \qquad \forall\, k$$

ce qui contredit la convergence de $\tilde{\mu}_k^{\ell}$ vers μ. On peut conclure que $\mu_n \xrightarrow{C} \mu$ et donc $\{\mu_n\}$ est équicontinue. ■

Nous avons besoin de deux théorèmes classiques de l'analyse complexe (voir par exemple [4]).

DÉFINITION. – Ω est une région si c'est un ensemble non vide, ouvert et connexe dans le plan complexe.

THÉORÈME. – Si f et g sont des fonctions holomorphes dans une région Ω et si $f(z) = g(z)$ $\forall\, z \in S$ où S est un ensemble ayant un point limite dans Ω, alors $f(z) = g(z)$ $\forall\, z \in \Omega$.

THÉORÈME (MORERA). – Soit f une fonction continue complexe dans un ouvert Ω telle que $\int_{\partial\Delta} f(z)\, dz = 0$ pour chaque triangle fermé Δ. Alors f est holomorphe dans Ω.

LEMME 4.4. – La transformée de Laplace d'une mesure finie est uniquement déterminée par une suite de points $\{z_n\} \subset \Omega$ où $\Omega = \{z \in \mathbf{C};\ \mathrm{Re}\, z > 0\}$ telle que $\{z_n\}$ a un point limite dans Ω.

DÉMONSTRATION. – Soit $L_\mu(z) = \int_0^\infty e^{-zx}\, d\mu(x)$ la transformée de Laplace d'une mesure finie μ. Soit Δ un triangle dans Ω.

$$\int_{\partial\Delta} \mathrm{L}_\mu\, dz = \int_{\partial\Delta} dz \left(\int_0^\infty \mathrm{e}^{-zx}\, d\mu(x) \right)$$

Puisque $|\mathrm{e}^{-zx}| \leq 1 \quad \forall\, x > 0$ et $z \in \Omega$, et puisque μ est fini, on peut appliquer le théorème de Fubini

$$\int_{\partial\Delta} \mathrm{L}_\mu(z)\, dz = \int_0^\infty d\mu(x) \int_{\partial\Delta} \mathrm{e}^{-zx}\, dz = 0$$

l'integrale est nulle parce que e^{-zx} est une fonction entière. Comme $\mathrm{L}_\mu(z)$ est évidemment une fonction continue dans Ω, le théorème de Morera donne que

$L_\mu(z)$ est holomorphe dans Ω et le théorème précédent nous permet de conclure que la transformée de Laplace est déterminée uniquement par une suite de points avec une limite dans Ω. ■

4.2. — POISSONISATION ET CONTINUITE DANS ℓ^1

On peut brièvement résumer les arguments de cette section et de celle qui suit. On démontre que la poissonisation transforme des suites faiblement convergentes en suites qui convergent point par point. La réciproque est également vraie. Pour l'espace ℓ^1, on démontre facilement que la convergence point par point implique une convergence forte et que $P(\mathcal{M}_b)$ est compact dans ℓ^1. Dans ℓ^2, $P(\mathcal{M}_b)$ n'est même pas complet mais il se trouve que l'image d'une famille équicontinue complète est compacte dans ℓ^2.

4.2.1. — Notation

On peut définir une application $P : \mathcal{M}_b \to \mathcal{M}_b$ par $\mu \in \mathcal{M}_b$; $P(\mu) = p$ est une mesure discrète sur les entiers positifs et sur zéro, donnée par

$$p(n) = \int_0^\infty e^{-x} \frac{x^n}{n!}\, d\mu(x).$$

Il s'agit bien sûr de l'application de poissonisation. Posons $\mathcal{P}$ l'espace image, $\mathcal{P} = P(\mathcal{M}_b)$. P se prolonge facilement sur $\overline{\mathcal{M}_b}$ et on appelle $\overline{\mathcal{P}} = P(\overline{\mathcal{M}_b})$.

4.2.2. — Existence et continuité de P^{-1}

LEMME 4.5. — L'application de la poissonisation $P : \overline{\mathcal{M}_b} \to \overline{\mathcal{M}_b}$ satisfait à

1) P est linéaire
2) P est injective.

DÉMONSTRATION. — La linéarité est évidente. Il reste à démontrer l'injectivité. Soit $\mu \in \overline{\mathcal{M}_b}$ et $p \in P(\mu)$, la fonction génératrice de p, $G_p(s) = E[s^p]$ est (avec un léger abus de notation) :

$$G_p(s) = E[s^p] = E[E[s^p|\mu]] = E[e^{(s-1)\mu}] = L_\mu(1-s) \qquad 0 \le s \le 1$$

où L_μ est la transformée de Laplace de μ. Si μ_1 et μ_2 sont telles que $P(\mu_1) = p = P(\mu_2)$, alors

$$L_{\mu_1}(s) = L_{\mu_2}(s) \qquad 0 \le s \le 1$$

et le lemme 4.4 nous donne que les transformées sont identiques pour $s \ge 0$ et par suite le théorème de l'inversion pour les transformées de Laplace donne $\mu_1 = \mu_2$ pp (presque partout) et l'injectivité est démontrée. ■

Il est clair que $\overline{\mathcal{P}} \subset \ell^1$ puisque $p \in \overline{\mathcal{P}} \Rightarrow \Sigma\, p(k) \le 1$ (par le théorème de Fubini). Soit $\{p_n\}$ une suite de points dans $\overline{\mathcal{P}} \subset \ell^1$; on dit que que p_n converge ponctuellement et on écrit $p_n \xrightarrow{\omega} p$ si $p_n(k) \to p(k)$ pour chaque $k = 0, 1, 2...$ On dit que la convergence est forte et on écrit $p_n \to p$ si $\| p_n - p \|_1 \to 0$ où $\| p \|_1$ est la norme sur ℓ^1. Notons que la convergence ponctuelle est générée par la topologie produite. Notons aussi le résultat classique [5] : Pour un espace ℓ^q où $q = 1, 2, 3...$, si $\{p_n\}$ et $p \in \ell^q$ et $p_n \xrightarrow{\omega} p$, alors $p_n \to p$ si et seulement si $\| P_n \|_q \to \| p \|_q$.

Si nous considérons $\overline{\mathcal{P}} \subset \ell^1$ nous avons le théorème suivant.

THÉORÈME. – Si $\mathcal{P}$ est considéré comme un ensemble de ℓ^1, alors $\mathcal{P}$ est compact dans ℓ^1 et l'application P est continue. De plus, l'application inverse est continue.

Ceci est une conséquence des deux lemmes suivants.

LEMME 4.6. – Si $\{\mu_n\} \subset \overline{\mathcal{M}}$ et $p_n = P(\mu_n)$ alors

$$\mu_n \xrightarrow{\omega} \mu \Leftrightarrow p_n \xrightarrow{\omega} p$$

De plus, chaque suite $\{p_n\} \subset \overline{\mathcal{P}}$ admet une sous-suite convergente dans $\overline{\mathcal{P}}$.

DÉMONSTRATION. – Soit $\mu_n \to \mu$. Choisissons $f \in C_0(-\infty, \infty)$ par

$$f(x) = \begin{cases} e^{-x}\dfrac{x^n}{n!} & x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Le théorème de Helly-Bray nous donne

$$p_n(m) = \mathrm{P}_\mu(m) = \mathrm{E}_{\mu_n}\left[e^{-y}\frac{y^m}{n!}\right] \to \mathrm{E}_\mu\left[e^{-y}\frac{y^m}{n!}\right]$$

Donc $\mathrm{P}(\mu_n) \xrightarrow{\omega} \mathrm{P}(\mu)$.

Réciproquement, si $p_n \xrightarrow{\omega} p$ dans $\overline{\mathcal{P}}$, alors $\mu_n = \mathrm{P}^{-1}(p_n)$ est une suite dans $\overline{\mathcal{M}_0}$. Soit μ_{n_k} une sous-suite convergente $\mu_{n_k} \xrightarrow{\omega} \mu$. Par la première partie, $\mathrm{P}(\mu_{n_k}) \xrightarrow{\omega} \mathrm{P}(\mu)$ et donc $\mathrm{P}(\mu) = p$. Si μ'_{n_k} est une autre sous-suite convergente $\mu'_{n_k} \xrightarrow{\omega} \mu'$, on a $\mathrm{P}(\mu) = p = \mathrm{P}(\mu')$ et le lemme 4.5 implique que $\mu = \mu'$ *pp*. Donc toutes les sous-suites convergentes ont la même limite d'où, par le théorème de Helly, $\mu_n \xrightarrow{\omega} \mu$.

Finalement, si $\{p_n\}$ est une suite dans $\overline{\mathcal{P}}$ alors $\left\{\mathrm{P}^{-1}(p_n)\right\} \subset \mathcal{M}_0$. On peut extraire une sous-suite convergente et en appliquant P, on trouve notre sous-suite convergente $\{p_{n_k}\}$. ■

LEMME 4.7. – Si $\{\mu_n\} \subset \mathcal{M}_0$ et $p_n = \mathrm{P}(\mu_n)$ alors $\mu_n \xrightarrow{C} \mu \Leftrightarrow p_n \to p$.

DÉMONSTRATION. –

$\Rightarrow$ $\mu_n \xrightarrow{C} \mu \Rightarrow p_n \xrightarrow{\omega} p$ par le lemme précédent. Il faut démontrer que $\|p_n\|_1 \to \|p\|$. Mais ceci est évident puisque μ_n et μ sont des mesures de probabilité et $\Sigma\, p(n) = \Sigma \int e^{-x}\frac{x^n}{n!}\, d\mu = \int d\mu = 1$. P est donc continue.

$\Leftarrow$ Si $p_n \to p \Rightarrow \mu_n \xrightarrow{\omega} \mu$ et $\mathrm{P}(\mu) = p$ mais $\|p\|_1 = 1 = \int \mu$, μ est une mesure de probabilité et $\mu_n \xrightarrow{C} \mu$. Donc P^{-1} est continue. ■

Le problème inverse pour la poissonisation n'est pas mal posé dans ℓ^1. Malheureusement, il n'existe pas de méthode simple pour associer à un échantillon $\{g(n)\}$ un point de $\mathcal{P}$ de manière continue. Il sera beaucoup plus facile d'avoir un théorème de projection. Nous allons donc chercher une solution dans ℓ^2 ou plus généralement dans $\ell^2(\nu)$ (moindres carrés avec ponderateur ν) pour un choix de norme ν.

4.3. — POISSONISATION ET CONTINUITE DANS ℓ^2

Il n'est malheureusement pas vrai que P^{-1} soit continu dans ℓ^2 comme on peut le voir avec l'exemple suivant.

EXEMPLE. — Soit $\mu_n = \delta_n$, $n = 1, 2, \ldots$ où δ_n est la mesure de probabilité qui attribue le poids 1 à n. Alors $\mu_n \in \mathcal{M}_0$ $\forall\, n$ et $\mu_n \xrightarrow{\omega} 0$.

Posons $P(\mu_n) = p_n$. Alors

$$p_n(m) = e^{-n}\,\frac{n^m}{m!}$$

Nous avons évidemment $p_n \xrightarrow{\omega} 0$ parce que $p_n(m) \to 0$. Mais il est vrai que $p_n \to 0$ dans ℓ^2 parce que

$$\|p_n\|_2^2 = \sum_{m=0}^{\infty}\left[e^{-n}\,\frac{n^m}{m!}\right]^2 = e^{-2n}\sum_{m=0}^{\infty}\frac{n^{2m}}{(m!)^2}$$

Mais $I_0(2n) = J_0(2in) = \sum_{m=0}^{\infty}\frac{n^{2m}}{(m!)^2}$, où $J_0(2in)$ est la fonction de Bessel de première espèce. Si l'on considère un développement asymptotique pour $I_0(z)$ pour $\frac{-\pi}{2} \le \arg z \le \frac{3\pi}{2}$, nous avons

$$I_0(z) \sim \frac{e^{-z}}{\sqrt{2\pi z}}\sum_0^{\infty}\frac{(-1)^k}{(2z)^k}\,\frac{\Gamma(k+\frac{1}{2})}{k!\,\Gamma(\frac{1}{2}-k)} + \frac{e^{-(z-\frac{i\pi}{2})}}{\sqrt{2\pi z}}\sum_0^{\infty}\frac{1}{(2z)^k}\,\frac{\Gamma(k+\frac{1}{2})}{k!\,\Gamma(\frac{1}{2}-k)}$$

et si l'on prend le terme principal, on obtient

$$\lim_{n\to\infty}\left\{I_0(2n)-\frac{e^{2n}+ie^{-2n}}{\sqrt{4\pi n}}\right\}\to 0$$

Donc
$$\lim_{n\to\infty} e^{-2n}\left(I_0(2n)-\frac{e^{2n}+ie^{-2n}}{\sqrt{4\pi n}}\right)\to 0$$

et
$$\lim_{n\to\infty} e^{-2n}\sum_{n=0}^{\infty}\frac{n^{2m}}{(m!)^2}=\lim_{n\to\infty}\frac{1+ie^{-4n}}{\sqrt{4\pi n}}=0$$

Nous avons donc $\|p_n\|_2^2\to 0$ et par conséquent $p_n\to 0$ dans ℓ^2.

On voit dans cet exemple que

$$p_n\to p \not\Rightarrow \mu_n \overset{C}{\to} \mu$$

et l'application inverse P^{-1} n'est pas continue.

NOTA. – Dans le dernier exemple, $\|p_1\|_1=1 \quad \forall\, n$ et il n'y a pas de convergence forte ℓ^1 comme prévu.

On peut réimposer la continuité si on se limite aux familles équicontinues.

LEMME 4.8. – Soit $\mathcal{M}_S$ une famille équicontinue complète et $P : \mathcal{M}_S \to \ell^2$. Alors

1) $P(\mathcal{M}_S)\subset \ell^2$
2) $P|_{\mathcal{M}_S}$ est continue
3) $P(\mathcal{M}_S)$ est complète
4) $P^{-1} : P(\mathcal{M}_S)\to \mathcal{M}_S$ est continue.

DÉMONSTRATION. –

1) $p\in P(\mathcal{M}_S)$ alors $\Sigma\, p(n)^2 \le \Sigma\, p(n) = 1$ et $p\in\ell^2$
2) On sait que $P(\mathcal{M}_S)\subset \ell^1$. Mais $p_n \overset{1}{\to} p \Rightarrow p_n \overset{2}{\to} p$ puisque

$\sum_n |p_n(m)-p(m)|^2 \leq \sum_n |p_n(m)-p(m)|$ pour N suffisamment grand, et le résultat suit facilement. (Voir lemme 4.10 pour la démonstration complète dans un cas un peu plus compliqué).

3 et 4) Si $p_n \xrightarrow{2} p$ alors $p_n \xrightarrow{\omega} p$ et donc $P^{-1}(p_n) \xrightarrow{\omega} P^{-1}(p)$. Mais $\mathcal{M}_S$ est équicontinue et par suite la convergence est complète. Donc $p \in \mathcal{P}$ et $P^{-1}(p) \in \mathcal{M}$, d'où la continuité de P^{-1}. ■

Pour des espaces $\ell^2(\nu)$, il existe une légère difficulté parce que $P(\mathcal{M}_S)$ n'est pas obligatoirement un sous-ensemble de $\ell^2(\nu)$. Admettons que ceci soit le cas, alors les parties 3) et 4) du lemme 4.8 restent valides sans changer l'argument et on a :

COROLLAIRE. – Si $\mathcal{M}_S$ est une famille équicontinue et si $P(\mathcal{M}_S) \subset \ell^2(\nu)$ pour une norme ν, alors

1) $P(\mathcal{M}_S)$ est complète
2) $P^{-1} : P(\mathcal{M}_S) \to \mathcal{M}_S$ est continue.

Pour que P soit elle-même continue, il faut que $P(\mathcal{M}_S)$ soit compacte dans $\ell^2(\nu)$.

LEMME 4.9. – Si $\mathcal{M}_S$ est une famille équicontinue complète et si $P(\mathcal{M}_S) \subset \ell^2(\nu)$ pour une norme ν, alors P est continue si et seulement si $P(\mathcal{M}_S)$ est compact.

DÉMONSTRATION. –

$\Rightarrow$ $\mathcal{M}_S$ est un espace métrique compact. Pour P continue, $P(\mathcal{M}_S)$ est compacte.

$\Leftarrow$ Soit $\mu_n \xrightarrow{C} \mu$ dans $\mathcal{M}_S$ et $p_n = P(\mu_n)$, $p = P(\mu)$. $\{p_n\} \subset P(\mathcal{M}_S)$ donc on peut extraire une sous-suite convergente. L'argument du lemme 4.6 nous montre que toute sous-suite convergente tend vers p. Il est facile de démontrer que $p_n \to p$ et par suite P est continue (sinon il existerait une sous-suite en dehors du voisinage de p, cette sous-suite ayant elle-même une sous-suite convergente qui devrait tendre vers p, ce qui est une contradiction). ■

L'ensemble des résultats nous fournit une méthode pour estimer μ connaissant un échantillon fini, g, qui est donc une perturbation de la mesure poissonisée $P(\mu)$.

La démonstration du théorème est une conséquence immédiate du théorème de projection et du fait que P est linéaire et donc elle transforme des ensembles convexes en ensembles convexes.

THÉORÈME. – Soit q un échantillon d'une mesure poissonisée $P(\mu) = p$. Si ν est une norme pour $\ell^2(\nu)$ et si $\mathcal{M}_S$ est une famille équicontinue convexe complète telle que $P(\mathcal{M}_S) \subset \ell^2(\nu)$, alors il existe une mesure $\hat{\mu} \in \mathcal{M}_S$ unique telle que $\chi^2_\nu = \sum (q(n) - P\hat{\mu}(n))^2 \nu(n)$ est minimisé. De même l'application $A : \ell^2(\nu) \to \mathcal{M}_S$ définie par $A(q) = \hat{\mu}$ est continue.

Avant de terminer, examinons un exemple de ν pour lequel $P(\mathcal{M}_S) \subset \ell^2(\nu)$. On sait déjà qu'avec ℓ^2 on peut choisir pour $\mathcal{M}_S$ n'importe quelle famille équicontinue convexe fermée. Le résultat suivant nous donne une classe des espaces $\ell^2(\nu)$ et nous démontre que les estimateurs utilisés au Chapitre III sont cohérents.

LEMME 4.10. – Soit $\mathcal{M}_K$ l'espace des mesures avec support compact dans $K = [0, M]$ et $\mathcal{P}_K = P(\mathcal{M}_K)$. Alors $P(\mathcal{M}_K)$ est compacte dans l'espace de Hilbert $\ell^2(1/f)$ où $\{f(n)\}$ est une suite telle que $f(n) > 0 \ \forall\, n$ et telle qu'il existe un entier N et $p \in \mathcal{P}$ (avec $p \neq P(\delta_0)$) tel que $\forall\, n \geq N$, $f(n) \geq p(n)$.

DÉMONSTRATION. – Si $q \in \mathcal{P}_K$ avec $q = P(\mu)$, alors supp $\mu \subset [0, M]$ et

$$q(n) = \int_0^M e^{-x} \frac{x^n}{n!}\, d\mu \leq \frac{M^n}{n!} \qquad \forall\, n .$$

Soient f, N et p définis comme dans l'énoncé du lemme, et soit ν tel que $p = P(\nu)$

$$p(n) = \int_0^\infty e^{-x} \frac{x^n}{n!}\, d\nu(x)$$

Puisque $\nu \neq \delta_0$ il existe $\epsilon > 0$ tel que $\nu([\epsilon, \infty[) > 0$ et ν n'étant pas une loi de probabilité dégénérée, il existe $K > 0$ tel que $\nu([\epsilon, K]) = \delta > 0$. Alors

$$p(n) \geq \int_\epsilon^K e^{-x} \frac{x^n}{n!}\, d\nu \geq e^{-K} \frac{\epsilon^n}{n!}\, \delta$$

et la norme de q dans $\ell^2(1/f)$ est

$$\| q \|_{1/f}^2 = \sum_0^\infty \frac{q(n)^2}{f(n)} = \sum_0^N \frac{q(n)^2}{f(n)} + \sum_{N+1}^\infty \frac{q(n)^2}{f(n)}$$

Posons $C = \sum_0^N \frac{q(n)^2}{f(n)}$.

Nous avons

$$\| q \|_{1/f}^2 \leq C + \sum_{N+1}^\infty \left[\left(\frac{M^n}{n!} \right)^2 / \left(e^{-K} \frac{\epsilon^n}{n!} \delta \right) \right]$$

$$\leq C + \frac{e^K}{\delta} \sum_{N+1}^\infty \frac{(M^2/\epsilon)^n}{n!} \tag{4.1}$$

$$\leq C + \frac{e^{-K}}{\delta} e^{M^2/\epsilon}$$

et donc $q \in \ell^2(1/f)$ pour chaque $q \in \mathcal{P}_K$.

Démontrons maintenant que $\mathcal{P}_K$ est compact. Soit $\{q_n\} \subset \mathcal{P}_K$. On sait qu'on peut extraire une sous-suite qui converge ponctuellement (lemme 4.6), $q_{n_k} \xrightarrow{\omega} q$. Nous allons démontrer que la convergence est forte.

Prenons $\epsilon > 0$. A cause de l'équation (4.1), il existe N' tel que, pour chaque $r \in \mathcal{P}_K$, le vecteur $r' \in \mathcal{P}_K$ défini par $r' = (r(1), \ldots, r(N'), 0, \ldots)$ est tel que

$$\| r - r' \| < \frac{\epsilon}{4}$$

et N' est indépendant de r.

Puisque $q_{n_k} \xrightarrow{\omega} q$, nous avons $q_{n_k}(m) \to q(m)$ $\forall\, m$. En particulier, choisissons N'' tel que

$$\max_{m \in [0,N']} \frac{|q_{n_k}(m) - q(m)|^2}{f(m)} < \frac{\epsilon}{2N'} \qquad \forall\, n_k \geq N''$$

ce qui est toujours possible parce que N' est fini.

Nous avons

$$\sum_{n=1}^{N'} \frac{|q_{n_k}(m) - q(m)|^2}{f(m)} < \frac{\epsilon}{2}$$

Donc pour $n_k \geq \max(N', N'')$ il est vrai que

$$\begin{aligned} \| q - q_{n_k} \| &= \| (q - q') + (q' - q'_{n_k}) + (q'_{n_k} - q_{n_k}) \| \\ &\leq \| q - q' \| + \| q' - q'_{n_k} \| + \| q'_{n_k} - q_{n_k} \| \\ &< \frac{\epsilon}{4} + \frac{\epsilon}{2} + \frac{\epsilon}{4} = \epsilon \end{aligned}$$

et la convergence est forte. $\mathcal{P}_K$ est compact. ■

Dans la pratique, nous admettrons que notre famille de mesures de probabilité $\mathcal{M}_{bK}$ ne dépend que d'un nombre fini de paramètres. Par exemple, dans le chapitre III, nous avons considéré des lois de la forme

$$F(dx) = \sum_{i=1}^{N} \lambda_i \, \delta_{x_i}(dx)$$

avec $\lambda_i \geq 0$ et $\Sigma \lambda_i = 1$. On peut bien sûr étudier des problèmes avec des mesures autres que δ_{x_i}, mais de toute manière, pour un calcul numérique, il faut admettre un nombre de paramètres fini. Si ce nombre est petit, il est fort possible qu'on aura de bonnes propriétés de continuité pour P^{-1}, mais s'il est grand, on sait que P^{-1} devient de plus en plus mal conditionnée. On peut se demander si un estimateur de maximum de vraisemblance est meilleur que l'estimateur χ^2. Cette question est traitée à la section suivante.

4.4. — ESTIMATION PAR MAXIMUM DE VRAISEMBLANCE

4.4.1. — Les conditions de Kuhn et Tucker

Rappelons les conditions nécessaire de Kuhn et Tucker pour qu'une fonction $f(\underline{a})$ réalise un minimum où $\underline{a} = (a_1, \ldots, a_n)$ doit respecter les contraintes

$$g_i(\underline{a}) \geq 0 \qquad i = 1,\ldots,\ell$$
$$h_j(\underline{a}) = 0 \qquad j = 1,\ldots,m \tag{4.2}$$

Les conditions de Kuhn et Tucker sont valables dès que les conditions de régularité suivantes sont respectées.

Soit I l'ensemble des indices par lesquels $g_i(a_{\min}) = 0$ (les contraintes actives). On admet que toutes les fonctions du problème sont dérivables en a_j et que les vecteurs

$$\left[\frac{\partial g_i(a_{\min})}{\partial a_0},\ldots,\frac{\partial g_i(a_{\min})}{\partial a_n}\right] \qquad \forall\, i \in \mathrm{I}$$

et

$$\left[\frac{\partial h_i(a_{\min})}{\partial a_0},\ldots,\frac{\partial h_j(a_{\min})}{\partial a_n}\right]_j \qquad \text{pour } i = 1,\ldots,m$$

sont linéairement indépendants.

Si on admet que ces conditions sont vérifiées, alors une condition nécessaire pour que f ait un minimum local sujet aux contraintes (4.2) est qu'il existe des paramètres de Lagrange $\mu_1,\ldots,\mu_\ell$ et $\lambda_1,\ldots,\lambda_m$ tels que

$$\begin{aligned}
&\frac{\partial f(\underline{a})}{\partial a_j} + \sum \mu_i \frac{\partial g_i}{\partial a_j}(\underline{a}) + \sum \lambda_i \frac{\partial h_i}{\partial a_j}(\underline{a}) = 0 && j = 1,\ldots,N\\
&\mu_i\, g_i(\underline{a}) = 0 && i = 1,\ldots,\ell\\
&\mu_i \geq 0 && i = 1,\ldots,\ell\\
&h_i(\underline{a}) = 0 && i = 1,\ldots,m
\end{aligned} \tag{4.3}$$

4.4.2. — Le maximum de vraisemblance

Supposons que la famille convexe équicontinue $\mathcal{M}_S$ des estimateurs possibles de mesure sous-jacente soit déjà choisie, et que chaque $\mu \in \mathcal{M}_S$ soit paramétrisé par n paramètres $(a_1,\ldots,a_n) = \underline{a}$. De plus nous admettons qu'il existe ℓ conditions de la forme

$$g_i(\underline{a}) \geq 0 \qquad i = 1,\ldots,\ell \tag{4.4}$$

qui nous assurent que si $\mu(\underline{a})$ est une mesure positive telle que

$$\int d\mu\ (\underline{a},x) = 1 \tag{4.5}$$

alors $\mu\ (\underline{a}) \in \mathcal{M}_S$. Il est nécessaire que notre estimation de la mesure vraie prenne en compte (4.4) et (4.5).

Maintenant admettons qu'on ait N données observées tirées selon la loi poissonisée, chaque observation étant un entier positif ou nul. On peut construire la fonction de vraisemblance pour cette occurrence.

Puisque $p(n) = \int e^{-x} \frac{x^n}{n!}\ d\mu\ (\underline{a},x)$, la fonction de vraisemblance est

$$\mathrm{L}(\underline{a}) = \prod_{n=0}^{K} \left[\int e^{-x} \frac{x^n}{n!}\ d\mu\ (\underline{a},x) \right]^{g(n)}$$

où K est le nombre de classes dans lesquelles il y a une observation et $g(n)$ est le nombre des observations dans la n$^{\text{ième}}$ classe (autrement dit $g(n)$ est le nombre des observations qui prend la valeur n).

Posons $$H(\underline{a},n) = \int e^{-x} \frac{x^n}{n!}\ d\mu\ (\underline{a},x)$$

et admettons que $H(\underline{a},n)$ est dérivable par rapport à a_i

$$H^{(i)}(\underline{a},n) = \frac{\partial}{\partial a_i} \int e^{-x} \frac{n^x}{n!}\ d\mu\ (\underline{a},x)$$

Si $\ell(\underline{a}) = \log L(\underline{a}) = \sum_{n=0}^{K} g(n) \log H(\underline{a},n)$, on a $\frac{\partial \ell(\underline{a})}{\partial a_i} = \sum_{n=0}^{K} g(n) \frac{H^{(i)}(\underline{a},n)}{H(\underline{a},n)}$

et les équations de Kuhn et Tucker deviennent

$$\sum_n g(n) \frac{H^{(j)}(\underline{a},n)}{H(\underline{a},n)} + \sum_i \mu_i \frac{\partial g_i}{\partial a_j}(\underline{a}) + \lambda \frac{\partial}{\partial a_j} \int d\mu(\underline{a},x) = 0 \qquad j = 1,\ldots,n$$

$$\mu_i\ g_i(\underline{a}) = 0 \qquad i = 1,\ldots,\ell$$

$$\mu_i \geq 0 \qquad i = 1,\ldots,\ell$$

$$\int d\mu(\underline{a},x) = 1$$

NOTA 1. – Le maximum de vraisemblance est unique parce que $\ell(\underline{a})$ est strictement concave par rapport à $H(\underline{a}, n)$ et par suite il admet un maximum unique sur le convexe compact $\mathcal{P}_{\mathcal{M}_S}$.

NOTA 2. – Le maximum de vraisemblance nous donne un point dans $\mathcal{P}_{\mathcal{M}_S}$ (les valeurs de $H(\underline{a}, n)$) et on passe à $\mathcal{M}_S$ indirectement à l'aide de P^{-1}. On attend donc des instabilités numériques pour l'estimateur de maximum de vraisemblance quand le nombre de paramètres $(a_1, \ldots, a_n)$ est élevé.

4.4.3. – Le χ^2 à point fixe

Soit $\mathcal{M}_S$ une famille équicontinue telle que $P(\mathcal{M}_S) \subset \ell^2(1/f)$ pour une suite $\{f(n)\}$. Si $\hat{\mu}$ l'estimateur de $\chi^2_{1/f}$ (§ 4.3) et si de plus $P(\hat{\mu}) = f$, on dit que f est un point fixe pour le problème χ^2.

Ecrivons les conditions de Kuhn et Tucker pour une solution χ^2 et appliquons les au cas où "par hasard" nous avons choisi une norme f qui est un point fixe.

Pour une norme ω, si l'on garde la notation développée pour le maximum de vraisemblance, on a

$$\chi^2_\omega = \sum_0^\infty \left(g(n) - \hat{G}(n)\right)^2 \omega(n)$$

où $\hat{G}(n) = P(\hat{\mu}) = H(\underline{a}, n)$. On veut minimiser χ^2_ω tel que

$$g_i(\underline{a}) \geq 0 \qquad i = 1, \ldots, \ell$$

$$\int d\hat{\mu}(\underline{a}, x) = 1$$

Donc $$\frac{\partial^2 \chi_\omega}{\partial a_i} = \sum_0^\infty 2(g(n) - H(\underline{a}, n))\ H^{(i)}(\underline{a}, n)\ \omega(n).$$

Nous voulons étudier le cas où nous avons un point fixe. Donc, "par hasard"

$$\omega(n) = \frac{1}{H(\underline{a}, n)}$$

$$\Rightarrow \qquad \frac{\partial^2 \chi_\omega}{\partial a_i} = \sum \frac{2(g(n) - H(\underline{a}, n))}{H(\underline{a}, n)} \; H^{(i)}(\underline{a}, n)$$

Notons que

$$\sum_0^\infty H^{(i)}(\underline{a}, n) = \sum \frac{\partial}{\partial a_i} \int e^{-x} \frac{x^n}{n!} \, d\mu(\underline{a}, n) = \frac{\partial}{\partial a_i} \int d\mu(\underline{a}, n)$$

et les équations de Kuhn et Tucker s'écrivent

$$2 \sum g(n) \frac{H^{(j)}(\underline{a}, n)}{H(\underline{a}, n)} + \sum_i \mu_i \frac{\partial g_i}{\partial a_j}(\underline{a}) + (\lambda - 2) \frac{\partial}{\partial a_i} \int d\mu(\underline{a}, n) = 0 \qquad j = 1, \ldots, n$$

$$\mu_i \; g_i(\underline{a}) = 0 \qquad i = 1, \ldots, \ell$$

$$\mu_i \geq 0 \qquad i = 1, \ldots, \ell$$

$$\int d\mu(\underline{a}, x) = 1$$

Si on pose $\mu'_i = \frac{\mu_i}{2}$ et $\lambda' = \frac{\lambda - 2}{2}$, on voit que ces équations sont les mêmes que celles de l'estimateur de maximum de vraisemblance.

Alors les points stationnaires pour les deux méthodes sont les mêmes. En particulier, chaque minimum local de $\ell(\underline{a})$ pour le maximum de vraisemblance sous contrainte correspond à une solution de χ^2 où on choisit pour norme $H(\underline{a}, n)$. Puisque le maximum de vraisemblance est unique, et si nous avons un estimateur χ^2 de point fixe, il faut simplement vérifier qu'il s'agit d'un maximum pour conclure que c'est le maximum de vraisemblance. Notons une deuxième fois que le maximum de vraisemblance est numériquement instable. On le voit ici parce qu'il s'agit d'une solution de χ^2 pour un choix particulier de norme.

Nous avons essayé de chercher expérimentalement un χ^2 point fixé en commençant avec une norme ν_1 arbitraire puis en itérant par $\nu_{n+1} = P(\mu_n)$ où μ_n est la solution du problème χ^2 avec la norme ν_n. Dans tous les cas que nous avons testés, il y avait effectivement une convergence, mais on n'offre pas de démonstration que cette dernière doit avoir lieu.

4.5. — NOTES ET COMMENTAIRES

NOTA 1. – Nous avons vu qu'un problème mal posé est tel que l'inverse A^{-1} existe mais n'est pas continu. Une méthode générale pour la solution des problèmes mal posés est donnée dans le lemme 4.1, où on réduit le domaine à un ensemble compact et dans ce cas l'inverse est continu. Cette méthode est valable uniquement si le domaine est "petit". Par exemple, dans le chapitre III nous avons travaillé dans un $\mathcal{M}_K$ compact, mais l'inverse n'était pas suffisamment continu pour obtenir de bons résultats. On peut essayer de réduire le nombre de paramètres si l'on considère l'espace $\mathcal{M}_S = \{\mu \; ; \; \mu = \Sigma \lambda_i g_i\}$ où les g_i sont des mesures plus lisses que les $\{\delta x_i\}$ utilisées au chapitre III. En effet, ceci est la méthode proposée dans [6] où on considère les $\{g_i\}$ comme une base de B–spline, voir [7]. Puisque n'importe quel spline cubique $\tilde{g}$ s'écrit $\tilde{g} = \Sigma \lambda_i g_i$ pour $\{g_i\}$ la base de B–spline, on obtient toujours des estimateurs lisses pour l'histogramme sous–jacent. Néanmoins, si le nombre de nœuds reste élevé, il se trouve que des instabilités subsistent dans cette méthode, et la solution proposée dans [6] est de choisir le nombre de nœuds minimal tel que la distance entre l'observé g et $\hat{G} = P(\hat{\mu})$ devienne stable. La méthode reste peu satisfaisante. Elle peut être améliorée si on l'applique avec une régularisation. Un inconvénient de cette dernière est que l'équivalence avec le krigeage perd son interprétation intuitive. La matrice de régularisation devient alors l'inverse d'une matrice de covariance entre les éléments de base de B–spline. Ceci nous empêche d'estimer la matrice de régularisation à partir des données, mais au moins cela n'est pas très important parce que de toute façon, pour le problème de l'histogramme, nous travaillons dans un voisinage unique, et la méthode d'inférence proposée dans le chapitre III ne s'applique pas. Un autre inconvénient apparaît lorsque l'histogramme a un pic important. Un mauvais choix des nœuds peut nous donner un surlissage considérable. Puisqu'on ne sait pas dans la pratique où il faut mettre des nœuds, ceci peut être un problème important. Le point d'intérêt principal de cette méthode est qu'elle permet de réduire le nombre de paramètres à estimer par un facteur de 4 ou 5 et par conséquent le temps de calcul s'en trouvera réduit de manière importante.

NOTA 2. – Nous avons discuté en détail du problème de poissonisation. Mais il est clair que la même analyse s'applique aux autre problèmes de randomisation

pour lesquels l'application inverse existe. Par exemple si Y suit une loi discrète et Z suit une binomiale $Z \sim B(Y,p)$ pour p fixé, alors

$$\mathrm{E}[s^Z] = \mathrm{E}[(q + ps)^Y] = G_Y(q + ps)$$

et la transformée de Laplace est connue sur un intervalle non dégénéré dès que $q \neq 1$. On voit que pour ce problème, la décimation de Y, on peut appliquer le même type d'analyse pour estimer la loi de Y à partir d'un histogramme expérimental de Z.

Par contre si Y est une variable aléatoire qui prend des valeurs entre 0 et 1 presque sûrement et $Z \sim B(N, Y)$ où N est fixé, il est très facile de voir que $A : Y \rightarrow Z$ n'est pas injective et la même analyse ne s'applique pas.

NOTA 3. – Comme exemple d'un cas où on peut espérer résoudre un problème mal posé sans utiliser une méthode de régularisation, on peut considérer l'estimation de la covariance Ω^{-1} pour une régularisation. Comme on l'a vu au chapitre III, ce problème est lui-même mal posé, mais puisqu'il dépend uniquement d'un faible nombre de paramètres, on peut espérer que l'estimation est acceptable. En effet des résultats obtenus dans l'application du krigeage à la déconvolution confirment cette hypothèse pour les cas déjà essayés [8].

CHAPITRE V

LA RÉGULARISATION

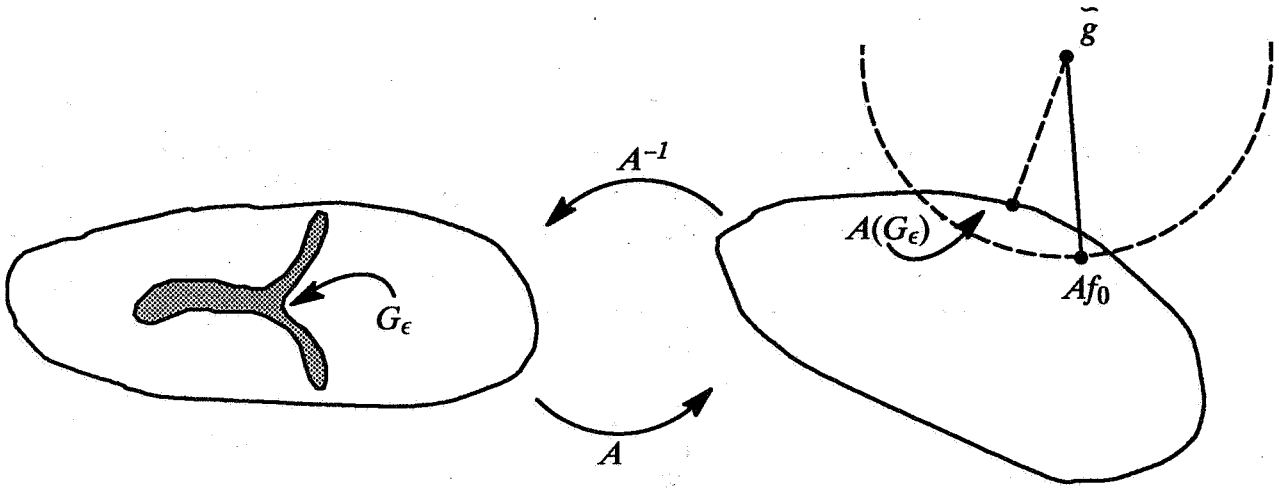
5.0. — *INTRODUCTION*

La méthode de régularisation a été introduite au chapitre III et a été appliquée à la construction de solutions stables du problème de poissonisation. On trouve une description complète de la méthode dans Tikhonov [1]. On ne donnera ici qu'un résumé très sommaire de sa démarche : ensuite seuls les problèmes linéaires dans un espace de Hilbert seront considérés. On démontrera l'équivalence de la méthode de régularisation avec celle du krigeage dans le cas du problème linéaire libre (sans contraintes de positivité). En toute rigueur donc l'équivalence ne s'applique pas dans le cas d'une poissonisation où on doit prendre en compte des conditions de positivité mais elle peut toujours être utilisée pour interpréter le sens de la norme de régularisation comme nous l'avons fait au chapitre III.

5.1. — *RÉGULARISATION DANS UN ESPACE MÉTRIQUE*

Soit F et G des espaces métriques et $A : F \to G$ un opérateur bijectif et continu sur $A(F)$ tel que A^{-1} n'est pas continu. A est donc un opérateur mal posé. Soit Ω une fonctionnelle positive et continue sur F telle que $\forall\ d \in \mathbb{R}$, l'ensemble $C_d = \{g \in F\ ;\ \Omega(g) \leq d\}$ est compact. Si $f_0 \in F$ est la "fonction vraie" à estimer et si $g = A\, f_0$ est la "fonction vraie déformée", notre but sera d'estimer f_0 à partir d'une perturbation $\tilde{g}$ de g. Supposons que $d(g, \tilde{g}) \leq \epsilon$ pour une valeur de ϵ, et appelons $G_\epsilon = \{f \in F\ ; d(Af, \tilde{g}) \leq \epsilon\}$, on peut démontrer que puisque C_d est compact pour chaque d, il existe au moins un point $\tilde{f}$ qui minimise $\Omega(f)$ dans G_ϵ . Une $\tilde{f}$ avec cette propriété s'appelle une solution régularisée et l'on peut démontrer que si $\epsilon \to 0$, c'est-à-dire quand $d(A\tilde{f}, \tilde{g}) \to 0$, alors $d(f_0, \tilde{f}) \to 0$ pour toute solution régularisée de $\tilde{f}$. On trouve alors une solution régularisée par

la minimisation d'une fonctionnelle continue Ω dans l'espace G_ϵ , G_ϵ étant définie telle que A_f est proche de $\tilde{g}$ pour chaque f dans G_ϵ .



Il se trouve que dans certaines conditions la minimisation se réduit à une minimisation sous contraintes linéaires : on dit que Ω est presque strictement convexe si pour tout point f tel que $\Omega(f) > \inf_{g \in F} \Omega(g)$, alors dans chaque voisinage ouvert V contenant f, il existe g tel que $\Omega(g) < \Omega(f)$. Par exemple une forme quadratique positive Ω dans un espace de Hilbert tel que $\Omega \neq 0$ est presque strictement convexe. Le théorème suivant est démontré dans [1].

THÉORÈME 5.1. – Si Ω est une fonctionnelle positive continue presque strictement convexe sur F alors un point $f \in G_\epsilon$ qui minimise $\Omega(f)$ sera tel que $d(Af, \tilde{g}) = \epsilon$ et le problème de régularisation se réduit à une minimisation sous contraintes linéaires.

La méthode de Lagrange pour la minimisation consiste donc à minimiser

$$M^\alpha = d^2(Af, \tilde{g}) + \alpha\, \Omega(f)$$

sujet à la condition $d(Af, \tilde{g}) = \epsilon$. Si la méthode de Lagrange est réalisable et qu'on trouve un $\alpha = \alpha(\epsilon)$ tel que la condition soit réalisée, alors la régularisation se réduit à la minimisation de M^α, sans aucune contrainte, avec ce choix de α. La régularisation est maintenant de la forme utilisée au chapitre III si on prend pour $d^2(Af, \tilde{g})$ la distance χ^2 et pour Ω une forme quadratique positive. Notons

que, bien que nous n'ayons considéré que le problème sans contraintes, il n'y a aucune difficulté théorique à en ajouter, sauf bien entendu au niveau du calcul.

A partir de maintenant, nous nous limitons aux problèmes mal posés linéaires dans un espace de Hilbert avec Ω fonctionnelle quadratique positive et continue. Sauf indication du contraire, la seule contraite imposée sur la minimisation sera que le minimum soit dans G_ϵ . Nous donnons les conditions sur Ω et A pour que la méthode de Lagrange ait une solution unique, c'est-à-dire pour qu'il existe une solution régularisée unique au problème. Notons qu'on n'impose plus la condition que les ensembles C_d définis ci-dessus soient compacts. Ceci était nécessaire pour démontrer l'existence d'une solution régularisée dans un espace métrique, mais dans le cas hilbertien, cette existence sera une conséquence des théorèmes de projection. Dans les conditions énoncées, la théorie de la régularisation n'est rien d'autre que la théorie des splines et on peut chercher des théorèmes d'équivalence avec le krigeage.

5.2. — *DÉFINITIONS ET THÉORÈMES*

Dans cette section, on énonce quelques définitions et théorèmes qui sont classiques en théorie élémentaire des espaces hilbertiens. Néanmoins il ne s'agit pas d'une liste complète et les théorèmes les plus connus tels que les théorèmes de projection, de Rietz, etc. sont admis ici sans énoncé. Nous admettons que les espaces sont des espaces réels.

Soit F et G deux espaces de Hilbert et $A : F \to G$ une application linéaire. A est continue s'il existe M tel que $\|Af\|_G \leq M \|f\|_F \quad \forall f \in F$. A est coercive s'il existe $c > 0$ tel que $\|Af\|_G \geq c \|f\|_F \quad \forall f \in F$. A partir de maintenant, à moins d'un risque de confusion, on n'utilisera plus les indices sur les normes hilbertiennes. Le lemme suivant, démontré dans [2], est très utile pour les calculs.

LEMME 5.1. — Si X et Y sont des sous-espaces fermés d'une espace de Hilbert F alors les propriétés suivantes sont équivalentes

1) $X + Y$ est fermé
2) $X^\perp + Y^\perp$ est fermé
3) $X + Y = (X^\perp \cap Y^\perp)^\perp$
4) $X^\perp + Y^\perp = (X \cap Y)^\perp$.

Le lemme suivant est souvent utilisé. Sa démonstration n'est pas difficile et sera omise.

LEMME 5.2. — Si X et Y fermés dans F et $X \cap Y = \{0\}$, alors $(X \oplus Y^{\perp}) \cap Y = \pi_Y X$.

Passons maintenant aux définitions de transposée et d'adjointe d'une application continue. Soit $A : F \to G$ une application linéaire entre des espaces de Hilbert F et G. Soit F' et G' des espaces duals de F et G. La transposée A' de A est définie de la manière suivante.

La transposée est l'application unique $A' : G' \to F'$ définie telle que $< A'g, f > = < g, Af > \quad \forall f \in F, g \in G'$. L'adjoint se définit à l'aide du théorème de Rietz. Soit h le représentant unique de g dans G, donc $< g, Af > = (h, Af)$. De même, $A'g$ a un représentant unique dans F, qu'on appelle A^*h. Nous avons

$$(A^*h, f) = (h, Af) \qquad \forall h \in G, f \in F$$

A^* est l'adjoint de A. On peut démontrer facilement que A' et A^* sont des applications continues.

Dans les espaces de dimension finie, A et A' ont des représentations matricielles. Il se trouve que $A' = A^t$ où A^t est la transposée de la matrice A. Par contre $A^* = A^t$ seulement quand nous avons une base orthogonale pour F et G. Ces résultats sont faciles à démontrer et nous mènent au lemme suivant.

LEMME 5.3. —

1) Soit F et G des espaces de Hilbert de dimension finie avec des espaces duals F' et G'. Si $\{f_1,\ldots,f_n\}$ est une base pour F, $\{f^1,\ldots,f^n\}$ la base duale de F' et si $\Omega_{ij} = (f_i, f_j)$ est la matrice du produit scalaire, alors $K^{ij} \equiv (f^i, f^j) = \left[\Omega^{-1}\right]_{ij}$.

2) Si $A : F \to G$ est continue et linéaire, si $x, y \in G$ sont des représentations de $\varphi, \eta \in G'$, et si D_{ij} est la matrice du produit scalaire de G alors

$$x^t(A^*)^t \, \Omega \, A^*y = (A^*x, A^*y) = (A'\varphi, A'\eta) = \varphi^t A \, K \, A^t\eta$$

où $A^* = \Omega^{-1}A^t \, D$ est l'adjoint de A.

DÉMONSTRATION. – Il s'agit d'un lemme d'algèbre linéaire élementaire, et on ne donnera que la démonstration de la première partie pour rappeler la notation utilisée dans le cas fini.

Supposons $f \in F$, alors $f = \Sigma\, a^i f_i$. On sait que f a une représentation unique dans F', que nous appelons également f, $f = \Sigma\, a_i f^i$. Le lien entre les composantes covariantes et contravariantes est le suivant

$$a_i = < f,\ f_i > = (f,\ f_i) = a^i\ \Omega_{ji}$$

et de la même manière

$$a^i = K^{ij}\ a_j$$

Maintenant si f et g sont deux vecteurs

$$(f,g)_{F'} = a_i\ K^{ij}\ \beta_j$$

et

$$(f,g)_F = a^i\ \Omega_{ij}\ \beta^j = a_k\ K^{ki}\ \Omega_{ij}\ K^{j\ell}\ \beta_\ell$$

Mais $(f,g)_{F'} = (f,g)_F \quad \forall\ f,g$

$$\Rightarrow \qquad K^{ij} = K^{ki}\ \Omega_{ij}\ K^{j\ell}$$

La deuxième partie vient après des calculs élémentaires.

Enonçons maintenant les théorèmes classiques des espaces de Hilbert utilisés par la suite.

THÉORÈME DE L'INVERSE CONTINU. – Soit F et G des espaces de Hilbert et A une application linéaire continue et bijective de F sur G. Alors A^{-1} est continue.

Le point crucial ici est que Im A est fermée. Les problèmes mal posés sont ceux pour lesquels Im A n'est pas fermée. Le théorème de l'inverse continu est vrai pour les espaces de Banach et c'est une conséquence simple du théorème de l'application ouverte [2].

Une forme bilinéaire $\Omega\,(u,\ v)$ sur $F \times F$ est continue s'il existe M tel que $|\Omega(u,v)| \leq M \| u \| \| v \|$ et est coercive s'il existe $c > 0$ tel que $|\Omega(u,u)| \geq c \| u \|^2$. Le théorème suivant fait le lien entre les formes bilinéaires

continue et coercive et les opérateurs linéaires continu et coercif. Le théorème est donné sous deux versions équivalentes dont la première est démontrée en [3]. La deuxième version ainsi que le théorème suivant sont démontrés dans [2].

THÉORÈME (LAX-MILGRAM). –

a) Si $\Omega\,(x,y)$ est une forme bilinéaire symétrique continue et coercive dans un espace de Hilbert F, alors il existe un opérateur continu, coercif, bijectif et autoadjoint unique $S : F \to F$ tel que

$$\Omega(x,y) = (x, Sy)$$

b) Pour chaque $\varphi \in F'$, il existe un $f \in F$ unique tel que $\Omega(f,g) = < \varphi, g >$ $\forall\ g \in F$ et f est caractérisé par la propriété

$$f \in H \qquad \text{et} \qquad \tfrac{1}{2}\,\Omega(f,f) - < \varphi, f > = \operatorname*{Min}_{g \in F} \{\tfrac{1}{2}\,\Omega(g,g) - < \varphi, g >\}$$

Notons que puisque la formulation a) donne S continue, coercive et bijective, on trouve que S^{-1} est continue et coercive. On peut généraliser le théorème de Lax–Milgram aux ensembles convexes.

THÉORÈME (STAMPACHIA). – Soit $\Omega\,(x,y)$ une forme bilinéaire symétrique continue et coercive dans un espace de Hilbert F. Soit K un ensemble convexe fermé non vide. Pour chaque $\varphi \in F'$, il existe un f unique dans K tel que

$$\Omega(f,\ g-f) \geq < \varphi,\ g-f > \qquad \forall\ g \in K$$

De plus f est caractérisé par la propriété

$$\tfrac{1}{2}\,\Omega(f,f) - < \varphi, f > = \operatorname*{Min}_{g \in K} \{\tfrac{1}{2}\,\Omega(g,g) - < \varphi, g >\}$$

Nous donnons maintenant une définition d'un espace de Hilbert autoreproduisant sur lequel est basée la théorie des splines [4], [5], puis nous introduirons l'idée d'une norme équivalente générée par une forme quadratique positive.

Soit X un espace métrique séparable (en pratique on prend $X = \mathbb{R}^n$ assez souvent) et F un espace de Hilbert de fonctions sur X. Il s'agit d'un espace de Hilbert autoreproduisant s'il existe une fonction $K(x,y) : E \times E \to \mathbb{R}$ tel que $K(\cdot,y) \in F$ et si pour chaque $f \in F$, $f(y) = (f,\ K(\cdot,y))$. On peut démontrer que

si $f_n \to f$ dans F, alors $f_n(x) \to f(x)$ pour chaque $x \in X$. De plus, on peut démontrer que si F est un espace de fonctions continues alors les vecteurs $K(\cdot, y)$ sont un ensemble total pour F quand y varie sur un ensemble dense de X. En particulier, puisque X est séparable, F est un espace de Hilbert séparable. Comme condition réciproque, si F est un espace hilbertien sur X, alors une condition suffisante pour qu'il y ait un noyau reproduisant est que $\forall\ x \in X$ la forme linéaire L_x définie par $L_x = f(x)$ est continue.

Si F est un espace de Hilbert et Ω une forme bilinéaire continue et coercive, alors Ω nous donne une nouvelle norme $\|\cdot\|_1$ sur F équivalente à l'ancienne norme $\|\cdot\|$ dans le sens qu'il existe c et M tels que $\|f\| \leq c\,\|f\|_1$ et que $\|f\|_1 \leq M\,\|f\|$. Plus généralement, il est facile de démontrer que si Ω est continue, $\ker\Omega = N$ et Ω est coercive sur $N^{\perp}$, alors la nouvelle norme sur F définie par

$$\|f\|_1^2 = \Omega(f,f) + \|\pi_N f\|^2$$

est équivalente à l'ancienne norme.

Cette situation se produit dans la théorie des splines où on a un opérateur continu D entre deux espaces de Hilbert $D : H_1 \to H_2$ qui est surjective. (Par exemple D est un opérateur associé à une équation différentielle [6]). Supposons que $\ker D = N$ alors $D : N^{\perp} \to H_2$ est une bijection continue et le théorème de l'inverse continu donne que $D|_{N^{\perp}}$ est coercive. Un autre moyen d'arriver à cette situation consiste à avoir une fonction aléatoire $Z(x)$, $x \in X$ qui génère un espace de Hilbert $H \subset L^2$ avec la fonction de covariance $C(x,y)$. Maintenant à chaque $Y \in H$ on associe la fonction $f \in H'$ telle que

$$f(x) = < f,\ Z(x) > = (Y,\ Z(x))$$

L'espace des fonctions H' ainsi généré est isomorphe à H et à un noyau reproduisant $C(x,y)$. Finalement on considère l'espace $F = H \oplus N$ pour un autre espace de Hilbert quelconque N et nous avons la situation décrite plus haut.

5.3. — RÉGULARISATION ET KRIGEAGE

5.3.1. — La régularisation dans un espace de Hilbert

Le problème général de régularisation de § 5.1 était de trouver f tel que $M^a = d^2(Af,\ g) + a\ \Omega(f)$ est un minimum.

Dans un espace de Hilbert, on prend Ω forme bilinéaire continue et positive et $d^2(Af, g)$ est une métrique hilbertienne définie sur l'espace engendré par les données. Nous allons ré-écrire le problème à l'aide des résultats du § 5.2.

Soit $A : F \to G$ un opérateur linéaire et continu, où F et G sont des espaces de Hilbert autoreproduisants sur X et X' respectivement. Nous admettons que les fonctionnelles d'évaluation $\tilde{H}_x$, $\tilde{H}_x(f) = f(x)$ $\forall$ $f \in F$, forment une partie totale pour F', et donc leurs représentations dans F, qu'on appelle $\tilde{H}_x$ également, forment une partie totale pour F. De même pour les fonctionnelles d'évaluation L_x dans G'. Soit Ω une forme quadratique positive symétrique et continue sur F avec $\ker \Omega = N$ un espace de dimension finie tel que Ω est coercive sur $N^{\perp}$. Alors si la norme de F est $\| \cdot \|_1$, on peut définir une norme équivalente comme on l'a vu à la section précédente

$$\| f \|^2 = \Omega(f,f) + \| \pi_N f \|_1^2$$

Les fonctionnelles d'évaluation, qu'on appelle maintenant H_x, restent continues puisque l'espace reste un espace de Hilbert autoreproduisant. De plus A reste continue. Pour la nouvelle norme $\| f \|^2 = \Omega(f,f) + \| \pi_N f \|^2$. Mais on a toujours $\| f \|^2 = \| \pi_{N^{\perp}} f \|^2 + \| \pi_N f \|^2$ et par identification $\| \pi_{N^{\perp}} f \|^2 = \Omega(f,f)$.

A partir de maintenant, on ne considérera que la nouvelle norme sur F parce que le problème de régularisation reste inchangé.

Admettons que $\tilde{S} \subset X'$ soit l'ensemble des points où l'on observe des données. Ceci peut être éventuellement défini comme dans le problème de poissonisation où l'espace $\tilde{S}$ comprend tous les entiers positifs ou nuls. Soit $S \subset G$ l'espace engendré par $\{L_{x_a}\}$, $x_a \in \tilde{S}$. Pour trouver une représentation de notre forme quadratique continue et coercive $d^2(Af, g)$ on suppose qu'on a une norme définie sur S. Par le théorème de Lax-Milgram, on peut trouver un opérateur $T : S \to S$ qui soit continu, coercif, autoadjoint et bijectif sur S tel que la norme sur S puisse s'écrire en termes de la norme de G. En particulier on peut représenter $d^2(Af, g)$ par $(g - Af, T\pi_S(g - Af))$ avec le produit scalaire de G. En effet, pour éviter des calculs inutiles, il est plus simple de changer la norme de G comme nous l'avons fait pour F, et de ne pas utiliser l'opérateur T. On introduit le nouveau produit scalaire sur G

$$(f,g)_1 = (f, T\pi_S\, g) + (\pi_{S^\perp} f, \pi_{S^\perp} g)$$

Ceci est une norme équivalente pour G qui respecte la norme imposée sur les données par T, en particulier A reste continue pour la nouvelle norme. Par conséquent le problème de régularisation reste invariant si l'on représente $d^2(Af,\ g)$ par $(g - Af,\ \pi_S(g - Af))_1$. Puisque nous allons utiliser uniquement cette norme sur G, il ne sera plus nécessaire d'écrire l'indice.

Le problème de régularisation prend la forme

$$M^a = \| \pi_{N^\perp} f \|^2 + ((g - Af), \pi_S(g - Af)) \qquad \textbf{(5.1)}$$

Nous n'utiliserons pas d'indices pour distinguer les normes utilisées (le premier terme se rapporte à F et le second à G) sauf s'il y a risque de confusion.

Notons que dans la régularisation (5.1) on ne fait pas appel à la norme sur F mais seulement sur $N^\perp \subset F$. En particulier, un autre choix de norme sur N tel que la nouvelle norme sur F reste équivalente à l'ancienne ne change ni la continuité de A ni $\| \pi_{N^\perp} f \|$, et par conséquent on trouve la même solution de (5.1). On peut mettre en évidence cette redondance plus facilement de la manière suivante. Supposons $f^1,\ldots,f^n$ une base pour l'espace N

$$(H_x,\ H_y) = (\pi_{N^\perp}\ H_x,\ \pi_{N^\perp}\ H_y) + (\pi_N\ H_x,\ \pi_N\ H_y)$$

mais $\pi_N\ H_x \in N \Rightarrow \exists\ C_\ell \in \mathbb{R}$ tel que $\pi_N\ H_x = \Sigma\ C_\ell\ f^\ell$ avec $C_m\ f^m = D_{\ell m}\ (H_x,\ f^\ell) f^m = D_{\ell m}\ f^\ell_x\ f^m$ où $D_{\ell m} = \left[(f^\ell, f^m)\right]^{-1}$ et $f^\ell_x = (H_x, f^\ell)$.

On a donc

$$(\pi_N\ H_x,\ \pi_N\ H_y) = D_{\ell m}\ f^\ell_x\ D_{\ell' m'}\ f^{\ell'}_y\ (f^m,\ f^{m'}) = D_{\ell\ell'}\ f^\ell_x\ f^\ell_y$$

et

$$(H_x, H_y) = (\pi_{N^\perp}\ H_x,\ \pi_{N^\perp}\ H_y) + D_{\ell\ell'}\ f^\ell_x\ f^\ell_y$$

Par conséquent, si l'on remplace $D_{\ell\ell'}$ par une autre matrice strictement positive, on obtient une norme équivalente sur F et le problème de régularisation n'est pas altéré.

5.3.2. — Le krigeage dans un espace de Hilbert

On garde la même notation qu'à la section précédente. On considère maintenant un problème de krigeage dans les espaces duals. Soit A' la transposée

de A. Alors A' : $G' \to F'$. On peut dire de manière approximative qu'un problème de krigeage consiste à estimer $H_x \in F'$ comme une somme pondérée "d'observations" dans $A'(S')$ où S' est le sous-espace de G' correspondant à S. Pour le problème de krigeage sans filtrage, il faut simplement trouver la projection de H_x sur $A'(S')$ parallèle à $N^{\perp}$. Pour ceci, il faut évidemment que $A'(S')$ soit fermé.

Pour voir que cette formulation correspond bien à l'idée intuitive, on peut considérer la situation suivante. Soit $Z = AY$ où Z,Y sont des fonctions aléatoires sur des supports $\{1,...,n\}$ et $\{1,...,r\}$ respectivement. Soit $\{ f^{\ell} \}$ une famille finie de fonctions sur $\{1,...,r\}$ et soit la fonction de covariance strictement conditionnellement définie positive pour des combinaisons linéaires autorisées, c'est-à-dire :

$$\mathrm{Var}(\sum \lambda^i Y_i) \geq 0$$

$$\text{si} \quad \sum \lambda^i f_i^{\ell} = 0$$

$$\text{avec} \quad \mathrm{Var}(\sum \lambda^i Y_i) = 0 \Leftrightarrow \lambda^i = 0 \quad \forall i$$

On veut estimer Y_x pour $x \in \{1,\ldots,r\}$ à l'aide des données $\{Z_k\}_{k \in \bar{S}}$ avec $\bar{S} \subset \{1,\ldots,n\}$. L'estimateur de krigeage est $Y_x^* = \sum_{k \in \bar{S}} \lambda^k Z_k$ avec les $\{\lambda^k\}$ choisies telles que $\mathrm{Var}(Y_x - \sum \lambda^k Z_k) = \mathrm{Var}(Y_x - \sum \lambda^k A_k^i Y_i)$ est un minimum sous les contraintes d'admissibilité $\sum \lambda^k A_i^k f_i^{\ell} = f_x^{\ell} \quad \forall \ell$.

Puisque la fonction de covariance admet des prolongements définis positifs sur $\{1,...,r\}$ (qui sont identiques à la covariance initiale sur les combinaisons linéaires autorisées) alors on peut en choisir un parmi eux et ensuite interpréter $\{Y_i\}$ comme des vecteurs de base d'un espace de Hilbert de norme $\| X \| = \mathrm{Var}(X)$. Donc le problème de krigeage devient : Trouver $Y_x^* = \sum \lambda^k Z^k$ tel que $\| Y_x - \sum \lambda^k A_k^i f_i^{\ell} \|$ est un minimum sujet aux contraintes d'admissibilité $\sum \lambda^k A_i^k f_i^{\ell} = f_x^{\ell} \quad \forall \ell$.

Maintenant considérons le formulation du problème de krigeage dans l'optique de la notation développée pour la régularisation. Soit F,G des espaces de Hilbert

avec des bases duales $\{H_i\}_{i=1}^r$ et $\{L_i\}_{i=1}^n$ respectivement. Soit $A : F \to G$ notre application linéaire et continue. Donc $A' : G' \to F'$, et en particulier pour $L_x \in G'$, $A'L_x \in F'$ admet la représentation $A'L_x = A_n^i H_i$. Le projection de H_x sur $A'(S)$ parallèle à $N^\perp$ est de la forme $H_x^* = \lambda^n A'L_n$ avec $\{\lambda^n\}$ les pondérateurs qui minimisent $\| H_x - \lambda^n A'L_n \| = \| H_x - \lambda^n A_n^i H_i \|$ sujet aux contraintes $< H_x - H_x^*, Y > = 0 \quad \forall Y \in N$.

Si maintenant l'on identifie l'espace F' avec l'espace engendré par les variables aléatoires $\{Y_i\}$ par $H_i \leftrightarrow Y_i$ et une identification des normes, et si de plus l'on interprète les fonctions $\{f^\ell\}$ comme base de l'espace $N \subset F$ (donc $< H_x - H_x^*, f^\ell > = 0 \quad \forall \ell \Leftrightarrow \lambda^k A_k^i f_i^\ell = f_x^\ell \quad \forall \ell$), alors on voit que notre krigeage correspond bien à un projection de H_x sur $A'(S')$ parallèle à $N^\perp$.

De la même manière, le problème de cokrigeage que l'on peut écrire formellement :

$$Z = AY + \epsilon \; ; \qquad \text{Estimer } Y_x \text{ à partir de } \{Z_a\}_{a \in S}$$

et dont nous démontrons l'équivalence avec le problème de régularisation dans le théorème 5.3, s'écrit comme la minimisation de

$$\| H_x - H_x^* \|^2 + \| L_x^* \|^2 \quad \text{où} \quad H_x^* = A' \, L_x^* \in A'(S')$$

On note que la norme sur l'espace G' (ou au moins sur S', la seule partie de G' utilisée) n'est rien d'autre que la fonction de covariance de l'erreur ϵ.

Comme le problème de régularisation a lieu dans les espaces F et G, il est plus simple de traiter les problèmes de krigeage également dans F et G. Donc, sans changer le problème, on peut considérer le krigeage comme la projection de $H_x \in F$ sur $A^*(S)$ au lieu d'une projection sur $A'(S')$ dans l'espace F'. Avec cette convention, on a les deux problèmes dans le même espace et on essaye de trouver les équations caractéristiques des solutions et ensuite de démontrer qu'elles sont identiques pour les deux problèmes.

5.4. — THÉORÈMES D'ÉQUIVALENCE DE LA RÉGULARISATION ET DU KRIGAGE

Le premier théorème ne s'applique pas aux problèmes de régularisation. Il s'agit en fait d'une équivalence entre krigeage et splines. Le résultat a été démontré à une dimension par G. Kimmeldorf et G. Wahba (1971) pour quelques cas particuliers [7]. Le résultat a été généralisé par J. Duchon (1977) [8]. G. Matheron (1980) a donné une équivalence générale dans [9]. Ici la présentation, qui suit les lignes générales de [9], prend en compte explicitement l'opérateur de distorsion A et le fait qu'on peut avoir un nombre infini de données. La section se termine avec un théorème d'existence et d'unicité des splines d'ajustement dans un convexe fermé.

THÉORÈME 5.2. — Quand N est de dimension finie, $N \cap \ker(\pi_S A) = \{0\}$ et $A^* |_S$ est un isomorphisme, alors les deux problèmes suivants ont une solution unique et identique.

1) $\underset{f \in F}{\text{Minimiser}} \| \pi_{N^\perp} f \|$ sous les contraintes $Af(x_a) = g_a$

2) Trouver $H_x^* \in I_m(A^* \pi_S)$ tel que $\| H_x - H_x^* \|$ est un minimum et tel que $(H_x - H_x^*, Y) = 0 \quad \forall \ Y \in N$.

Avant de démontrer ce théorème, on considère la signification des contraintes, $N \cap \ker(\pi_S A) = \{0\}$ et $A^* |_S$ est un isomorphisme. La première condition $N \cap \ker(\pi_S A) = \{0\}$ admet la représentation algébrique suivante. Puisque $\{f^\ell\}$ est une base pour N, pour chaque $C_\ell f^\ell \in N$, on a $C_\ell f^\ell \in \ker \pi_S A$ si et seulement si

$$(L_a, A(C_\ell f^\ell)) = 0 \qquad \forall \ x_a \in \bar{S}$$

$$\Leftrightarrow \quad C_\ell (A f^\ell)(x_a) = 0 \qquad \forall \ x_a \in \bar{S}$$

Donc $N \cap \ker(\pi_S A) = \{0\}$ implique la condition suivante

$$C_\ell (A f^\ell)(x_a) = 0 \quad \forall \ x_a \in \bar{S} \quad \Rightarrow \quad C_\ell = 0 \quad \forall \ \ell$$

et dans le cas où il n'y a qu'un nombre fini de points dans X et X', alors

$$C_\ell A_a^i f_i^\ell = 0 \quad \forall\ a \quad \Rightarrow \quad C_\ell = 0 \quad \forall\ \ell$$

La deuxième condition impose que $A^*|_S$ soit un isomorphisme donc $S \cap \ker A^* = \{0\}$ et Im $A^*\pi_S$ est fermé. Mais la condition $S \cap \ker A^* = \{0\}$ est une condition très naturelle. En effet elle est satisfaite si, pour chaque point $x_a \in \bar{S}$, nous avons $Af(x_a) \neq 0$ pour au moins une $f \in F$, parce que dans ce cas

$$(L_{x_a},\ Af) \neq 0$$

$$\Rightarrow \qquad (A^*\ L_{x_a},\ f) \neq 0$$

et

$$L_a \notin \ker A^*$$

Si cette condition est satisfaite $\forall\ x_a \in \bar{S}$, alors S, qui est engendré par $\{L_{x_a}\}_{x_a \in \bar{S}}$, satisfait à $S \cap \ker A^* = \{0\}$. L'image de $A^*\pi_S$ est fermée si S est de dimension finie. En effet, puisque A est continu, il est vrai que A^* est continu De plus l'image d'un espace complet de dimension finie est complète. Donc si S est de dimension finie et $S \cap \ker A^* = \{0\}$, alors $A^*|_S$ est un isomorphisme. Dans le cas général, il faut admettre que Im $(A^*\pi_S)$ est fermée. Par exemple, pour le problème de poissonisation, ceci est le cas si nous choisissons pour l'espace F un espace de Hilbert des mesures avec support dans le même compact $K \subset \mathbb{R}^+$ et $S = G = \ell^2(\mathbb{N})$ (voir chapitre IV).

DÉMONSTRATION DU THÉORÈME 1. – On démontre d'abord l'existence de la solution du problème spline. Puisque $A^*|_S$ est un isomorphisme, son dual $\pi_S A : A^*(S) \to S$ est aussi un isomorphisme et par conséquent, il existe une solution unique $h \in A^*(S)$ de l'équation

$$Ah(x_a) = g_a \tag{5.2}$$

En effet, il existe une solution unique de $\pi_S Ah(x_a) = g_a$ et donc

$$g_a = (L_a,\ g) = (L_a, \pi_S Ah) = (L_a,\ Ah)$$

et

$$Af(x_a) = g_a$$

Notons que pour chaque $f \in \ker \pi_S A$, nous avons $(L_a, Af) = 0$ et puisqu'il est vrai que $A^*(S)^\perp = \ker \pi_S A$, la solution générale de (5.2) est de la forme $h + \ker \pi_S A$. Donc toute solution particulière est de la forme

$$f = h + f' \quad \text{avec} \quad f' \in \ker \pi_S A$$

Le problème des splines est de trouver f tel que $\| \pi_{N^\perp} f \|$ soit un minimum.

Notons que $\ker \pi_S A + N = \ker \pi_S A \oplus N$ puisque leur intersection est nulle. De plus c'est un ensemble fermé parce que N est de dimension finie et $\ker \pi_S A$ est fermé (π_S et A étant continus). Par le lemme 5.2 on trouve que

$$(\ker \pi_S A \oplus N) \cap N^\perp = \pi_{N^\perp}(\ker \pi_S A)$$

et $\pi_{N^\perp}(\ker \pi_S A)$ est fermé parce que c'est l'intersection d'ensemble fermés. Alors l'ensemble des solutions projeté sur $N^\perp$, soit $\pi_{N^\perp} h + \pi_{N^\perp}(\ker \pi_S A)$ est fermé et il existe un point de norme minimale. De plus, puisque $\ker \pi_S A \cap N = \{0\}$, la restriction de $\pi_{N^\perp}$ à $\ker \pi_S A$ est bijective sur son image dans $N^\perp$ et on peut conclure qu'il existe une solution unique de l'équation (5.2) telle que sa projection sur $N^\perp$ est de norme minimale. Ceci est la solution unique du problème des splines.

Nous allons maintenant chercher un système d'équations qui caractérise la solution spline. Pour cela, il faut d'abord noter que $\pi_{N^\perp} f \in \text{Im}\,(A^* \pi_S)$ où f est la solution du problème des splines. En effet, on sait que $\pi_{N^\perp} f$ est orthogonale à $\pi_{N^\perp}(\ker \pi_S A)$ puisque $\pi_{N^\perp} f$ est le point de norme minimale dans $\pi_{N^\perp} h + \pi_{N^\perp}(\ker \pi_S A)$. Donc

$$(\pi_{N^\perp} f,\ \pi_{N^\perp} g) = 0 \quad \forall g \in \ker \pi_S A$$

$$\Rightarrow (\pi_{N^\perp} f,\ g) = 0 \quad \forall g \in \ker \pi_S A$$

$$\Rightarrow \pi_{N^\perp} f \in \text{Im}(A^* \pi_S)$$

Puisque $\{L_a\}$ forme un ensemble total pour S et puisque $A^* : S \to A^*(S)$ est un isomorphisme, il est vrai que $\{A^* L_a\}$ forme un ensemble total pour $A^*(S)$. Donc chaque $g \in A^*(S)$ s'écrit

$$g = \lim_n \sum \lambda_n^a A^* L_a$$

Nous profitons d'une notation introduite dans [10] pour écrire

$$g = \int \mathfrak{L}(dx)\, A^* L_x$$

Il s'agit d'une notation symbolique et en général $\mathfrak{L}(dx)$ n'admet pas d'interprétation en termes de mesure.

Pour la solution du problème du spline f, nous avons $\pi_{N^\perp} f \in A^*(S)$ et donc

$$\pi_{N^\perp} f = \int \mathfrak{L}(dx)\, A^* L_a$$

De plus, $\pi_{N^\perp} f \in N^\perp$ et donc

$$< \int \mathfrak{L}(dx)\, A^* L_a,\ f^\ell > = 0 \qquad \ell = 1,\ldots,n$$

Puisque $\lim_n \sum \lambda_n^a A^* L_a$ converge fortement vers $\int \mathfrak{L}(dx) A^* L_x$, par la convergence faible induite, nous avons

$$\lim_n \sum \lambda_n^a\, (A^* L_a,\ f) = 0$$

où, en utilisant la notation symbolique

$$\int \mathfrak{L}(dx)\, Af(x) = 0 \tag{5.4}$$

La solution du problème f prend la forme

$$f = \int \mathfrak{L}(dx)\, A^* L_x + C_\ell f^\ell \tag{5.5}$$

et est sujette à la condition (5.4) ainsi que

$$(Af,\ L_\beta) = g_\beta \tag{5.6}$$

ce qui est simplement $(f,\ A^* L_\beta) = g_\beta$, et si l'on injecte l'équation (5.5), on trouve en utilisant la convergence faible et la notation symbolique

$$\begin{aligned} \int \mathcal{L}(dx)\ (A^*L_\alpha,\ A^*L_x) + C_\ell\ Af^\ell(x_\beta) &= g_\beta \\ \int \mathcal{L}(dx)\ Af^\ell(x) &= 0 \end{aligned} \tag{5.7}$$

Ces équations sont caractéristiques de la solution du problème des splines. On peut le démontrer directement, mais on préfère ici démontrer qu'elles sont caractéristiques du problème de krigeage et donc qu'elles ont une solution unique, d'où la conclusion qu'elles sont également caractéristiques pour le problème des splines.

Pour le deuxième problème, on cherche le meilleur estimateur linéaire qui soit intrinsèque dans le sens que $(H_x - H_x^*,\ Y) = 0 \quad \forall\ y \in N$. Cette théorie a été développée dans [10], et on la reprend ici pour la terminologie développée jusqu'à présent.

On veut estimer $H_x^* \in \text{Im}\ (A^*\pi_S)$ tel que $\| H_x - H_x^* \|$ est un minimum et tel que $(H_x - H_x^*,\ Y) = 0 \quad \forall\ y \in N$. La deuxième condition implique que $H_x - H_x^* \in N^\perp$, et intuitivement on peut donc penser au krigeage comme à la projection sur $\text{Im}(A^*\pi_s)$ parallèle à $N^\perp$. Plus précisément

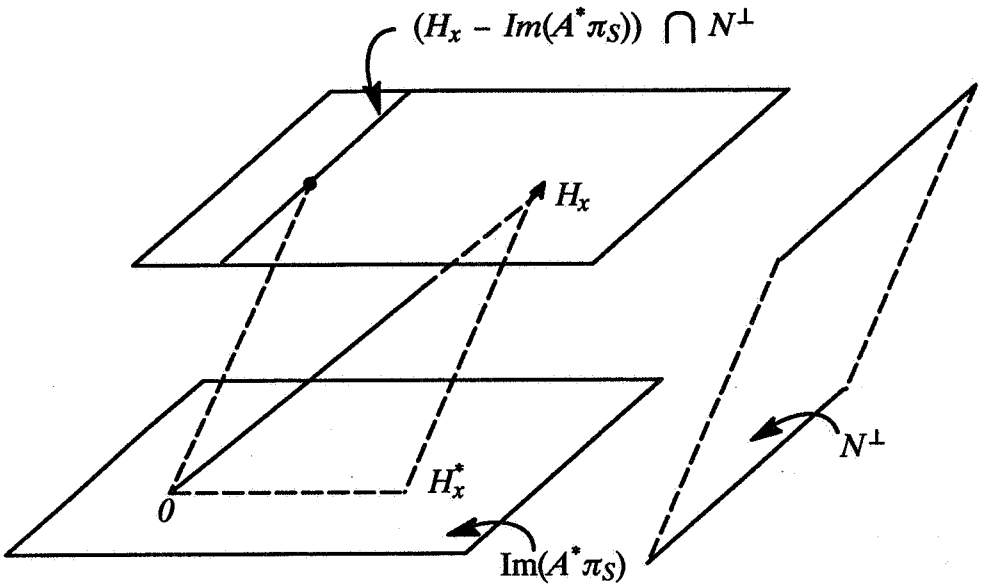
$$H_x^* \in \text{Im}\ (A^*\pi_S) \Rightarrow H_x - H_x^* \in H_x - \text{Im}\ (A^*\pi_S).$$

Mais $H_x - H_x^* \in N^\perp$ et donc

$$H_x - H_x^* \in \left[H_x - \text{Im}\ (A^*\pi_S)\right] \cap N^\perp$$

Par conséquent $H_x - H_x^*$ est la projection de 0 sur la variété

$$\left[H_x - \text{Im}\ (A^*\pi_S)\right] \cap N^\perp .$$



Pour voir qu'une telle projection existe et est unique, on remarque que la variété est fermée puisque $\mathrm{Im}\,(A^*\pi_S) \cap N^\perp$ est fermée. De plus elle est non vide. En effet supposons que $\mathrm{Im}\,(A^*\pi_S) \cap N^\perp \neq 0$, alors $\mathrm{Im}\,(A^*\pi_S) \subset N^\perp$ ou $N^\perp \subset \mathrm{Im}\,(A^*\pi_S)$ (les variétés $\mathrm{Im}\,(A^*\pi_S)$ et $N^\perp$ étant "parallèles"). Soit X l'ensemble majorant, $X = \mathrm{Im}(A^*\pi_S) + N^\perp$. Alors $X^\perp = \ker \pi_S A \cap N$. Il est non vide, donc en contradiction avec l'hypothèse, sauf si $X = F$. Donc, on a soit $\mathrm{Im}(A^*\pi_S) = F$, soit $N^\perp = F$. Dans ce cas, l'autre sous-espace doit être vide (sinon $[H_x - \mathrm{Im}\,(A^*\pi_S)] \cap N^\perp$ sera non vide). Il reste donc deux possibilités :

1) $N^\perp = F$ et $\mathrm{Im}\,(A^*\pi_S) = 0$. Il existe une solution unique $H_x^* = 0$

2) $N^\perp = 0$ et $\mathrm{Im}\,(A^*\pi_S) = F$. Il existe une solution unique $H_x^* = H_x$.

On voit donc qu'une solution unique existe qui est donnée par la projection de 0 sur $[H_x - \mathrm{Im}\,(A^*\pi_S)] \cap N^\perp$. On a donc

$$(H_x - H_x^*,\ Y) = 0 \qquad \forall\ Y \in H_x - \left[H_x - \mathrm{Im}(A^*\pi_S)\right] \cap N^\perp$$

Finalement H_x^* est caractérisé parce qu'il est le seul vecteur dans $\mathrm{Im}(A^*\pi_S)$ tel que $(H_x - H_x^*,\ Y) = 0 \quad \forall\ Y \in N$ qui de plus satisfait à

$$(H_x - H_x^*,\ Y) = 0 \qquad \forall\ Y \in \mathrm{Im}(A^*\pi_S) \cap N^\perp \qquad (5.8)$$

Maintenant appelons $V = \text{Im } (A^* \pi_S) \cap N^\perp$ et soit V' l'orthogonal de V dans Im $(A^* \pi_S)$

$$V' = \left[\text{Im } (A^* \pi_S) \cap N^\perp\right]^\perp \cap \text{Im } (A^* \pi_S)$$

Par le lemme 5.1 on trouve que

$$V' = \pi_{\text{Im } (A^* \pi_S)} N$$

Il est vrai que $\pi_{\text{Im } (A^* \pi_S)} : N \to V'$ est bijective parce que si $u \in N$, alors

$$u = \pi_{\text{Im } (A^* \pi_S)} u + \pi_{\ker (\pi_S A)} u$$

et si $\pi_{\text{Im } (A^* \pi_S)} u = 0$, nous trouvons $u = \pi_{\ker (\pi_S A)} u$

$$\Rightarrow \qquad u \in \ker \pi_S A \cap N \quad \text{et donc} \quad u = 0.$$

Prenons le produit scalaire de $H_x - H_x^*$ avec un vecteur quelconque $z \in \text{Im } (A^* \pi_S)$

$$z = \pi_V z + \pi_{V'} z$$

L'équation (5.8) implique que

$$\begin{aligned} (H_x - H_x^*,\ z) &= (H_x - H_x^*,\ \pi_{V'} z) \\ &= (\pi_{V'}(H_x - H_x^*),\ z) \end{aligned}$$

mais $\pi_{V'}(H_x^* - H_x) \in V' \quad \Rightarrow$ il existe $n \in N$ unique tel que

$$\pi_{\text{Im } (A^* \pi_S)} n = \pi_{V'}(H_x^* - H_x)$$

Alors

$$(n, z) = \left(\pi_{\text{Im } (A^* \pi_S)} n, z\right) = (\pi_{V'}(H_x - H_x^*),\ z)$$

et donc

$$(H_x - H_x^*,\ z) = (n, z) \qquad \forall\ z \in \text{Im}(A^* \pi_S) \tag{5.9}$$

et l'on a toujours

$$(H_x - H_x^*,\ y) = 0 \qquad \forall\ y \in N \tag{5.10}$$

Ces deux équations caractérisent $H_x^* \in \text{Im}\ (A^*\pi_S)$ qui est la projection de H_x $\text{Im}(A^*\pi_S)$ parallèle à $N^\perp$ parce qu'elles sont satisfaites si et seulement si

$$(H_x - H_x^*,\ y) = 0 \qquad \forall\ y \in \text{Im}\ (A^*\pi_S) \cap N^\perp$$

En effet, si (5.10) et (5.9) sont vérifiées, alors (5.10) implique que $H_x - H_x^* \in N^\perp$ et donc pour chaque $z \in \text{Im}\ (A^*\pi_S) \cap N^\perp$, l'équation (5.9) se réduit à $(H_x - H_x^*,\ z) = (n,z) = 0$.

Rappelons que les $\{A^*L_\alpha\}$ est un ensemble total pour $\text{Im}(A^*\pi_S)$ et puisque $n \in N$, il existe $C_\ell \in \mathbb{R}$ tel que $n = \Sigma\ C_\ell f^\ell$. Maintenant si on réintroduit la notation symbolique, les équations (5.9) et (5.10) prennent la forme suivante sur l'ensemble total $\{A^*L_\alpha\}$ (et ils restent donc caractéristiques de la solution)

$$\begin{aligned} \int \mathcal{L}_x'(dy)\ (A^*L_y,\ A^*L_\beta) + C_\ell'\ Af^\ell(x_\beta) &= (H_x,\ A^*L_\beta) \\ \int \mathcal{L}_x'(dy)\ Af^\ell(y) &= f^\ell(x) \end{aligned} \tag{5.11}$$

pour l'estimateur $H_x^* = \int \mathcal{L}_x'(dy)\ A^*L_y$.

Ces équations étant caractéristiques pour le krigeage, cela implique que les équations (5.7) sont caractéristiques pour le problème de régularisation. Pour démontrer qu'ils ont la même solution, on utilise l'argument classique pour démontrer l'équivalence entre krigeage et krigeage dual.

On peut interpréter H_x^* en prenant son produit scalaire avec une fonction. En particulier, pour chaque f tel que $Af(x_\alpha) = g_\alpha$, on a

$$h_x^* = (H_x^*,\ f) = \int \mathcal{L}_x'(dy)\ (A^*L_y,\ f) = \int \mathcal{L}_x'(dy)\ g_y$$

Donc la fonction h_x^* reste la même pour chaque f qui satisfait aux contraintes. Nous démontrons que h_x^* est la solution donnée par le problème spline

$$h_x^* = \int \mathcal{L}_x'(dy)\, g_y$$

Par l'équation (5.7)

$$h_x^* = \int \mathcal{L}_x'(dy)\left\{\int \mathcal{L}(dz)\ (A^*L_z\, A^*L_y) + C_\ell\, Af^\ell(y)\right\}$$

On peut renverser l'ordre des symboles parce que si $x_n \to x$ et $y_n \to y$ alors $(x_n,y_n) \to (x,y)$ dans un espace de Hilbert et le résultat suit si on applique ce résultat ici avec x_n une somme finie qui tend vers le symbole $\int \mathcal{L}_x'(dy)\, g_y$. Par l'équation (5.11), on trouve que

$$h_x^* = \int \mathcal{L}(dz)\left\{(H_x,\ A^*L_z) - C_\ell'\, Af^\ell(z)\right\} + C_\ell\, f_x^\ell$$

et encore par l'équation (5.7)

$$h_x^* = \int \mathcal{L}(dz)(H_x,\ A^*L_z) + C_\ell\, f_x^\ell$$

h_x^* est donc (f,H_x) où f est donné par l'équation (5.5), et le krigeage et les splines coïncident. ■

NOTA 1. – Nous avons démontré que pour chaque problème de spline, on peut trouver une interprétation en termes de krigeage : estimer Y_x à partir de $\{Z_a\}_{a\in S}$ quand $Z(x) = AY(x)$. Il faut tout simplement identifier $Y(x)$ avec H_x et utiliser comme covariance $C(x,y) = (H_x,H_y)$.

NOTA 2. – Réciproquement, si $Y(x)$ est une fonction aléatoire sur X et qu'on veuille kriger $Y(x)$ en présence des fonctions de dérive $\{f^\ell\}_{i=1}^n$, on peut construire une spline qui donne le même résultat. En effet supposons que $H \subset L^2(\Omega, \mathcal{A}, p)$ est l'espace de Hilbert engendré par $\{Y(x)\}$. Construisons l'espace $\tilde{F}$ des fonctions isomorphes à H par

$$\tilde{F} = \left\{\, g_y\ ; g_y(x) = (Y,\ Y(x)) \text{ quand } Y \text{ parcourt } H\right\}$$

$\bar{F}$ est un espace autoreproduisant, et en particulier il est généré par $\{g_{Y(x)},\ x \in X\}$. Soit N un espace de dimension n avec comme base $\{f^\ell\}_{i=1}^n$ et produit scalaire $(f^\ell, f^m) = B^{\ell m}$ une matrice définie positive mais arbitraire. Maintenant posons $F = \bar{F} \oplus N$. Cet espace est engendré par les vecteurs $g_{Y(x)} + C_\ell f^\ell$. F est l'espace sur lequel on construit la solution spline équivalente au krigeage. Pour le voir, supposons que $\{B_1,\ldots,B_n\}$ est la base duale à $\{f^1,\ldots,f^n\}$. Il est facile de voir que les vecteurs $Y(x) + f_x^m B_m \in H \oplus N'$ sont des fonctionnelles d'identification sur F. C'est-à-dire

$$< Y(x) + f_x^m B_m,\ g > = g(x) \qquad \forall\ g \in F$$

De plus, ces fonctionnelles engendrent $H \oplus N'$ qui est le dual de F. Il reste à considérer B_m comme une variable aléatoire dans $L^2(\Omega, \mathcal{A}, p)$ indépendante de $Y(x)$ $\forall x$ et avec Cov $(B_n, B_m) = D_{nm} = (B^{-1})^{nm}$ qui a été choisi de manière arbitraire. Il est clair que le krigeage de $Y(x)$ est le même que celui de $Y(x) + B_m f_x^\ell$ du fait que les fonctions de dérive $\{f^\ell\}$ sont filtrées. Par le théorème qui vient d'être démontré, le krigeage est donc équivalent au problème des splines dans F.

NOTA 3. – Quand les espaces sont de dimension finie, on peut appliquer le lemme 5.3 pour voir que les équations de (5.7) se réduisent à

$$\lambda^T A\, K\, A^t + C^T A\, f^\ell = g$$

$$\lambda^T A\, f^\ell \qquad\qquad = 0$$

où $K = \text{Cov}\,(Y_i, Y_j) = (H_i, H_j)$ et $f^\ell = (f^1,\ldots,f^n)^t$ et l'estimateur est $f^* = \lambda^t A^t L_a + b_\ell f^\ell$.

Sans restreindre la généralité, on peut admettre dans le théorème suivant que $\ker A^* \cap S = \{0\}$. Pour le voir, changeons legèrement la notation. Maintenant supposons que les données observées génèrent l'espace $\bar{S}$ et $\ker A^* \cap \bar{S} \neq \{0\}$.

Posons $S = \overline{\text{Im}\ \pi_{\bar{S}} A}$. Notons que $S \subset \bar{S}$. Soit S' l'orthogonal de S dans $\bar{S}$.

Notons que $\ker A^* \cap S = \{0\}$ et que les problèmes de régularisation et de krigeage restent inchangés si l'on travaille dans S au lieu de $\tilde{S}$. En effet, le problème de régularisation est de trouver le f qui minimise le terme suivant

$$\begin{aligned} & f \text{ tel que } f \text{ minimise } \{ \| \pi_{N^\perp} f \|^2 + (g - Af,\ \pi_{\tilde{S}}(g - Af)) \} \\ = & f \text{ tel que } f \text{ minimise } \{ \| \pi_{N^\perp} f \|^2 + (g - Af,\ \pi_S(g - Af)) + (g,\ \pi_{S'} g) \} \\ = & f \text{ tel que } f \text{ minimise } \{ \| \pi_{N^\perp} f \|^2 + (g - Af,\ \pi_S(g - Af)) \} \end{aligned}$$

Pour le problème de krigeage décrit dans le théorème, il est clair que $L_x^* \in S$ et $L_x^* \notin \tilde{S} \setminus S$ parce que $\pi_S L_x^*$ est toujours plus petit que L_x^* dans le sens de la forme quadratique à minimiser.

THÉORÈME 5.3. – Quand N est de dimension finie et $N \cap \ker \pi_S A = 0$ où $S = \overline{\text{Im } \pi_{\tilde{S}} A} \cap \tilde{S}$, alors les deux problèmes suivants ont des solutions uniques et identiques

1) $\underset{f}{\text{Min}}\ M = ((g - Af),\ \pi_{\tilde{S}}(g - Af)) + \| \pi_{N^\perp} f \|^2$

2) Trouver $L_x^* \in \tilde{S}$ et $H_x^* = A^* L_x^*$ tels que $M' = \| H_x - H_x^* \|^2 + \| L_x^* \|^2$ est un minimum et tel que $(H_x - H_x^*,\ Y) = 0 \ \ \forall\ Y \in N$.

DÉMONSTRATION. – On a déjà remarqué qu'on ne change pas les problèmes si l'on travaille exclusivement dans S. Les problèmes deviennent :

1) $\underset{f}{\text{Min}}\ M = ((g - Af),\ \pi_S(g - Af)) + \| \pi_{N^\perp} f \|^2$

2) Trouver $L_x^* \in S$ et $H_x^* = A^* L_x^*$ tels que $M' = \| H_x - H_x^* \|^2 + \| L_x^* \|^2$ est un minimum et tel que $(H_x - H_x^*,\ Y) = 0 \ \ \forall\ y \in N$.

De plus, on sait que $\ker A^* \cap S = \{0\}$.

Démontrons d'abord l'existence d'une solution unique pour le premier problème.

Notons que M s'écrit

$$M = (g,g) - 2(f,\ A^* g) + (\pi_S Af,\ \pi_S Af) + (\pi_{N^\perp} f,\ \pi_{N^\perp} f)$$

(g,g) est constant, et n'intervient pas dans la minimisation. M est donc de la forme $M = < \varphi,\ f > + B(f,f)$ si l'on prend pour φ le représentant de $A^* g$ et si

$$B(f,f) = (\pi_{N^\perp} f,\ \pi_{N^\perp} f)_F + (\pi_S Af,\ \pi_S Af)_G$$

Mais $B(f,f)$ est continue et sera coercive si $(\pi_S Af, \pi_S Af)_G$ est strictement positive sur N. Supposons que $f \in N$ et $(\pi_S Af, \pi_S Af)_G = 0$. Alors $\pi_S Af = 0$.

Mais $N \cap \ker \pi_S A = 0 \quad \Rightarrow \quad f = 0$.

Donc $B(f,f)$ est coercive et le théorème de Lax-Milgram s'applique. Le problème de régularisation a donc une solution unique f. De plus f est déterminée par l'équation $B(f,u) = < \varphi, u > \quad \forall\ u \in F$.

Ici donc

$$(\pi_S Af,\ \pi_S Au) + (\pi_{N^\perp} f,\ \pi_{N^\perp} u) = (A^* g,\ u) \qquad \forall\ u \in F$$

d'où

$$(A^* \pi_S Af + \pi_{N^\perp} f - A^* g,\ u) = 0 \qquad \forall\ u \in F$$

et l'on trouve que

$$\pi_{N^\perp} f + A^* \pi_S Af - A^* g = 0 \tag{5.12}$$

Mais le deuxième et le troisième terme sont dans $A^*(S)$. Donc

$$\pi_{N^\perp} f \in A^*(S)$$

et on peut écrire

$$\pi_{N^\perp} f = \int \mathfrak{L}(dx)\ A^*\ L_x$$

Par la continuité de A^* et la définition de de la notation symbolique, on trouve que

$$\pi_{N^\perp} f = A^* \int \mathfrak{L}(dx)\ L_x$$

et l'équation (5.12) devient

$$A^* \left(\int \mathfrak{L}(dx)\ L_x + \pi_S Af - g \right) = 0$$

Mais $\ker A^* \cap S = \{0\}$, et donc on trouve que

$$\int \mathfrak{L}(dx)\ L_x + \pi_S Af - g = 0 \tag{5.13}$$

Notre estimateur s'écrit sous la forme

$$f = \pi_{N^\perp} f + \pi_N f$$
$$f = \int \mathcal{L}(dx)\, A^*\, L_x + C_\ell\, f^\ell \tag{5.14}$$

et puisque $\pi_{N^\perp} f \in N^\perp$

$$\int \mathcal{L}(dx)\, A f_x^\ell = 0 \tag{5.15}$$

Notons que $\{L_a\}$ forme un ensemble total pour S. Donc si on substitue (5.14) dans (5.13) et si on prend le produit scalaire avec L_a, on obtient avec l'équation (5.15) le système

$$\begin{aligned} &\int \mathcal{L}(dx)\, \left\{(L_x,\ L_a) + (A^* L_x,\ A^* L_a)\right\} + C_\ell\, A f^\ell(x_a) = g_a \\ &\int \mathcal{L}(dx)\, A f_x^\ell = 0 \end{aligned} \tag{5.16}$$

qu'on peut résoudre pour obtenir l'estimation de f dans l'équation (5.14). On démontre que ces équations sont caractéristiques de la solution unique à l'aide du formalisme du krigeage.

Le problème du krigeage nécessite l'estimation de $L_x^* \in S$ et $H_x^* = A^* L_x^*$. Ce problème s'interprète dans le graphe (F,S). Choisissons la norme suivante sur l'espace produit $F \times S$

$$(f,g) \in F \times S \qquad \|(f,g)\|^2 = \|f\|_F^2 + \|g\|_G^2$$

Soit $\varphi : S \to F \times S$ une application linéaire et continue définie par $\varphi(f) = (A^* f, f)$. Il est facile de démontrer que $\varphi^*(f,g) = \pi_S Af + g$ et donc $\mathrm{Im}(\varphi^*) = S$ qui est fermée. Par conséquent $\varphi(S)$ est fermé dans $F \times S$. Maintenant le problème se réduit à une répétition du théorème 5.2 si on choisit

$$\tilde{H}_x = (H_x, 0) \qquad \tilde{H}_x^* = \left(\int \mathcal{L}_x'(dy)\, A^* L_y,\ \int \mathcal{L}_x'(dy)\, Ly \right)$$

et $\bar{Y} = (Y, 0)$ pour $Y \in N$.

Alors le problème s'écrit

$$\underset{H_x^* \in \varphi(S)}{Min} \| \bar{H}_x - \bar{H}_x^* \| \quad \text{tel que} \quad < \bar{H}_x - \bar{H}_x^*, \bar{Y} > = 0 \quad \forall\ \bar{Y} \in N \times \{0\}$$

L'argument suit exactement les lignes du théorème 5.2, sauf que maintenant on a

$$V = \varphi(S) \cap (N^{\perp} \times S)$$

$$V' = \pi_{\varphi(S)} (N \times \{0\})$$

et la projection $\pi_{\varphi(S)} : N \times \{0\} \to V'$ est bijective parce que $\ker \pi_S A \cap N = \{0\}$. Nous avons donc $\varphi(S) = V \oplus V'$ et

$$(\bar{H}_x - \bar{H}_x^*, Y) = 0 \qquad \forall\ Y \in V$$

Comme dans le théorème 5.2, pour un $Z \in \varphi(S)$, on trouve un n unique, $n \in N \times \{0\}$ tel que

$$(\bar{H}_x - \bar{H}_x^*, Z) = (n, Z) \qquad \forall\ Z \in \varphi(S)$$

Puisque $(A^* L_a, L_a)$ est une famille totale pour $\varphi(S)$, on trouve que le problème de krigeage a une solution

$$\bar{H}_x^* = \left(\int \lambda_x'(dy)\, A^* L y,\ \int \lambda_x'(dy)\, L y \right)$$

qui est caractérisée par

$$\int \lambda_x'(dy) \left[(A^* L_\beta,\ A^* L_y) + (L_\beta,\ L_y) \right] + C_\ell'\, A f^\ell(x_\beta) = (H_x,\ A^* L_\beta) \tag{5.17}$$

$$\int \lambda_x'(dy)\, A f^\ell(y) = f_x^\ell$$

On peut conclure que le système d'équations (5.16) caractérise le problème de régularisation. Il est facile de démontrer comme dans le théorème (5.2) que les deux méthodes ont la même solution. ■

NOTA 1. – En appliquant le lemme 5.3, on peut trouver les équations dans le cas fini

$$\lambda^t [A\ K\ A^T + B] + C^t\ A\ f^\ell = g$$

$$\lambda^t\ A\ f^\ell \qquad\qquad = 0$$

où $K = \text{Cov}\ (Y_i,\ Y_j) = (H_i,\ H_j)$

$B = \text{Cov}\ (\epsilon_i,\ \epsilon_j)$.

NOTA 2. – Les conditions d'existence d'une solution unique du problème de régularisation (cokrigeage) sont plus faibles que pour le problème des splines (krigeage). Ceci est encore plus marqué dans le cas d'une minimisation sous contraintes (voir théorème 5.4).

NOTA 3. – On a démontré que chaque problème de régularisation se réduit à un problème de cokrigeage. Il est vrai que chaque cokrigeage de ce type (filtrage) s'interprète également en termes de régularisation. Par contre, le cokrigeage général n'admet pas d'interprétation évidente en termes de problème de régularisation.

NOTA 4. – Revenons brièvement sur la notation originale de la régularisation

$$\min_f\ \Omega(f) + B(g - Af)$$

où Ω et B sont des formes quadratiques positives. Dans le chapitre III, on a noté que dans la littérature concernant la régularisation, on utilise parfois un estimateur *a priori* $h(x)$ pour $f(x)$. Il faut donc choisir f tel qu'il minimise

$$\min_f\ \Omega(f - h) + B(g - Af) \qquad \textbf{(5.18)}$$

En termes de krigeage, on peut considérer h comme une dérive complètement spécifiée (au lieu d'être spécifiée à un paramètre près, ce qui est le cas le plus courant et correspond à la méthode de dérive externe [11]). Pour examiner de plus près cette interprétation, nous allons modifier le problème de régularisation (5.18).

Soient $\tilde{f} = f - h$ et $\tilde{g} = g - Ah$. Alors (5.18) s'écrit

$$\min_{\tilde{f}} \Omega(\tilde{f}) + B(\tilde{g} - A\tilde{f}) \qquad \textbf{(5.19)}$$

et on voit que les solutions des deux problèmes (5.18) et (5.19) sont liées par $f^* = \tilde{f}^* + h$.

Mais on sait que (5.19) est équivalent à un krigeage. Explicitons lequel. Si

$$Z = A\,Y + \epsilon$$

Appelons $\tilde{Y}(x) = Y(x) - h(x)$ et $\tilde{Z}(x) = Z(x) - Ah(x)$ et choisissons l'équivalence spline/krigeage entre (5.19) et le krigeage des résidus $\tilde{Y}(x)$ par les observés résiduels $\tilde{Z}(x)$. Mais $h(x)$ étant connue, on sait que le krigeage de $Y(x)$ n'est autre que

$$Y^*(x) = \tilde{Y}^*(x) + h(x)$$

et par conséquent, le krigeage de Y(x) en présence d'une dérive fixée $h(x)$ est équivalent à la régularisation avec un estimateur *a priori* (5.17).

En ce qui concerne la méthode de dérive externe avec filtrage, il est facile de voir qu'elle est équivalente à la régularisation suivante

$$\min_{f,a} \Omega(f - ah) + B(g - Af)$$

où a est un paramètre à déterminer. En effet, $\Omega(f - ah)$ est nulle quand $f \in ah + N$. Donc Ω est coercive sur le sous-espace $\tilde{N}^{\perp}$, où $\tilde{N} = N + E$ et E est l'espace d'une dimension générée par h. Comme nous l'avons fait au début de ce chapitre, nous pouvons redéfinir la norme sur F, et la régularisation devient

$$\min_{f} \| \pi_{\tilde{N}^{\perp}} f \|^2 + (g - Af,\ \pi_S\,(g - Af))$$

ce qu'on peut résoudre comme précédemment. Notons qu'il y aura des conditions de dérive supplémentaires dues à h. Elles ne sont rien d'autre que les conditions de la dérive externe.

Bien que nous ayons considéré ici les résultats obtenus pour la dérive externe avec filtrage, il est clair qu'on a le même type de résultats pour le krigeage avec dérive externe sans filtrage.

Dans la littérature sur les splines, on s'intéresse aux problèmes des splines d'interpolation dans un convexe. Il existe des théorèmes d'existence et d'unicité pour ces problèmes [11], mais ils sont nettement plus difficiles, et beaucoup plus contraignants que le théorème 5.2. Grâce au théorème de Stampachia, on peut donner ici des conditions très simples pour l'existence et l'unicité d'un problème de régularisation (spline d'ajustement) dans un convexe. Malheureusement, ce n'est pas un théorème constructif, et comme pour le problème des splines dans un convexe, il semble qu'il soit nécessaire de faire appel aux méthodes d'analyse convexe en dimension finie pour expliciter une solution. Par exemple, nous avons résolu le problème de poissonisation à l'aide d'une programmation quadratique dans le chapitre III.

THÉORÈME 5-4. – Soit K un ensemble convexe, fermé, non vide. Alors le problème de régularisation

$$\min_{f\in K} Q = (g-Af),\ \pi_S(g-Af)) + \| \pi_{N^\perp} f \|$$

a une solution unique $f\in K$ quand $N \cap \ker \pi_S A = \{0\}$.

DÉMONSTRATION. – Comme pour le théorème 5.3, on peut écrire

$$Q = (g,g) + 2 < \varphi, f > + B(f,f)$$

où (g,g) est une constante,

φ est le représentant de A^*g dans F'

et $B(f,f) = (\pi_{N^\perp} f,\ \pi_{N^\perp} f)_F + (\pi_S Af,\ \pi_S Af)_G$ est une forme bilinéaire continue et coercive.

On peut donc appliquer le théorème de Stampachia pour conclure qu'une solution unique $f\in K$ existe. De plus cette solution est caractérisée par

$$f\in K$$

$$(\pi_{N^\perp} f + A^*\pi_S Af - A^*g,\ h-f) \geq 0 \quad \forall\ h \in K \qquad (5.18)$$

■

Cette procédure de régularisation n'admet pas de formulation évidente en termes de krigeage, dans le sens qu'on ne trouve pas facilement une fonction

aléatoire et une minimisation de variance qui donnent le même résultat. Néanmoins on voit dans l'équation (5.18) et dans le développement du théorème 5.3 qu'une interprétation des normes comme l'inverse d'une matrice de covariance reste valable, ce qui justifie notre démarche dans l'application de la régularisation au problème de la poissonisation.

Dans le cas fini dimensionnel, il doit être possible de trouver une caractérisation en termes "d'équations de krigeage dual sous contraintes" du problème de régularisation de la même manière que dans [12], pour la définition du "krigeage positif". Nous n'avons pas suivi cette voie parce qu'il n'y avait pas d'applications directes aux problèmes pratiques qui sont traités ici.

CHAPITRE VI

LE FILTRAGE NON LINÉAIRE

6.0. — *INTRODUCTION*

Jusqu'ici nous avons traité de deux problèmes : l'estimation du signal sous-jacent $Y(x)$, qui a été traité par des méthodes spatiales linéaires (krigeage linéaire), et l'estimation de la loi de $\{Y(x)\}$ dans laquelle la distribution spatiale de $Y(x)$ n'a joué aucun rôle. Si l'on veut maintenant estimer une fonction non-linéaire de $Y(x)$ qui est elle-même une variable spatiale, il semblera assez naturel d'utiliser un estimateur non linéaire. Par exemple, si l'on veut estimer l'histogramme local dans un domaine v

$$F_v(z) = P[Y(x) < z | x \in v] = \frac{1}{v} \int_v \mathbb{1}_{(Y(x)<z)}\, dx$$

il faudra alors estimer $\mathbb{1}_{(Y(x)<z)}$ qui est une fonction non linéaire de $Y(x)$. Dans ce chapitre, nous traiterons de deux estimateurs qui sont l'espérance conditionnelle et le krigeage disjonctif.

Il est intéressant de voir si, dans la pratique, un filtrage non-linéaire donne de meilleurs résultats que le krigeage linéaire pour l'estimation du signal sous-jacent $Y(x)$. Ceci n'est nullement évident. Par exemple l'espérance conditionnelle peut donner de mauvais résultats si le modèle choisi est mal adapté. Dans le krigeage linéaire, on n'utilise que des fonctions de covariance qui, comme on l'a vu, sont assez faciles à estimer à partir des données observées. Pour l'espérance conditionnelle, par contre, il faudra connaître la loi multivariable, laquelle n'est pas accessible à l'aide des seules données. L'estimation par krigeage linéaire peut être considérée comme une procédure objective, alors que l'espérance conditionnelle ne sera pas en général objective, sauf peut-être dans les cas où la physique du phénomène est bien connue, et c'est le modèle objectif physique qui suggère la loi multivariable avec laquelle on travaillera (on peut penser par exemple à une modélisation avec un processus de diffusion).

En position intermédiaire entre l'espérance conditionnelle et le krigeage linéaire, se trouve le krigeage disjonctif, [1], qui dépend seulement des lois bivariables et est par conséquent plus facile à modéliser à partir des données.

6.1. — ESTIMATEURS PROVENANT D'UNE PROJECTION

Deux types d'estimateurs non-linéaires sont introduits à titre de rappel dans ce chapitre, le krigeage disjonctif et l'espérance conditionnelle. Ces estimateurs appartiennent avec le krigeage linéaire à une famille d'estimateurs provenant d'une projection dans un sens donné sans rigueur mathématique dans les paragraphes suivants.

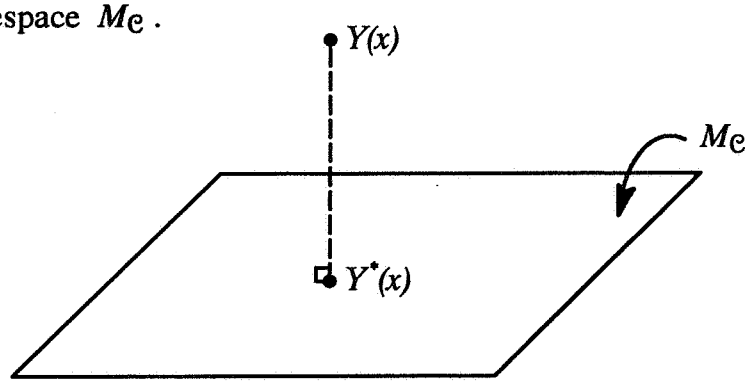
Comme précédemment, $Y(x)$ désignera la valeur à estimer connaissant des observations des variables aléatoires $Z(x_1),\ldots,Z(x_n)$. $Y(x)$ et $Z(x)$ sont des fonctions aléatoires strictement stationnaires. On aura l'occasion d'utiliser la notation $Z(x_i) = Z_i$. Soit $M_{\mathcal{C}}$ un espace linéaire contenant une certaine classe $\mathcal{C}$ de fonctions de $Z_1,\ldots, Z_n$ admettant le moment d'ordre deux.

$$M_{\mathcal{C}} = \{ f(Z_1,\ldots,Z_n) \; ; \; f \in \mathcal{C} \}$$

$M_{\mathcal{C}}$ sera muni du produit scalaire de la covariance non-centrée

$$< f_1, f_2 > = \mathrm{E}\,[f_1(Z_1,\ldots,Z_n)\, f_2(Z_1,\ldots,Z_n)]$$

et $M_{\mathcal{C}}$ est supposé fermé pour la topologie induite par le produit scalaire. On peut donc considérer la projection de $Y(x)$, ou plus généralement d'une fonction measurable $F(Y(x))$, dans l'espace $M_{\mathcal{C}}$.



La projection $Y^*(x)$ (ou $F^*(Y(x))$) est un estimateur de $Y(x)$. L'estimateur sera bon si $Y(x)$ est "proche" de $M_{\mathcal{C}}$ en termes de norme. On peut voir que si l'espace $M_{\mathcal{C}}$ devient plus grand, la norme de l'erreur $\| Y(x) - Y^*(x) \|$ devient plus petite. Néanmoins, comme on le verra dans les exemples suivants, ceci nécessite un renforcement des hypothèses à faire sur les données. En particulier si l'on choisit un espace $M_{\mathcal{C}}$ trop grand, on peut être amené à faire des hypothèses qu'on ne pourra pas vérifier en pratique.

Rappelons qu'une projection est caractérisée par le fait que l'erreur $(Y(x) - Y^*(x))$ est orthogonale à tous les vecteurs de $M_{\mathcal{C}}$.

$$< Y(x) - Y^*(x),\ f(Z_1,\ldots,Z_n) > = 0 \qquad \forall\ f(Z_1,\ldots,Z_n) \in M_{\mathcal{C}}$$

Mais $Y^*(x)$ est dans $M_{\mathcal{C}}$. Elle prend donc la forme $Y^*(x) = g(Z_1,\ldots,Z_n)$ et $g(Z_1,\ldots,Z_n)$ sera alors caractérisé par l'équation :

$$< g(Z_1,\ldots,Z_n),\ f(Z_1,\ldots,Z_n) > = < Y(x),\ f(Z_1,\ldots,Z_n) > \qquad \forall\ f \in M_{\mathcal{C}}$$

C'est cette procédure qu'on utilisera maintenant pour définir le krigeage simple, le krigeage disjonctif et l'espérance conditionnelle.

6.1.1. — Krigeage simple

Ici la classe $\mathcal{C}$ représente les combinaisons linéaires de $\{Z_i\}$ et du vecteur constant 1.

$$M_{\mathcal{C}} = \left\{ C + \sum_{i=1}^{n} \mu_i\, Z_i\ ;\ C,\ \mu_i \in \mathbb{R} \right\}$$

$M_{\mathcal{C}}$ est un espace de dimension $(n+1)$ avec pour base $\{1,\ Z_1,\ldots,Z_n\}$. Puisque $Y^*(x) \in M_{\mathcal{C}}$, il existe $\lambda_0,\ \lambda_1,\ldots,\lambda_n$ tels que $Y^*(x) = \lambda_0 + \Sigma\ \lambda_i\ Z_i$ et $Y^*(x)$ est caractérisé par le fait que $(Y(x) - Y^*(x))$ est orthogonal à tous les vecteurs de $M_{\mathcal{C}}$. Pour cela, il suffit que $(Y(x) - Y^*(x))$ soit orthogonal à chaque vecteur de base.

$$\begin{aligned} &< Y(x) - Y^*(x),\ 1 > = 0 \\ &< Y(x) - Y^*(x),\ Z_i > = 0 \qquad\qquad i = 1,\ldots,n \end{aligned} \tag{6-1}$$

Introduisons maintenant la covariance centrée :

$$\text{Cov}\ [Z(x),\ Z(y)] = \text{E}\ [Z(x), Z(y)] - m_Z^2 = < Z(x),\ Z(y) > - m_Z^2$$

$$\text{Cov}\ [Z(x),\ Y(y)] = \text{E}\ [Z(x), Y(y)] - m_Z\ m_Y$$

où $m_Z = \text{E}\ [Z(x)]$ et $m_Y = \text{E}\ [Y(x)]$.

La première équation de *(6-1)* donne alors :

$$\lambda_0 = \sum_{i=1}^{n} \lambda_i \, m_Z + m_Y \qquad (6\text{–}2)$$

et en utilisant *(6–2)*, la deuxième équation de *(6–1)*

$$\sum_{i=1}^{n} \lambda_i \, \mathrm{Cov}\,[Z_i, \, Z_j] = \mathrm{Cov}\,[Y(x), \, Z_j] \qquad (6\text{–}3)$$

qui sont les équations du krigeage linéaire avec moyenne connue. Le terme $(Y(x) - Y^*(x))$ est l'erreur résiduelle et la variance de krigeage est donc :

$$\sigma_K^2 = \| \, Y(x) - Y^*(x) \, \|^2 = \mathrm{Var}(Y(x)) - \sum_{i=1}^{n} \lambda_i \, \mathrm{Cov}(Y(x), \, Z_i) \, .$$

La solution des équations de krigeage linéaire nécessite seulement les fonctions de covariance $\mathrm{Cov}(Z(x), Z(y))$ qu'on peut estimer directement avec les données et $\mathrm{Cov}(Y(x), Z(y))$ qu'on obtient avec deux acquisitions ou bien si l'on connaît explicitement la loi du bruit (voir Chapitre II). Si, par contre, l'on veut estimer une fonction $F(Y(x))$, alors on peut aussi facilement écrire le système des équations à résoudre :

$$\Sigma \, \lambda_i \, \mathrm{Cov}(Z_i, Z_j) = \mathrm{Cov}(F(Y(x)), Z_j)$$

mais en général il n'y aura pas de moyen simple d'estimer $\mathrm{Cov}(F(Y(x)), Z_j)$ sans connaître la loi bivariable (Y,Z). Puisqu'on est obligé d'utiliser la loi de (Y,Z), on peut aussi bien utiliser un estimateur qui l'exploite complètement. C'est le krigeage disjonctif.

6.1.2. — Le krigeage disjonctif

Soit $F_Z(dx)$ la loi stationnaire de $Z(x)$ et $L^2(F_Z)$ l'espace des fonctions à carré intégrable par rapport à F_Z

$$f \in L^2(F_Z) \Leftrightarrow \int |f(x)|^2 \, F_Z \, (dx) < \infty$$

On prend maintenant l'espace $M_{\mathfrak{C}}$ de la forme :

$$M_{\mathfrak{C}} = \left\{ \sum_{i=1}^{n} f_i(Z_i) \; ; \; f_i \in L^2(F_Z) \right\}$$

Cette classe est évidemment beaucoup plus grande que celle utilisée pour le krigeage linéaire. Notre estimateur prend la forme $Y^*(x) = \Sigma f_i(Z_i)$ et il faut maintenant déterminer les fonctions f_i au lieu d'estimer uniquement les pondérateurs comme dans le cas précédent. Les f_i sont déterminées si pour chaque $\tilde{Z} \in M_{\mathcal{C}}$ on a :

$$< Y(x) - Y^*(x),\ \tilde{Z} > = 0$$

Puisque $\tilde{Z} = \Sigma g_k(Z_k)$ pour des $g_k \in L^2(F_Z)$, il suffit que :

$$< Y(x) - Y^*(x),\ g_k(Z_k) > = 0 \qquad k = 1,\ldots,n\ ;\ g_k \in L^2(F_Z)$$

$$\Leftrightarrow \mathrm{E}\left[(Y(x) - Y^*(x))\ g_k(Z_k)\right] = 0$$

$$\Leftrightarrow \mathrm{E}\left[g_k(Z_k)\ E\left[Y(x) - Y^*(x) \mid Z_k\right]\right] = 0$$

$$\Leftrightarrow \mathrm{E}\left[Y(x) - Y^*(x) \mid Z_k\right] = 0 \qquad \forall\ k = 1,\ldots,n$$

Les f_i sont donc déterminées par les équations :

$$\mathrm{E}\left[Y(x) \mid Z_k\right] = \sum_{i=1}^{n} E\left[f_i(Z_i) \mid Z_k\right] \qquad k = 1,\ldots,n \qquad \textbf{(6–4)}$$

Notons que l'on peut estimer aussi facilement $f(Y(x))$ pour une fonction arbitraire f. En effet, si $f^*(Y(x)) = \Sigma f_i(Z_i)$, les f_i satisfont à :

$$E\left[f(Y(x)) \mid Z_k\right] = \sum_{i=1}^{n} E\left[f_i(Z_i) \mid Z_k\right] \qquad k = 1,\ldots,n \qquad \textbf{(6–5)}$$

La variance d'estimation est $\| Y(x) - f^*(Y(x)) \|^2$. Les équations à résoudre dépendent des lois bivariables $(Y(x), Z(y))$ et $(Z(x), Z(y))$, et il convient de les modéliser avant de procéder à l'estimation.

6.1.3. — L'espérance conditionnelle

Le troisième exemple d'un estimateur provenant d'une projection est aussi le plus puissant. Il s'agit de l'espérance conditionnelle. On a besoin ici de la loi de

$(Y(x), Z(x_1),\ldots, Z(x_n))$. Supposons que $F_Z(dx_1,\ldots,dx_n)$ soit la loi de $(Z_1,\ldots,Z_n)$. Avec la même notation que précédemment, on appelle $L^2(F_Z)$ l'espace des fonctions à carré intégrable par rapport à $F(dx_1,\ldots,dx_n)$

$$f \in L^2(F_Z) \Leftrightarrow \int |f(x_1,\ldots,x_n)|^2 \, F(dx_1,\ldots,dx_n) < \infty$$

et l'on considère :

$$M_{\mathcal{C}} = \left\{ f(Z_1,\ldots,Z_n),\ f \in L^2(F) \right\}$$

L'estimateur devient maintenant $f(Y)^* = g(Z_1,\ldots,Z_n)$. L'estimateur est déterminé si, pour tout $\tilde{Z} \in M_{\mathcal{C}}$:

$$< Y(x) - Y^*(x), \tilde{Z} > = 0$$

et comme pour le krigeage disjonctif, ceci est vrai si et seulement si :

$$E\,[Y(x) - Y^*(x) | Z_1,\ldots,Z_n] = 0$$

6.2. — LE KRIGEAGE DISJONCTIF

Retournons au krigeage disjonctif pour l'étudier plus en détail. Comme on l'a vu, les équations à résoudre sont :

$$E\,[Y(x)|Z_i] = \sum_{k=1}^{n} E[f_k(Z_k)|Z_i] \qquad i = 1,\ldots,n$$

Soit $P^{Z_iZ_j}(dx,dy)$ la loi bivariable de $(Z_i,\ Z_j)$, et $P^{Y_iZ_j}(dx,dy)$ la loi de $(Y_i,\ Z_j)$. Supposons que la loi conditionnelle de $(Z_i|Z_j)$ existe et soit $P^{Z_i|Z_j}(dx)$, et que de plus $(Y|Z_j)$ existe avec pour loi $P^{Y|Z_j}(dx)$. Les équations du krigeage disjonctif pour l'estimation de $f(Y(x))$ prennent la forme :

$$\int f(y)\ P^{Y/Z_j}\,dy = \sum_{k=1}^{n} \int f_k(y)\ P^{Z_k/Z_j}(dy) \qquad j = 1,\ldots,n \qquad (6\text{–}6)$$

Le krigeage disjonctif s'écrit alors en termes d'un système d'équations intégrales. Si on voulait résoudre ces équations directement – et on en verra un

exemple dans la section suivante –, une procédure possible serait de commencer avec un modèle de $(Y(x),Y(y))$. Le plus facile sera de prendre un modèle assez simple, par exemple l'anamorphose d'une loi bivariable gaussienne ou gamma (voir 6.3 et 6.4). Parfois on connaît un modèle de l'origine physique de l'image (par exemple une processus de diffusion, un modèle de feuilles mortes [2], ou des fonctions booléennes [3]), et on connaît ainsi la loi de $(Y(x),Y(y))$ à quelques paramètres près qu'on peut estimer à l'aide des données observées $(Z(x),Z(y))$ et de notre connaissance de la loi du bruit $(Z(x)/Y(x))$. On peut même essayer une estimation directe de la loi bivariable à partir des variables observés $(Z(x),Z(y))$ par une méthode d'inversion comme celle que l'on a vue au chapitre V. Mais cette dernière est trop lourde en pratique, et il faudra estimer les lois bivariables au moins pour plusieurs couples $(Z(x),Z(y))$ de séparation $(x–y)$ et ensuite trouver une famille de lois bivariables cohérentes (provenant d'une fonction aléatoire) qui ressemble à nos lois estimées.

Quand on a choisi les lois bivariables $(Y(x),Y(y))$, on peut facilement calculer $\mathrm{P}^{Z_i|Z_j}$ et $\mathrm{P}^{Y|Z_j}$ en fonction de la randomisation Y|Z (par exemple avec une poissonisation ou par un autre type de randomisation). Ceci nous donnera au moins une vérification partielle lors de la comparaison avec la version expérimentale de $\mathrm{P}^{Z_i|Z_j}$. Pour finir, on discrétise $\mathrm{P}^{Z_i|Z_j}$ et $\mathrm{P}^{Y|Z_j}$ et on résoud les équations linéaires dérivant de *(6–6)*.

Actuellement, cette méthode semble trop coûteuse pour être utilisée systématiquement dans des applications actuelles sauf, peut être, dans quelques cas particuliers. C'est néanmoins la seule méthode qui assure le libre choix des lois bivariables et en même temps un bon modèle de la loi de randomisation $Z(x)|Y(x)$.

Dans la pratique de la géostatistique minière, on ne résoud pas directement les équations du KD. Les données sont anamorphosées en une loi isofactorielle pour laquelle le krigeage disjonctif admet une forme plus simple. Nous considérerons les lois isofactorielles au paragraphe 6–3.

6.2.1. — Un exemple simple

Pour essayer de résoudre directement les équations du krigeage disjonctif, nous allons considérer l'image suivante à une dimension. Supposons la droite divisée en segments par un processus de Poisson. Sur chaque intervalle, on tire la valeur

1 ou 0 avec la probabilité p et 1-p respectivement, de façon à ce que les valeurs prises par les intervalles soient indépendantes deux par deux. Ceci constitue un exemple simple d'un processus mosaïque. On suppose maintenant que cette "image" est observée avec un bruit poissonien. Sur la Fig. 6-1, on voit l'image sous-jacente et sur la Fig. 6-2 l'image bruitée est superposée à l'image sous-jacente. Le but est ici de comparer les estimations de l'image sous-jacente obtenues par krigeage linéaire et par krigeage disjonctif. Nous ne disons pas que ce soient les meilleures méthodes pour résoudre ce problème particulier. Le but est tout simplement de comparer les krigeages linéaire et disjonctif dans un cas d'école. Notons bien que le bruit est très bien corrélé à l'image sous-jacente. En particulier, il ne peut y en avoir que lorsque l'image sous-jacente prend des valeurs non-nulles.

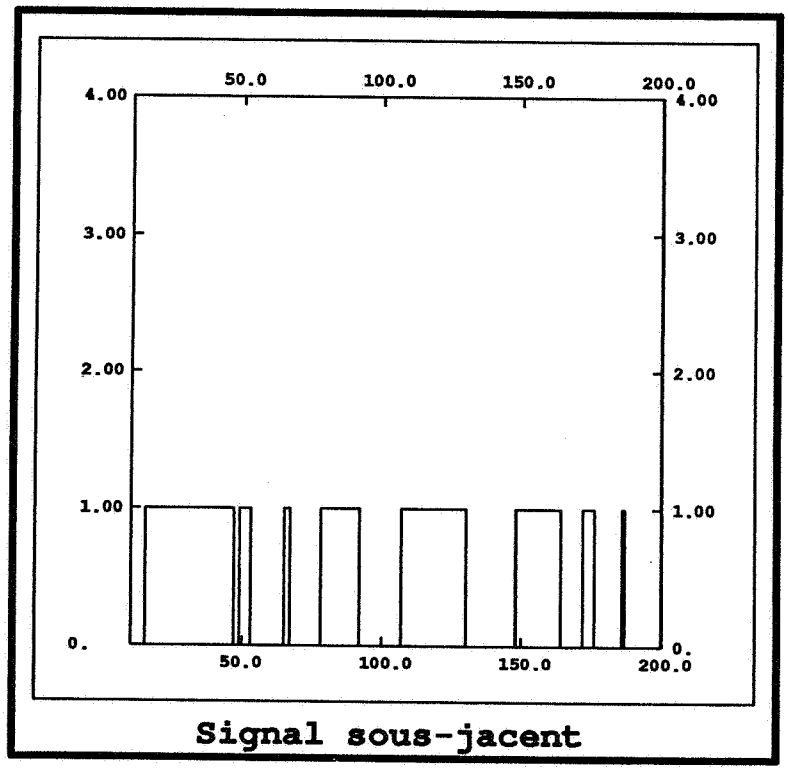


Figure 6.1 Signal sous-jacent

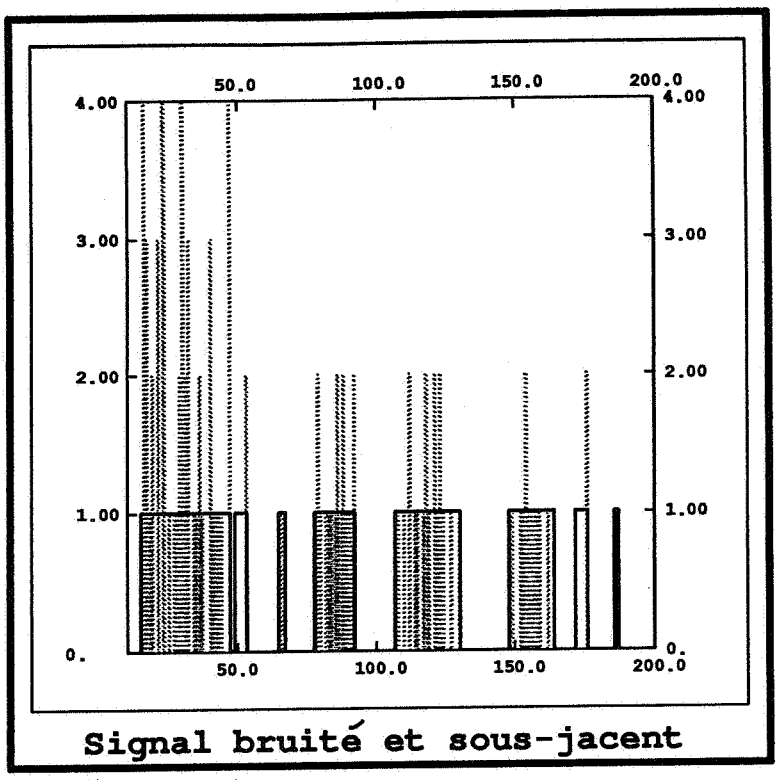


Figure 6.2 Signal sous-jacent ... trait continu
Signal bruité ... pointillé

Il est facile de démontrer que la fonction de covariance de l'image sous-jacente prend la forme

$$C^Y(h) = (p - p^2)\, e^{-|h|/d}$$

et que la moyenne est $m = p$. Par conséquent, la fonction de covariance pour l'image bruitée est

$$C^Z(h) = (p - p^2)\, e^{-|h|/d} + p\delta_{h0}$$

et le rapport signal sur bruit SSB $= 1 - p$. Soit $\varrho(h)$ la fonction de corrélation de Y, $\varrho(h) = e^{-|h|/d}$. Il est facile de vérifier que la loi bivariable pour deux points d'un processus mosaïque séparé par une distance h est ([5]) :

$$P_h(dx, dy) = \varrho(h)\, \delta_x(dy)\, F(dx) + (1 - \varrho(h))\, F(dx)\, F(dy)$$

où $F(dx)$ est la loi univariable.

Dans ce cas $$F(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x \le 1 \\ 1 & x \le 1 \end{cases}$$

d'où on tire la loi bivariable pour $Y(x)$ et $Y(x+h)$ séparés par h.

$$P_{00}(h) = P\,[Y(x) = 0,\ Y(x+h) = 0] = (1-p)\,e^{-|h|/d} + (1-p)^2\,(1-e^{-|h|/d})$$

$$P_{10}(h) = P\,[Y(x) = 1,\ Y(x+h) = 0] = p(1-p)(1-e^{-|h|/d}) = P_{01}(h)$$

$$P_{11}(h) = P\,[Y(x) = 1,\ Y(x+h) = 1] = p\,e^{-|h|/d} + p^2\,(1-e^{-|h|/d})$$

Il est très facile de trouver la poissonisée de cette loi :

$$P[Z(x) = n,\ Z(x+h) = m] = \begin{cases} \delta_{n0}\,\delta_{m0}\,P_{00}(h) + \dfrac{1}{e}\left(\delta_{m0}\dfrac{P_{10}(h)}{n!} + \delta_{n0}\dfrac{P_{01}(h)}{m!}\right) + \dfrac{1}{n!m!e^2}P_{11}(h) & \text{si } h > 0 \\ \delta_{nm}\left[(1-p)\,\delta_{n0} + \dfrac{p}{n!e}\right] & \text{si } h = 0 \end{cases}$$

où $n,\ m = 0,\ 1,\ 2,\ 3,\ldots$

On trouve aussi

$$P[Y(x) = i,\ Z(x+h) = m] = \delta_{m0}\,P_{i0}(h) + \frac{1}{m!\,e}P_{i1}(h) \qquad \forall h$$

où $i = 0,1\ ;\ m = 0,\ 1,\ 2,\ldots$

On peut maintenant essayer de résoudre les équations du krigeage disjonctif directement. Notons d'abord que

$$P^{Y|Z_i=m}(j) = \frac{P\,[Y = j,\ Z_i = m]}{P\,[Z_i = m]}.$$

$$\text{donc} \quad E[Y|Z_i = m] = \frac{P[Y = 1, Z_i = m]}{P[Z_i = m]}$$

Les équations du krigeage disjonctif deviennent

$$\sum_{\beta=1}^{n} \sum_{j=0}^{\infty} f_\beta(j)\ \mathrm{P}\ [Z_\alpha = j,\ Z_\beta = i] = \mathrm{P}\ [Y_x = 1,\ Z_\alpha = j]$$

$$\alpha = 1,\ldots,n$$
$$i = 0,\ 1,\ 2,\ldots$$

Dans cet exemple, nous construisons une mosaïque avec $p = ½$. Autrement dit, 50% des valeurs prennent la valeur 1. Les fonctions de covariance sont $C^Y(h) = ¼\ e^{-|h|/6}$ et $C^Z(h) = ¼\ e^{-|h|/6} + ½\delta_{h0}$ et donc le rapport signal sur bruit est égal à ½. Le nombre des données est de 400 et le voisinage utilisé est de 15 points. Les résultats du krigeage disjonctif et du krigeage linéaire avec moyenne connue sont montrés aux figures 6–3 et 6–4 respectivement. Les erreurs d'estimation et les rapports signal sur bruit sont résumés au Tableau 6–1. Les valeurs du krigeage linéaire avec moyenne inconnue ont été ajustées pour permettre la comparaison.

Tableau 6–1.

	Variance d'erreur	SSB
Image bruitée	0.512	0.487
Krigeage linéaire (moyenne inconnue)	0.089	2.80
Krigeage linéaire (moyenne connue)	0.084	2.98
Krigeage disjonctif	0.058	4.26

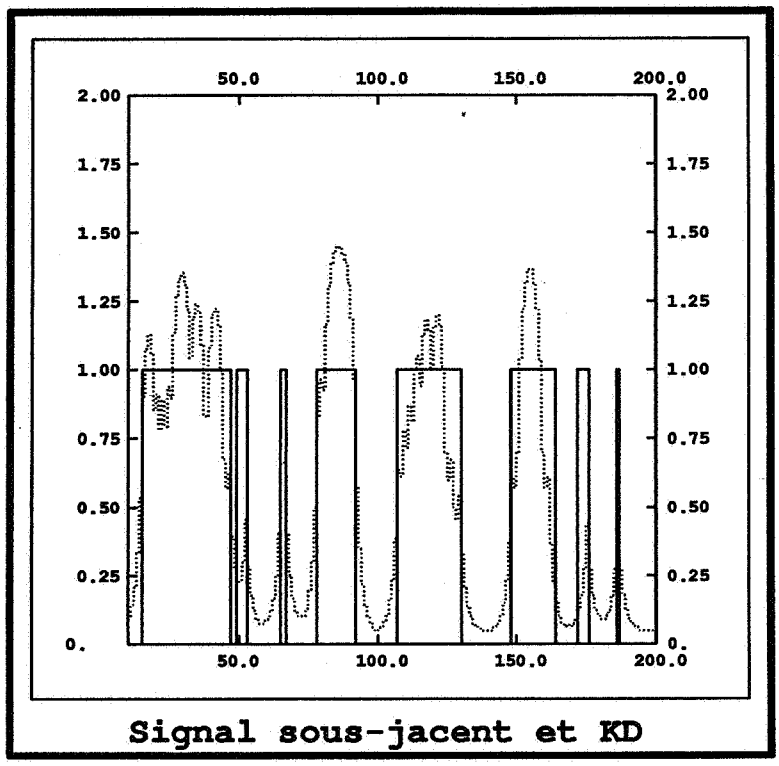


Figure 6.3 Estimateur par krigeage disjonctif

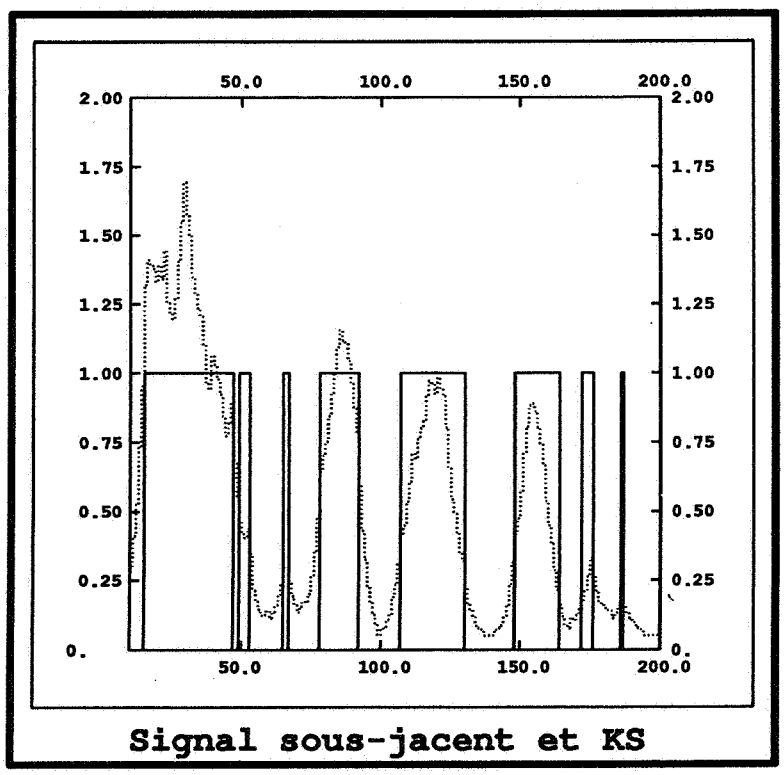


Figure 6.4 Estimateur par krigeage linéaire
(krigeage avec moyenne connue)

Pour ce type d'image, nous voulons avoir un résultat en termes de 0 et 1, et nous proposons donc de seuiller l'image. Toutes les valeurs supérieures à q $(0 \leq q \leq 1)$ prendront la valeur 1.

Si l'on fait un krigeage disjonctif avec un voisinage d'un point, l'estimateur en un point x où l'observation est non–nulle sera 1. On perd cette propriété lorsqu'on utilise un voisinage de plus d'un point. Pour la réimposer, nous avons décidé de choisir la coupure q telle que l'estimation en un point x où l'observation est non–nulle mais où tous les points du voisinage de x ont des observations nulles, sera mise égale à 1. On trouve $q = 0.438$.

Pour le krigeage linéaire, il n'existe pas de moyen aussi simple pour choisir une coupure parce qu'on ne prend pas en compte la loi de l'image sous–jacente. On a décidé de choisir $q = 0.5$ et il se trouve qu'en effet ce choix n'est pas loin de l'optimal trouvé expérimentalement (puisque l'on connait l'image réelle on peut faire varier q pour trouver la valeur optimale. Malheureusement dans un cas pratique nous n'aurons pas cette possibilité). On compare alors ces estimateurs avec l'estimateur naïf pour lequel on avait mis toutes les observations non–nulles à 1. Dans les figures 6–5 – 6–8, on voit ces trois estimateurs sur l'intervalle [1, 200] avec le processus sous–jacent comme référence.

Comme mesure d'erreur, on n'utilise pas la variance d'erreur mais le pourcentage de points mal classés qui est résumé pour tout le champ [8,392] au Tableau 6–2.

Tableau 6–2.

	Nombre d'erreurs	Pourcentage
Estimateur naïf	66	17.2%
Krigeage linéaire (moyenne connue)	31	8.1%
Krigeage disjonctif	14	3.6%

On peut voir sur les deux tableaux que le krigeage disjonctif donne des résultats qui sont meilleurs que ceux obtenus par le krigeage linéaire à la fois pour l'erreur quadratique et pour le nombre d'erreurs de classification après seuillage.

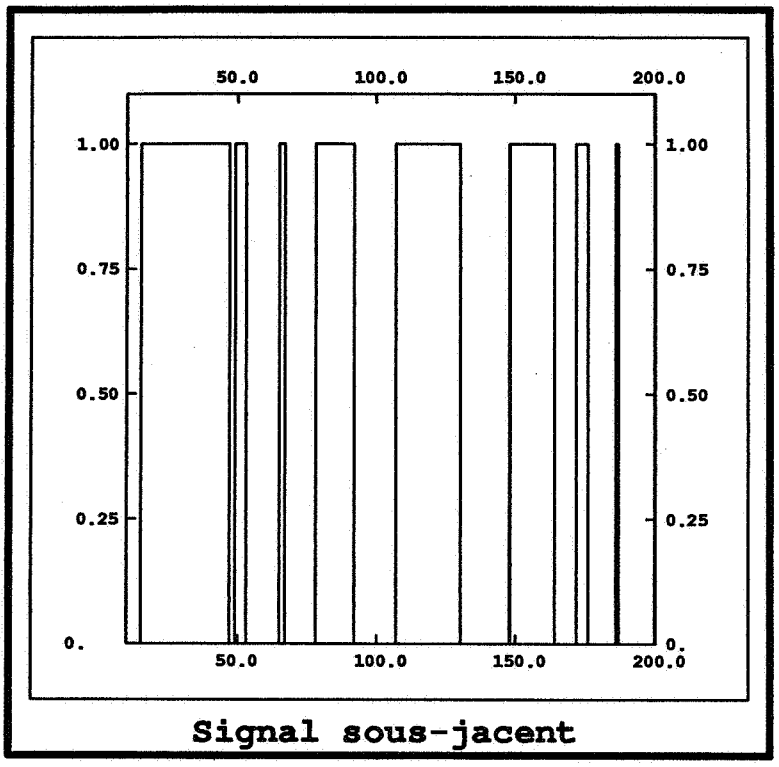


Figure 6.5 Signal sous-jacent

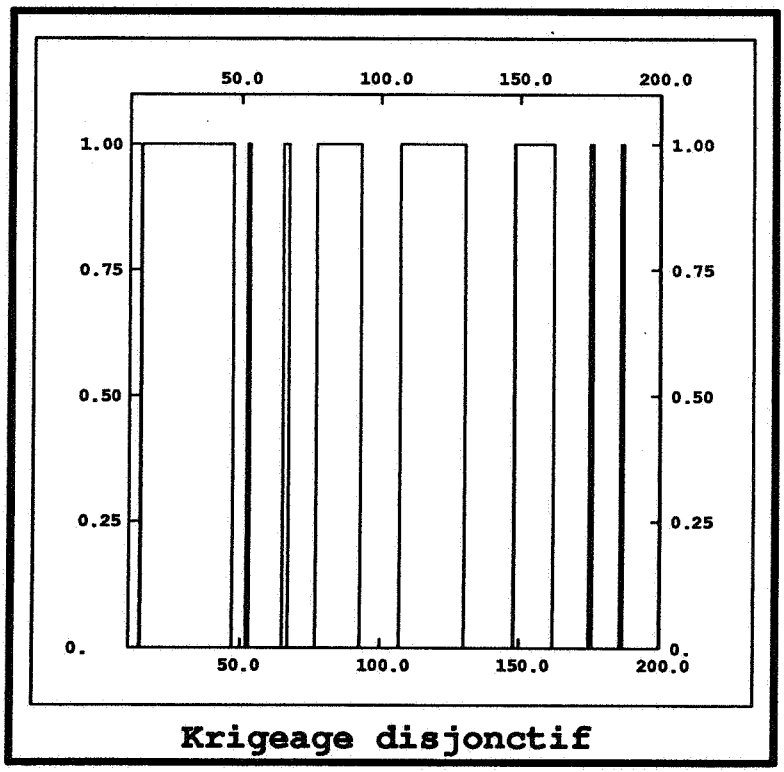


Figure 6.6 Estimateur par krigeage disjonctif
(et seuillage)

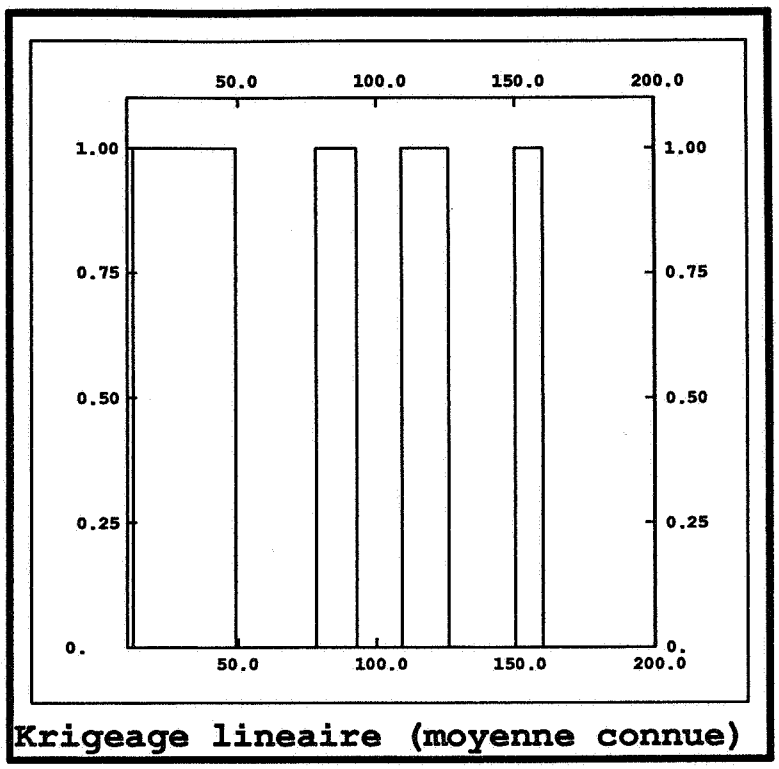


Figure 6.7 Estimateur par krigeage linéaire
(et seuillage)

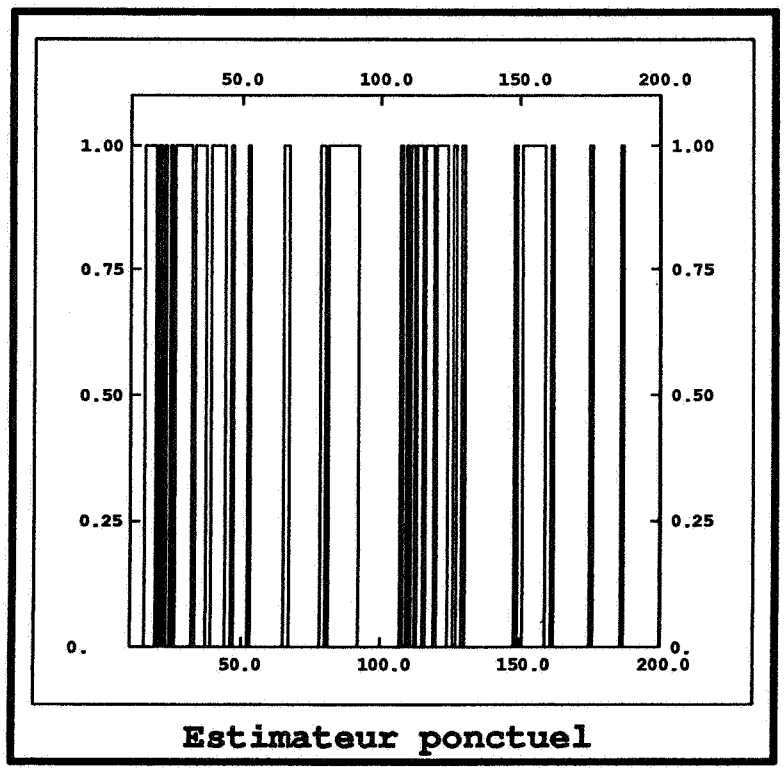


Figure 6.8 Estimateur sans filtrage spatial

6.3. — MODÈLES ISOFACTORIELS

Nous avons déjà remarqué que la solution directe des équations du krigeage disjonctif est souvent assez coûteuse en temps calcul. Ceci est une des raisons pour lesquelles en géostatistique minière on utilise des lois spéciales, qu'on appelle lois isofactorielles, pour lesquelles le krigeage disjonctif admet une forme simple. Nous notons que les lois $(Y(x),Y(y))$ ou $(Z(x),Z(y))$ sont des lois symétriques, et la loi bivariable entre le signal sous-jacent et le signal observé $(Y(x),Z(y))$ est une loi dissymétrique. En général, il n'est pas trop difficile de trouver des lois isofactorielles symétriques, mais étant donnée une loi isofactorielle $(Y(x),Y(y))$ et une loi de randomisation $(Z(y)|Y(x))$, la loi de $(Y(x),Z(y))$ ne sera pas en général isofactorielle.

Pour l'instant, nous allons supposer que toutes les lois bivariables sont isofactorielles puis nous verrons quelques exemples où cette condition est vraie.

Supposons que $Y(x)$ et $Z(x)$ sont des fonctions aléatoires strictement stationnaires. Si la loi bivariable pour $(Z(x),Z(x+h))$ prend la forme

$$P_h^Z(dx, dy) = \Phi_h(x,y)\ F^Z(dx)\ F^Z(dy)$$

où $F^Z(dx)$ est la loi univariable de Z, et si Φ_h est de la forme

$$\Phi_h(x,y) = \sum_{n=0}^{\infty} \mu_n(h)\ \chi_n(x)\ \chi_n(y)$$

où $\chi_n(x)$ est un système orthonormal complet pour $L^2(F^Z)$

$$\int \chi_n(x)\ \chi_m(x)\ F^Z(dx) = \delta_{nm}\,,$$

on dit que $Z(x)$ a des lois bivariables isofactorielles (ou souvent que Z est isofactorielle). Notons que les $\mu_n(h)$ ne sont pas des fonctions quelconques de h. En particulier, on voit que

$$\text{Cov}\ [\chi_n(Z(x)),\ \chi_n(Z(x+h))] = \mu_n(h)$$

Il n'est pas nécessaire de donner ici plus de détails sur les fondations des lois isofactorielles. Le lecteur intéressé pourra se reporter à [1] ou [6].

Supposons également que la loi dissymétrique est isofactorielle

$$P_h^{YZ}(dx, dy) = \sum_{n=0}^{\infty} \nu_n(h)\ \eta_n(x)\ \chi_n(y)\ F^Y(dx)\ F^Z(dy)$$

Notons que $\nu_n(h)$ est une fonction de covariance

$$\nu_n(h) = \text{Cov}\ [\eta_n(Y(x)),\ \chi_n(Z(x+h))]$$

où $\eta_n(x)$ est un système orthonormal complet pour $L^2(F^Y)$.

On considère maintenant à nouveau les équations du krigeage disjonctif. Puisque η_n est une famille complète, on a

$$\varphi(Y(x)) = \sum_{0}^{\infty} \varphi_n\ \eta_n(Y(x))$$

avec $\varphi_n = \int \varphi(y)\ \eta_n(y)\ F^Y(dy)$.

Notre estimateur est de la forme $\varphi^*(Y(x)) = \sum_{\alpha=1}^{N} f_\alpha(Z_\alpha)$. On peut développer chaque fonction f_α par

$$f_\alpha(Z_\alpha) = \sum_{n=0}^{\infty} f_\alpha^{\,n}\ \chi_n(Z_\alpha)$$

avec $f_\alpha^{\,n} = \int f_\alpha(x)\ \chi_n(x)\ F^Z(dx)$.

L'estimation de $\varphi(Y(x))$ se réduit alors à l'estimation des coefficients $\{f_\alpha^{\,n}\}$.

Notons les résultats importants :

$$\text{E}\ [\chi_n(Z_\alpha)|Z_\beta] = \int \chi_n(x) \left\{ \sum_{0}^{\infty} \mu_n(x_\alpha - x_\beta)\ \chi_m(x)\ \chi_m(Z_\beta)\ F^Z(dx) \right\} = \mu_n(x_\alpha - x_\beta)\ \chi_n(Z_\beta)$$

Egalement : $\text{E}[\eta_n(Y_X)|Z_\beta] = \nu_n(x - x_\beta)\ \chi_n(Z_\beta)$.

Maintenant les équations du krigeage disjonctif, qui sont

$$\text{E}\ [\varphi(Y_X)|Z_\beta] = \sum_{\alpha=1}^{N} \text{E}\ [f_\alpha(Z_\alpha)|Z_\beta] \quad \beta = 1,\ldots,n$$

se transforment à l'aide des résultats ci-dessus en :

$$\sum_{n=0}^{\infty} \varphi_n \nu_n(x-x_\beta)\, \chi_n(Z_\beta) = \sum_{\alpha=1}^{N} \sum_{n=0}^{\infty} f_\alpha^n \mu_n(x_\alpha - x_\beta)\, \chi_n(Z_\beta)$$

Les coefficients de $\chi_n(Z_\beta)$ doivent être égaux; nous obtenons donc un système d'équations linéaires pour chaque $n = 0, 1, 2, \ldots$ Soit :

$$\sum_{\alpha=1}^{N} \mu_n(x_\alpha - x_\beta)\, f_\alpha^n = \varphi_n\, \nu_b(x-x_\beta) \qquad \beta = 1,\ldots,N \qquad (6\text{–}7)$$

Rappelons que Cov $[\eta_n(Y_\alpha),\ \chi_n(Z_\beta)] = \nu_n(x_\alpha - x_\beta)$. On voit alors que le système d'équations *(6–7)* n'est que le krigeage linéaire de $(\varphi_n \eta_n(Y_x))$ à partir des données $\{\chi_n(Z_\beta)\}_{\beta=1}^{N}$. Autrement dit, l'estimateur de $\varphi(Y_x)$, prend la forme simple $\varphi^*(Y(x)) = \Sigma\ \varphi_n\ \eta_n^*(Y(x))$ où $\eta^*(Y(x))$ n'est que le krigeage linéaire de $\eta(Y(x))$ à partir de $\{\chi_n(Z_\beta)\}$.

On utilise essentiellement trois méthodes pour obtenir les modèles isofactoriels. La première consiste à régulariser des mesures aléatoires orthogonales [1]. La deuxième considère les processus de Markov, plus particulièrement ceux de naissance et de mort , et les processus de diffusion [7, 8, 9]. Pour finir, on a les processus de type mosaïque [10] et leur généralisation [11, 12].

Nous avons déjà dit qu'il n'y a en général pas grande difficulté à trouver des modèles isofactoriels pour modéliser les lois symétriques, au moins pour la variable observée $Z(x)$, et on trouve un grand nombre de modèles dans la littérature citée. Pour les lois dissymétriques, nous allons emprunter les lois utilisées en géostatistique minière pour la raison suivante. Normalement, dans une étude minière, on connaît des échantillons quasi-ponctuels, et on veut estimer la fonction de répartition des unités minières (ou blocs) soit globalement soit dans une zone assez grande (un panneau). Nous sommes alors intéressés par une estimation des variables avec un certain support, connaissant des données sur un autre support. Bien qu'il soit plus difficile de les trouver que dans le cas symétrique, la littérature déjà citée contient un grand nombre de modèles de changement de support qui sont isofactoriels. Dans la plupart des cas, ces modèles sont développés en considérant que l'échantillon est une randomisation de la valeur du bloc dont il provient. Pour ce type de modèle, si l'on considére le bloc comme étant notre variable sous-jacente et l'échantillon notre variable randomisée $Z(x)$, on voit que ces modèles restent valables pour le problème qui nous concerne. Nous allons tout de suite considérer quelques exemples.

6.3.1. — Modèles isofactoriels hermitiens

Si $Y(x)$ est une fonction aléatoire stationnaire et gaussienne ayant pour fonction de corrélation $\varrho(h)$, la densité bivariable $(Y(x),Y(x+h))$ est de la forme :

$$g_h^Y(x,y) = \frac{1}{\sqrt{1-\varrho^2}}\, e^{\frac{-1}{2(1-\varrho^2)}(\varrho^2x^2+\varrho^2y^2-2\varrho xy)}\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}$$

Posant $g(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$, on peut démontrer que

$$g_h^Y(x,y) = \sum_{n=0}^{\infty} \varrho^n(h)\, \eta_n(x)\, \eta_n(y)\, g(x)\, g(y)$$

où $\eta_n(x)$ est le polynôme d'Hermite normalisé d'ordre n

$$\eta_x = \frac{1}{\sqrt{n!}}\, H_n(x)$$

et $e^{-\frac{x^2}{2}}\, H_n(x) = \frac{d^n}{dx^n}\, e^{-\frac{x^2}{2}}$ est la formule de Rodrigues utilisée pour les générer. Il est néanmoins plus facile de générer ces polynômes à l'aide des relations de récurrence

$$\eta_0(x) = 1 \; ; \; \eta_1(x) = -x$$

$$\eta_{n+1}(x) = -\frac{1}{\sqrt{n+1}}\, x\, \eta_n(x) - \sqrt{\frac{n}{n+1}}\, \eta_{n-1}(x)$$

Notons aussi que, pour un couple bigaussien (X,Y) de corrélation ϱ, on a :

$$\mathrm{E}\,[H_n(X)\,|\,Y] = \varrho^n H_n(Y)$$

On construit maintenant la loi bivariable de $(Y(x),Z(x+h))$ à partir de la randomisation suivante :

$$Z(x) = r\, Y(x) + \sqrt{1-r^2}\, \epsilon(x)$$

où $0 \leq r \leq 1$, $Z(x),Y(x)$ et $\epsilon(x)$ sont tous $N(0,1)$ et ϵ est indépendant de Y. On a donc Corr $[Y(x),Z(x)] = r$. Il suit que $(Y(x),Z(x+h))$ est bigaussien avec la corrélation $r\varrho(x_1-x_2)$ et on peut immédiatement l'écrire sous sa forme isofactorielle

$$g_h^{Y,Z}(x,y) = \sum_{n=0}^{\infty} (r\varrho(x_1-x_2))^n \, \eta_n(x) \, \eta_n(y) \, g(x) \, g(y)$$

De plus, le couple $(Z(x),Z(x+h))$ est bigaussien avec la corrélation

$$\varrho_Z(h) = \begin{cases} r^2\varrho(h) & \text{si } h > 0 \\ 1 & \text{si } h = 0 \end{cases}$$

et la loi bivariable de $(Z(x),Z(x+h))$ s'écrit :

$$g_h^Z(x,y) = \sum \varrho_Z(h) \, \eta_n(x) \, \eta_n(y) \, g(x) \, g(y)$$

Nous avons maintenant les lois symétriques et dissymétriques sous forme isofactorielle. Le krigeage disjonctif de $\varphi(Y_x)$ pour une fonction $\varphi \in L^2(F^Y)$ se fait selon la méthode donnée au paragraphe 6.2.

Si $$\varphi(Y_x) = \sum \varphi_n \, \eta_n(Y_x)$$

alors $$\varphi^*(Y_x) = \sum \varphi_n \, \eta_n^*(Y_x)$$

où $\eta_n^*(Y_x) = \sum \lambda_\alpha^n \, \eta_n(Z_\alpha)$ et les λ_α^n sont les solutions de

$$\sum_{\alpha=1}^{N} \lambda_\alpha^n \, \varrho_Z^n(x_\alpha - x_\beta) = r^n \varrho^n(x - x_\beta) \qquad \beta = 1,\ldots,N$$

On peut démontrer facilement que la variance d'estimation s'écrit

$$\text{Var}\left[\varphi(Y_x) - \varphi^*(Y_x)\right] = \sum_{n=0}^{\infty} \varphi_n \left(\varphi_n - \sum_{\alpha=1}^{N} \lambda_\alpha^n \, r^n \varrho^n(x - x_\alpha) \right)$$

On peut généraliser les lois isofactorielles de lois marginales gaussiennes de la manière suivante ([13]).

Si X et Y sont des gaussiennes réduites et si les polynômes d'Hermite satisfont à

$$\mathrm{E}\,[\eta_n(X)|Y] = T_n \, \eta_n(Y)$$

et

$$\mathrm{E}\,[\eta_n(Y)|X] = T_n \, \eta_n(X)$$

alors nous avons une loi isofactorielle et tous les résultats énoncés restent valables pour cette loi plus générale (appelée loi hermitienne). La loi bivariable de X et Y est

$$g^{X,Y}(x,y) = \sum T_n \, \eta_n(x) \, \eta_n(y) \, g(x) \, g(y)$$

On peut démontrer que les coefficients $\{T_n\}$ proviennent d'une loi hermitienne si et seulement si

$$T_n = \int \varrho^n \, \omega(d\varrho)$$

pour $\omega(d\varrho)$ une loi de probabilité sur $[-1,1]$. Autrement dit la loi hermitienne est un mélange des lois gaussiennes par rapport à la loi $\omega(d\varrho)$. L'exemple le plus simple est la mosaïque gaussienne (sur chaque intervalle, la fonction prend une valeur tirée selon une loi de Gauss) pour laquelle $\omega(d\varrho)$ est concentrée en 0 et 1.

6.3.2. — Modèles isofactoriels de Laguerre

On dit que Y suit une loi gamma de paramètre a si sa densité $g_a(y)$ est

$$g_a(y) = \frac{1}{\Gamma(a)} e^{-y} \, y^{a-1}$$

où $\Gamma(a) = \int_0^\infty e^{-x} \, x^{a-1} \, dx$.

On veut utiliser des lois isofactorielles gamma. Pour démontrer qu'une telle loi existe, on considère une construction explicite à l'aide d'un processus de diffusion, $Y(x)$, de générateur infinitésimal

$$Af(x) = x \, f''(x) + (a - x) \, f'(x)$$

Il est bien connu que le processus associé a une loi marginale gamma. Les vecteurs propres du générateur sont une base orthogonale pour l'espace $L^2(g_a)$ soit $Af_n = \lambda_n f_n$ où $\lambda_n = n$ et $f_n(x) = L_n^a(x)$, $L_n^a(x)$ étant le polynôme de Laguerre d'indice a. Le polynôme de Laguerre normé ℓ_n^a est donné par

$$\ell_n^a = b_n^a \, L_n^a(x)$$

où $b_n^a = \sqrt{\frac{\Gamma(a+n)}{\Gamma(a)n!}}$, et est defini par la formule de Rodrigues

$$g_a(y)\, \ell_n^a(y) = b_n^a \frac{d^n}{dy^n}\, g_{a+n}(y)\,.$$

On peut utiliser les relations de récurrence pour calculer rapidement les polynômes de Laguerre.

$$L_0^a(x) = 1\ ,\ L_1^a(x) = \left(1 - \frac{y}{a}\right)$$

$$-(a+n)\, L_n^a(y) = \big(y - a - 2(n-1)\big)\, L_{n-1}^a + (n-1)\, L_{n-2}^a(y)\,.$$

Si l'on considère l'équation d'évolution du processus de diffusion, on peut démontrer que les densités bivariables $g_h^Z(x,y)$ sont données par

$$g_h^Z(x,y) = \sum_0^\infty \varrho^n(h)\, \ell_n^a(x)\, \ell_n^a(y)\, g_a(x)\, g_a(y)$$

où $\varrho\ (h)$ est la fonction de corrélation du processus qui admet une décomposition isofactorielle pour ses lois bivariables.

On sait donc que l'usage d'un modèle isofactoriel gamma est légitime au moins à une dimension quand la fonction de covariance est exponentielle ! En effet, on ne sait pas construire explicitement une fonction aléatoire à n dimensions avec des lois bivariables gamma mais nous avons utilisé ce modèle à deux dimensions sans remarquer de problèmes particuliers. De plus il est toujours possible d'obtenir un modèle isofactoriel à n dimensions de loi marginale gamma si l'on randomise le processus de diffusion $Y(x)$ par un FAI-0 $Z(x)$ [7]. Le nouveau processus

$$\tilde{Y}(x) = Y(Z(x))$$

a des lois marginales gamma et des lois bivariable de Laguerre que nous définissons maintenant.

Comme dans le cas gaussien, on peut généraliser les lois isofactorielles ayant pour loi marginale un loi gamma. La forme générale est

$$g^Z(dx, dy) = \sum \varrho_n\, \ell_n^a(x)\, \ell_n^a(y)\, g_a(x)\, g_a(y)$$

où les ϱ_n sont des moments d'une variable aléatoire avec pour loi $\omega(dx)$ sur [0,1]

$$\varrho_n = \mathrm{E}(R^n) = \int_0^1 r^n \, \omega(dr) .$$

Quand la loi bivariable prend cette forme, on parle d'une loi de Laguerre. Si $\omega(dx)$ est concentré à 0 et 1, on trouve un processus mosaïque, et si on prend pour $\omega(dx)$ une loi Beta de paramètres $(\varrho a, (1-\varrho)a)$ où ϱ est la corrélation et a le paramètre de la loi Gamma, alors la loi disjonctive correspond à celle d'un processus généré par la régularisation d'une mesure aléatoire orthogonale de loi Gamma de paramètre a.

Nous allons maintenant considérer un changement de support ([14]). Si les données randomisées $Z(x)$ ont des lois bivariables de Laguerre, on s'attend à ce que la loi de Y, si elle reste gamma, soit moins dissymétrique que celle de Z – son indice a' sera plus grand que a.

Si $$g_h^Z(x,y) = \sum \varrho_n(h) \, \ell_n^a(x) \, \ell_n^a(y) \, g_a(x) \, g_a(y)$$

et si $$Y(x) = Z(x) + \mathrm{U}$$

où U suit une loi gamma de paramètre $a' - a$, $Y(x)$ suit alors une loi gamma de paramètre a' et on peut démontrer que les lois bivariables (Y,Z) et (Z,Z) restent isofactorielles et prennent les formes

$$g_h^{Y,Z}(x,y) = \sum_n \varrho_n(h) \, C_n(a,a') \, \ell_n^a(x) \, \ell_n^{a'}(y) \, g_a(x) \, g_{a'}(y)$$

$$g_h^{Y}(x,y) = \sum_n \varrho_n(h) \, C_n^2(a,a') \, \ell_n^{a'}(x) \, \ell_n^{a'}(y) \, g_{a'}(x) \, g_{a'}(y)$$

avec $C_n(a,a') = \sqrt{\dfrac{\Gamma(a+n)\,\Gamma(a')}{\Gamma(a'+n)\,\Gamma(a)}}$.

Pour ce modèle, la loi de randomisation $Z/Y = y$ est donnée par $(Z|Y=y) \backsim yV$ où V est une loi beta de paramètres $(a, a'-a)$.

6.3.3. — *Modèle isofactoriel binomial discret*

On dit que $Y(x)$ suit une loi binomiale négative de paramètres (ν,p) si $Z(x)$ est discrète et

$$p_i = \mathrm{P}[Y(x) = i] = q^{\nu}\frac{\Gamma(\nu + i)}{\Gamma(\nu)\, i!}p^i$$

et $p + q = 1$.

On peut construire une fonction aléatoire isofactorielle avec des marginales binomiales négatives comme dans le cas gamma sauf qu'ici on utilise un processus de naissance et de mort avec un générateur infinitésimal

$$(Af)_i = -\big(p(\nu + i) + i\big)\, f_i + p(\nu + i)\, f_{i+1} + \mathrm{if}_{i-1}$$

et on peut démontrer que les lois bivariables sont

$$\mathrm{P}_h^Y(i,j) = \sum_{n=0}^{\infty} \varrho^n(h)\, \chi_n^{p,\nu}(i)\, \chi_n^{p,\nu}(j)\, p_i\, p_j$$

où les $\chi_n^{p,\nu}(i)$ sont des polynômes de Meixner normés. Les polynômes non-normés $H_n(i)$ sont liés aux polynômes normés par :

$$\chi_n(i) = C_n^{\nu p}\, H_n(i)$$

avec $C_n^{\nu p} = \sqrt{\dfrac{\Gamma(n + \nu)p^n}{\Gamma(\nu)n!}}$.

Les polynômes sont définis par la formule de Rodrigues

$$\chi_n^{p,\nu}(i) = \frac{1}{n!\; C_n^{\nu} p}\,\frac{1}{p_\nu(i)}\nabla^n\, p_{\nu,n}(i)$$

où $p_{\nu,n}(i) = p_\nu(i + n)\dfrac{(i + n)!}{i!}$ et $\nabla\, p_{\nu,n}(i) = p_{\nu,n}(i) - p_{\nu,n}(i-1)$.

On les calcule facilement à l'aide des relations de récurrence

$$H_0(i) = 1 \qquad H_1(i) = \frac{p\nu - iq}{p\nu}$$

$$p(\nu + n)\, H_{n+1}(i) = [\nu p + n(p+1) - iq]\, H_n(i) - n\, H_{n-1}(i)$$

Comme dans les cas précédents, on peut généraliser ce modèle isofactoriel par

$$\mathrm{P}_h^Y(i,j) = \sum \varrho_n(h)\, \chi_n^{p,\nu}(i)\, \chi_n^{p,\nu}\, p_i\, p_j$$

où les $\varrho_n(h)$ sont des moments d'une variable aléatoire avec la loi $\omega(dx)$ sur 0,1

$$\varrho_n = \int_0^1 \varrho^n\, \omega(d\varrho)$$

Comme dans le cas gamma, on peut obtenir le modèle dissymétrique si l'on considère $Y(x) = Z(x) + N$ avec N binomial négatif de même paramètre ϱ, mais avec a' au lieu de a et $a' - a > 0$. En effet, pour la géostatistique minière, c'est cette méthode qui donne le modèle le plus intéressant pour le problème du changement de support.

Il existe un autre type de randomisation pour lequel on commence avec la variable sous-jacente $Y(x)$, et on fait une décimation de ce processus pour obtenir $Z(x)$. On peut démontrer que ce nouveau processus reste binomial négatif et isofactoriel et la loi dissymétrique est

$$\mathrm{P}_h^{Y,Z}(i,j) = \sum \varrho_n(h) \left(\frac{p'}{p}\right)^{n/2} \chi_n^{p,\nu}(i)\, \chi_n^{p',\nu}(j)\, p_i\, p'_j$$

où p'_j est la loi marginale pour le processus décimé de paramètre (a,p') et où $p' < p$.

La loi bivariable pour le processus randomisé est

$$\mathrm{P}_h^Z(i,j) = \sum \varrho_n(h) \left(\frac{p'}{p}\right)^{n} \chi_n^{p',\nu}(i)\, \chi_{n'}^{p',\nu}(j)\, p'_i\, p'_j$$

La loi conditionnelle de $Z(x)|Y(x)$ est binomiale, $\mathrm{B}\left(Y(x), \frac{p'q}{pq'}\right)$.

6.3.4. — Un modèle mixte

Commençons avec les modèles isofactoriels négatifs binomiaux de *6.3.3.* avec le changement de support provenant d'une décimation. Puisque p et p' sont

différents pour les lois sous-jacentes et et les lois randomisées, on peut faire varier le premier sans affecter le second. En particulier on fait tendre $q = 1-p$ vers zéro, et alors Nq a comme limite une variable aléatoire gamma de paramètre ν puisque nos variables sont binomiales négatives (p, ν). La loi isofactorielle tend vers une loi de Laguerre isofactorielle

$$g_h^Y(x,y) = \sum \varrho_n(h)\, \ell_n^\nu(x)\, \ell_n^\nu(y)\, g_\nu(x)\, g_\nu(y)$$

et la loi dyssimétrique tend vers un modèle mixte discret/continu

$$\mathrm{P}^{Y,Z}(y,i) = \sum \varrho_n(h)\, p'^{\frac{n}{2}}\, \ell_n^\nu(x)\, \chi_n^{p',\nu}(i)\, g_\nu(x)\, p'_i$$

et la loi de la variable randomisée n'est évidemment pas changée

$$\mathrm{P}_h^Z(i,j) = \sum_n \varrho_n(h)\, p'^n\, \chi_n^{p'\nu}(i)\, \chi_n^{p'\nu}(j)\, p'_i\, p'_j$$

Pour ce modèle, il se trouve que la loi conditionnelle $Z(x)|Y(x) = y$ est une loi de Poisson de paramètre $(\frac{p'}{q'}y)$, et présente donc un intérêt particulier pour le filtrage des données poissonisées.

6.4. — *MISE EN OEUVRE D'UN KRIGEAGE DISJONCTIF*

L'histogramme d'une image risque fortement de ne pas correspondre directement à une loi pour laquelle un modèle isofactoriel est connu (gaussien, gamma, binomial négatif, etc.). Il faut donc trouver une transformation des données dans en une loi convenable et vérifier que les lois bivariables correspondent au modèle isofactoriel. Cette procédure est bien détaillée dans la littérature ([14]), et on n'en donnera ici qu'un bref rappel. Nous passerons ensuite à une modélisation de la loi dissymétrique (Y,Z) et aux équations du krigeage disjonctif pour le filtrage de Y et l'estimation de l'histogramme local.

6.4.1. — *L'anamorphose empirique*

Supposons pour l'instant qu'on veuille transformer les données en une loi $G(dx)$ avec une densité continue (par exemple gaussienne, gamma). Si notre loi observée a n valeurs distinctes, elle est de la forme

$$F^*(dz) = \sum_{i=1}^{N} p_i \, \delta_{z_i}(dz) .$$

Si maintenant $y_i = G^{-1}\left(\sum_{j=1}^{i} p_j \right)$, on peut considérer la fonction $\phi(y)$

$$\phi(y) = \Sigma \, z_i \, 1_{[y_{i-1}, y_i[}(y)$$

et ϕ est la transformation désirée, appelée anamorphose empirique. Elle n'est malheureusement pas bijective et on est obligé de choisir un pseudo-inverse que nous avons pris comme

$$\varphi^{-1}(z_i) = \frac{1}{p_i} \int_{y_{i-1}}^{y_i} y \; G(dy)$$

Pour plus de détail, on se reportera à [14], [15].

Pour la mise en oeuvre du krigeage disjonctif, il est utile d'avoir un développement de la fonction d'anamorphose en termes de polynômes orthogonaux pour $G(dx)$, soit $\chi_n(x)$.

$$\phi(y) = \sum \phi_n \, \chi_n(x)$$

et

$$\phi_n = \sum_{i=1}^{N} z_i \int_{y_{i-1}}^{y_i} \chi_n(y) \; F(dy)$$

Les intégrales (ou les sommes dans le cas fini) se calculent facilement à l'aide de la formule de Rodrigues pour χ_n.

Si la fonction de répartition $G(dx)$ est discrète, ni la fonction d'anamorphose ni son inverse ne sont définies univoquement. Une procédure a été développée dans [16]. Elle donne des resultats à peu près corrects si la loi des données observées et la loi qu'on veut obtenir après l'anamorphose ont un nombre élevé de classes de probabilités non-nulle, et s'il n'y a pas de classe avec une très grande probabilité d'occurrence. Si par contre il n'y a qu'un nombre relativement petit de classes pour une des deux lois, l'anamorphose risque d'être très peu efficace. Par conséquent, dans ce cas, l'ajustement d'une loi isofactorielle doit se faire directement sur les données sans anamorphose.

Lorsqu'on a trouvé une anamorphose convenable, il faut vérifier que la loi bivariable des données transformées est correctement modélisée par la loi

isofactorielle. Pour faire ceci correctement, il faut étudier l'espérance conditionnelle des facteurs $E[\chi_n(Z_\alpha)|Z_\beta]$ mais notons qu'il existe des tests qui utilisent le variogramme d'ordre 1 et qui, bien que non exhaustifs, permettent cependant de vérifier partiellement l'hypothèse ([5], [14]).

Notons ici que l'usage des modèles isofactoriels avec une anamorphose n'est pas tout à fait satisfaisant pour le problème de randomisation. Par exemple, si nous observons des données avec un bruit poissonien, nous voudrons modéliser la loi de randomisation sur les données avant d'effectuer l'anamorphose. Néanmoins nous trouvons dans les études numériques de §6.6 qu'une modélisation en termes de lois isofactorielles peut donner de bons résultats si l'on utilise la loi sous-jacente obtenue par la méthode de régularisation du chapitre III. Il s'agit d'un compromis pratique qui semble assez robuste pour l'application d'un krigeage disjonctif mais un peu moins robuste pour l'espérance conditionnelle.

6.4.2. — Ajustement des lois dissymétriques

Après avoir ajusté un modèle isofactoriel aux données observées (éventuellement après une anamorphose), la prochaine étape consistera à obtenir un modèle pour la loi dissymétrique. Soit $\tilde{Z}(x)$ les données observées et $Z(x)$ les données anamorphosées. Si l'on prend une anamorphose gaussienne $\tilde{Z} = \varphi(Z)$, si la loi bivariable $(Z(x), Z(x+h))$ s'écrit

$$g_h^Z(x,y) = \sum_n \varrho_Z^n(h)\ \eta_n(x)\ \eta_n(y)\ g(x)\ g(y)$$

et si on modélise la loi sous-jacente $Y(x)$ selon la randomisation

$$Z(x) = rY(x) + \sqrt{1-r^2}\ \epsilon(x)$$

il nous faut une estimation de r. On peut considérer deux méthodes pour l'obtenir.

1. — Si l'on suppose que $Y(x)$ est une fonction aléatoire continue, elle n'aura pas d'effet de pépite, et celui des données anamorphosées doit être $1 - r^2$

$$(\text{pépite})_Z = 1 - r^2$$

d'où on obtient un estimateur de r. Plus généralement, si nous avons deux acquisitions $\tilde{Z}_1(x)$ et $\tilde{Z}_2(x)$, après anamorphose des deux acquisitions, on a

$$Z_1(x) = r\ Y(x) + \sqrt{1-r^2}\ \epsilon_1(x)$$

$$Z_2(x) = r\ Y(x) + \sqrt{1-r^2}\ \epsilon_2(x)$$

et le variogramme croisé entre $Z_1(x)$ et $Z_2(x)$ aura un palier à r^2

$$\text{Cov}\,[Z_1(x), Z_2(x)] = r^2\,.$$

2. — La deuxième méthode est plus couramment utilisée en géostatistique minière. On a

$$\bar{Z}(x) = \varphi(Z(x)) = \sum_0^\infty \varphi_n\ \eta_n(Z(x))$$

Soit $\psi(Y) = \bar{Y}$ une anamorphose pour la fonction aléatoire sous-jacente. $\psi(Y) = \Sigma\ \psi_n\ \eta_n(Y)$.

Si l'on applique la condition de Cartier

$$\text{E}\,[\bar{Z}\,|\,\bar{Y}] = \bar{Y}$$

et le fait que

$$\text{E}\,[\eta_n(Z(x))\,|\,Y(x)] = r^n\ \eta_n(Y(x))\,,$$

on trouve

$$\bar{Y}(x) = \psi(Y(x)) = \Sigma\ r^n\ \varphi_n\ \eta_n(Y)$$

(d'où $\psi_n = r^n\ \varphi_n$). En particulier, on trouve

$$\text{Var}\ \bar{Y} = \text{Cov}(\bar{Z}(x), \bar{Y}(x)) = \sum_{n=1}^\infty r^n\ \varphi_n\psi_n\,.$$

Puisqu'on connaît la variance des données réelles $\bar{Y}(x)$, on peut trouver r.

Si le modèle isofactoriel est adapté les deux méthodes doivent donner des estimateurs de r comparables. Si ces derniers ne concordent pas, on choisit la deuxième méthode qui a le mérite d'utiliser l'estimateur de la loi sous-jacente, mais il faut remarquer que le désaccord des estimateurs peut indiquer un problème dans le choix du modèle isofactoriel.

Pour les autres types de modèles isofactoriels, on aura essentiellement les mêmes types de résultats pour estimer les paramètres. Un cas particulièrement

intéressant est le modèle mixte gamma/binomial négatif. Supposons la loi de $\bar{Z}(x)$ de moyenne m et de variance σ_Z^2. Pour une variable de loi binomiale négative de paramètres (p, δ) on doit trouver une moyenne de $\frac{\delta p}{q}$ et une variance de $\frac{\delta p}{q^2}$. Si l'on compare avec les données observées on trouve que

$$p = \frac{\sigma_Z^2 - m}{\sigma_Z^2} \qquad \text{et} \qquad \delta = \frac{m^2}{\sigma_Z^2 - m}.$$

On peut donc considérer une anamorphose de notre variable observée en un binomial négatif de paramètres $\left(\frac{\sigma_Z^2 - m}{\sigma_Z^2} , \frac{m^2}{\sigma_Z^2 - m} \right)$.

Si l'anamorphose directe est faisable et ne déforme pas trop la loi, cette modélisation donne une bonne approximation à une randomisation par une loi de Poisson.

6.4.3. — Filtrage et histogramme local

Dans le cas du filtrage de $\bar{Y}$, où $\bar{Y} = \psi(Y)$ représente la fonction aléatoire sous-jacente, il existe deux méthodes pour choisir la fonction ψ.

1. — Nous avons déjà vu que la condition de Cartier peut nous donner l'anamorphose de Y en termes d'anamorphose de Z,

$$\bar{Y} = \psi(Y) = \sum r^n \varphi_n \eta_n(Y).$$

2. — Puisque les méthodes décrites au Chapitre V nous donnent un estimateur direct de la loi $F^Y(y)$, on peut ajuster une fonction d'anamorphose sur l'estimation $\bar{Y} = \bar{\psi}(Y) = \Sigma \bar{\psi} \eta_n(Y(x))$ et faire ensuite le krigeage disjonctif de cette fonction $\bar{\psi}(Y)$.

Nous utilisons la deuxième méthode qui dépend explicitement de la loi sous-jacente au lieu de la première méthode qui dépend plus du modèle.

Si l'on veut estimer l'histogramme local

$$\mathrm{P}\left[\tilde{Y}(x) < y \,\middle|\, x \in v\right] = \frac{1}{v}\int_v 1_{\left\{\tilde{Y}(x)<y\right\}} dx$$

$$= \frac{1}{v}\int_v 1_{\{Y(x)<\psi^{-1}(y)\}}\, dx$$

où ψ est l'anamorphose choisie pour Y. Par linéarité,

$$\left[\frac{1}{v}\int_v 1_{\{Y(x)<\psi^{-1}(y)\}} dx\right]^* = \frac{1}{v}\int_v 1^*_{\{Y(x)<\psi^{-1}(y)\}}\, dx$$

On peut donc estimer l'histogramme local soit directement, soit le calculer avec une moyenne mobile des estimateurs de $1_{\{Y(x)<\psi^{-1}(y)\}}$. Si les données ne sont pas sur une grille régulière, la première méthode sera en général préférable, mais pour les images, la deuxième méthode avec estimation de l'indicatrice à chaque pixel et ensuite de la moyenne mobile est une approximation acceptable.

L'estimation de l'indicatrice est simple à partir de son développement en termes de polygones orthogonaux

$$1_{\{Y(x)<\psi^{-1}(y)\}} = \Sigma\, \varphi_n\, \chi_n(Y(x))$$

où les $\varphi_n = \int_0^{\psi^{-1}(y)} \chi_n(x)\, dF^Y(x)$ se calculent facilement à l'aide de la formule de Rodrigues.

6.5. — L'ESPÉRANCE CONDITIONNELLE

Après l'anamorphose gaussienne de $\tilde{Z}$ et $\tilde{Y}$, si les fonctions aléatoires transformées $Z(x)$, $Y(x)$ sont multigaussiennes (au lieu de bigaussiennes) et si

$$Z(x) = r\, Y(x) + \sqrt{1-r^2}\, \epsilon(x)$$

où $\epsilon\,(x)$ est également multigaussien, on peut obtenir un estimateur plus fort que le krigeage disjonctif, à savoir l'espérance conditionnelle. Notons que ceci peut

être mis en oeuvre facilement pour le cas gaussien uniquement (en effet, pour les fonctions aléatoires gaussiennes, l'espérance conditionnelle et le krigeage linéaire avec moyenne connue coïncident), mais que ce n'est pas vrai pour les autres lois utilisées en pratique. Notons qu'une méthode pour obtenir l'espérance conditionnelle dans le cadre d'un modèle de champ markovien a été récemment proposée [18]. Elle utilise le "sampler" de Gibbs pour obtenir des simulations conditionnelles de l'image sous-jacente. La moyenne des ces simulations est donc un estimateur de l'espérance conditionnelle. Cette méthode est certainement d'une grande intérêt, mais semble malheureusement être beaucoup trop coûteuse actuellement pour être utilisée dans les applications qui nous concernent.

Nous avons vu que $Y(x)$ connaisssant $Z_1,\ldots,Z_n$ est une variable aléatoire gaussienne de moyenne $\Sigma\ \lambda_i Z_i$ et de variance $\sigma_E^2 = \mathrm{Var}(\Sigma\ \lambda_i Z_i)$ où les $\{\lambda_i\}$ sont les solutions des équations du krigeage simple de $Y(x)$.

On peut maintenant estimer n'importe quelle fonction $f(Y(x))$, et particulièrement $\tilde{Y} = \varphi(Y)$ ou $P[\tilde{Y}(x) < y | x \in v]$ par le calcul de son espérance conditionnelle selon la loi $(Y(x)|Z_1,\ldots,Z_n)$.

Ceci peut se faire directement ou en termes de polynômes d'Hermite de la manière suivante :

$$f(Y(x)) = \sum f_n\ \eta_n(Y(x))\,.$$

Si on utilise les quatre résultats simples suivants :

1) $$\mathrm{Cov}\left[Y(x),\ \frac{\Sigma\ \lambda_i Z_i}{\sigma_E}\right] = \frac{\mathrm{Var}\ (\Sigma\ \lambda_i Z_i)}{\sigma_E} = \sigma_E$$

2) $\left(\dfrac{\Sigma\ \lambda_i Z_i}{\sigma_E},\ Y(x)\right)$ est une bigaussienne réduite de corrélation σ_E

3) $$E\,[Y(x)|Z_1,\ldots,Z_n] = E\left[Y(x)\ \middle|\ \frac{\Sigma\ \lambda_i Z_i}{\sigma_E}\right]$$

4) $$E\left[\eta_n(Y(x))\ \middle|\ \frac{\Sigma\ \lambda_i Z_i}{\sigma_E}\right] = \sigma_E^n\ \eta_n\left(\frac{\Sigma\ \lambda_i Z_i}{\sigma_E}\right),$$

on obtient

$$\mathrm{E}\left[f(Y(x))\,|\,Z_1,\ldots,Z_n\right] = \sum \sigma_{\mathrm{E}}^n\, f_n\, \eta_n\left(\frac{\sum \lambda_i Z_i}{\sigma_{\mathrm{E}}}\right)$$

et la variance d'estimation est donnée par

$$\mathrm{Var}\left[f(Y(x)) - f^*(Y(x))\right] = \sum \left(1 - \sigma_{\mathrm{E}}^{2n}\right) f_n^2 .$$

L'espérance conditionnelle a deux avantages par rapport au krigeage disjonctif. Si on veut estimer l'histogramme local $\mathrm{P}[\tilde{Y} < y | x \in v]$, on est assuré d'obtenir une loi de probabilité, autrement dit des estimations entre 0 et 1. Avec le krigeage disjonctif, on n'a aucune garantie de cette contrainte et on peut avoir des valeurs légèrement plus faibles que 0 ou plus grandes que 1. Deuxièmement, quand les données sont sur une maille régulière, l'espérance conditionnelle est plus rapide à calculer que le krigeage disjonctif.

En revanche, l'espérance conditionnelle dépend de la loi multivariable de la fonction aléatoire, et on ne peut pas connaître cette loi dans la pratique. Dans beaucoup de cas (par exemple quand la loi a un atome à l'origine), une anamorphose gaussienne n'est pas du tout adaptée et il nous faudra des modèles autres que gaussiens pours lesquels on ne sait pas appliquer l'espérance conditionnelle. Si l'on utilise les équations du krigeage disjonctif directement sans passer par un modèle isofactoriel, on peut avoir des estimateurs cohérents avec la randomisation, ce qui n'est généralement pas possible pour l'espérance conditionnelle. Enfin, si l'on veut estimer les histogrammes locaux et que les données ne sont pas sur une maille régulière, le krigeage disjonctif permet de faire l'estimation directement, mais avec l'espérance conditionnelle, il faudra discrétiser chaque bloc qui sera ensuite estimé avec cette discrétisation. Le krigeage disjonctif est donc plus efficace dans ce cas.

6.6. — EXEMPLES

1. — Commençons avec l'image simulée qui est montrée à la figure 6.9. Elle a une moyenne de 200 et une variance de 143. Le variogramme est sphérique avec une portée de 5 pixels. L'image a été randomisée par un bruit poissonien (fig. 6.10). Le filtrage linéaire de cette image est montré à la figure 6.11. La variance d'estimation est de 44.23 et l'erreur vraie à pour variance 44.67.

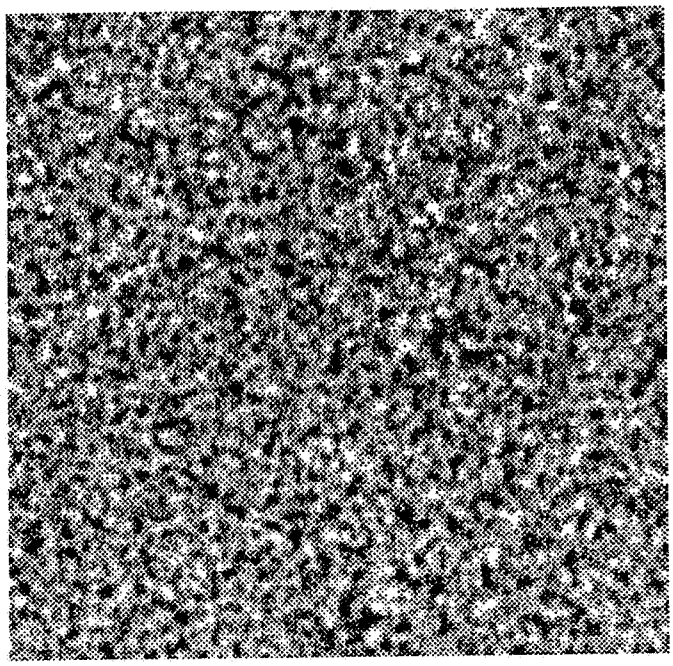


Figure 6.9 Image réelle

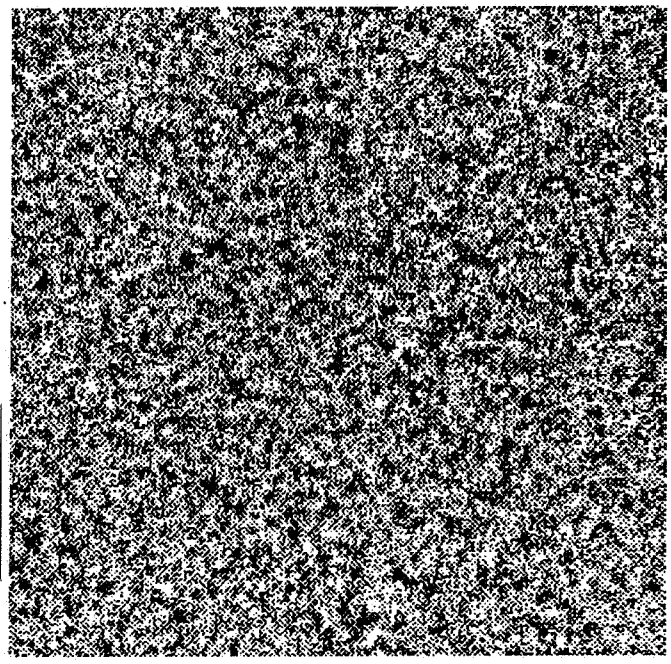


Figure 6.10 Image bruitée

Calculons maintenant un estimateur disjonctif gaussien. Tout d'abord il faut estimer la loi sous-jacente de $\tilde{Y}$ à partir de celle de $\tilde{Z}$ en utilisant les méthodes développées au chapitre III. On trouve ensuite une fonction d'anamorphose ψ telle que $\tilde{Y} = \psi(Y)$ et Y est gaussien réduit. Pour $\tilde{Z}$, on trouve aussi une fonction d'anamorphose φ telle que $\tilde{Z} = \varphi(Y)$ et Z est gaussien réduit. Le coefficient de corrélation ϱ est estimé en utilisant l'une des deux méthodes décrites à la section 6.4.2, qui sont basées sur l'effet de pépite du variogramme sous-jacent et sur la covariance croisée entre $\tilde{Z}(x)$ et $\tilde{Y}(x)$. Les résultats sont de 0.6461 et 0.6459 et ils sont donc très proches dans ce cas. Le krigeage disjonctif peut être effectué et le résultat est présenté sur la figure 6.12. La variance d'estimation est de 44.17 et la variance de l'erreur réelle de 44.516. Le résultat est donc comparable au krigeage ordinaire, ce qui était prévisible, les données sous-jacentes étant gaussiennes.

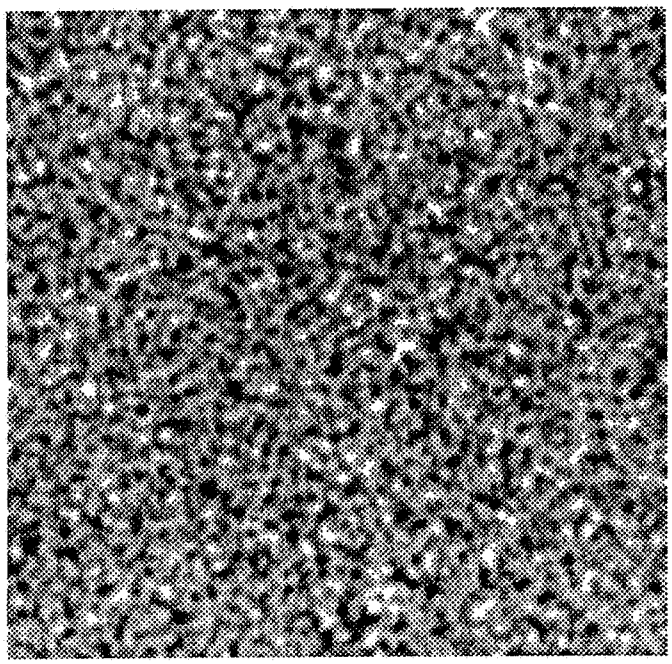


Figure 6.11 Krigeage linéaire

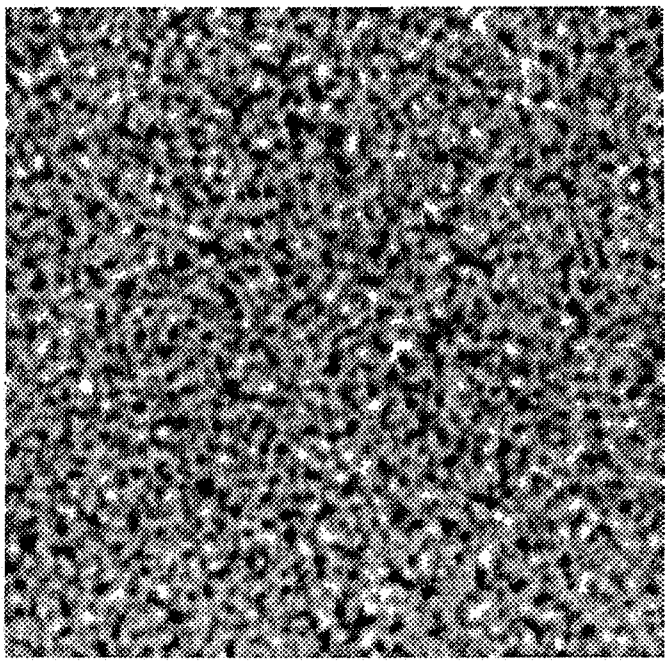


Figure 6.12 Krigeage disjonctif — Gaussien

Il est intéressant d'essayer ici le modèle mixte binomial négatif/gamma. On devrait pouvoir obtenir un bon ajustement des modèles binomial négatif et gamma directement sur les histogrammes, mais les paramètres sont trop grands pour qu'on puisse avoir des calculs stables par la suite. Il fallait donc choisir un paramètre pour le modèle gamma qui soit grand (pour que la loi ressemble à une loi de Gauss) mais pas trop (pour éviter des instabilités numériques). On a choisi $a = 100$, et l'anamorphose a été faite sur l'histogramme sous-jacent estimé par régularisation. Pour les données observées, on a fait une pseudo-anamorphose dans une loi binomiale négative de paramètre (100, p) où on a choisi p tel que $\text{Cov}\,(Z(x), Y(x)) = \Sigma\, p^{n/2}\, \chi_n^{p,\nu}(i)\, g_\nu(x)$. Celle-ci doit se faire par itération. On choisit une valeur initiale de p, on fait la pseudo-anamorphose et on obtient une nouvelle version de p avec la formule de la covariance croisée, et ainsi de suite. Normalement une itération suffit pour obtenir un p stable parce que la forme de l'anamorphose ne dépend pas fortement de p s'il n'y a pas d'atome dans la fonction de répartition (un phénomène similaire a été observé dans le cas du krigeage disjonctif gamma, [14]). Les résultats du filtrage binomial négatif/gamma sont montrés à la figure 6.13. La variance d'erreur estimée est de 44.51 et l'erreur réelle est de 45.113.

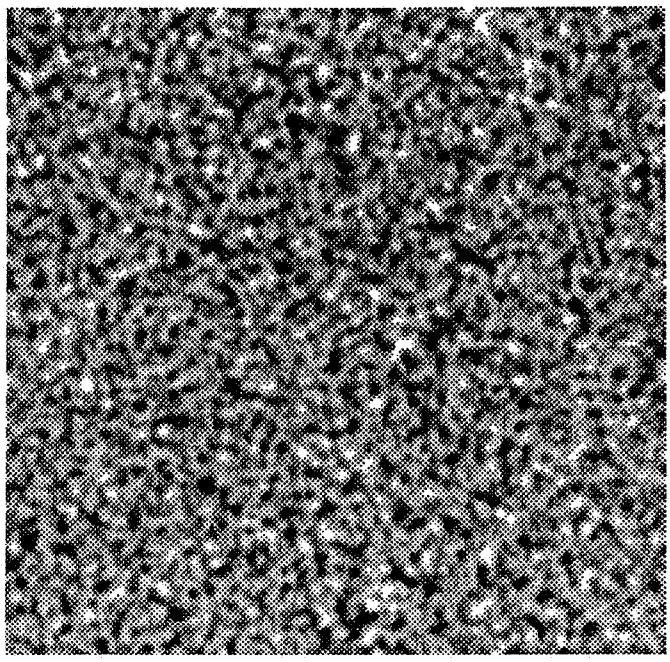


Figure 6.13 Krigeage disjonctif — Binomial négatif / Gamma

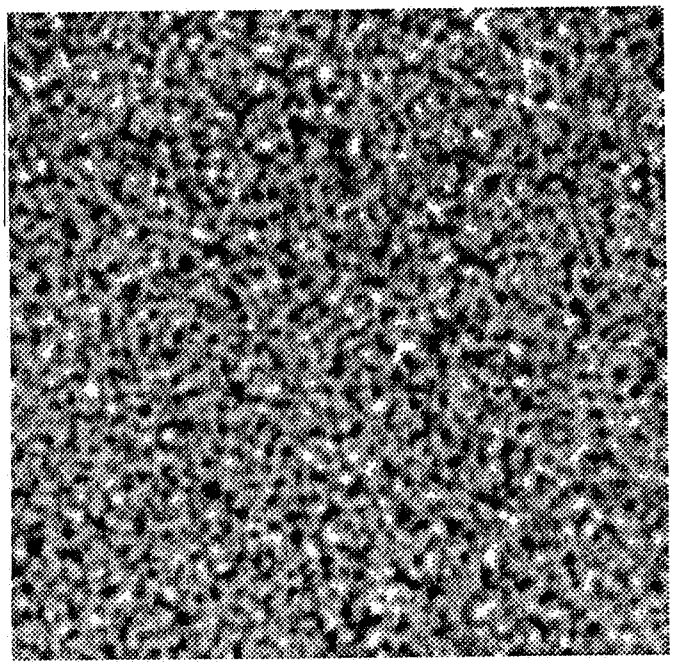


Figure 6.14 Espérance conditionnelle

On construit également un estimateur par espérance conditionnelle qui est présenté à la figure 6.14. (Bien entendu, il ne s'agit pas d'une véritable espérance conditionnelle prenant en compte la forme poissonienne du bruit, mais c'est l'estimateur qui traite Y et Z comme des anamorphosées d'une loi gaussienne, voir § 6.5). Ici l'erreur d'estimation vraie est de 44.526.

On note que tous les estimateurs sont remarquablement proches, ce qui était presque obligatoire par construction.

Nous allons maintenant calculer un seuillage de l'histogramme local. Fixons la valeur à 220, ce qui correspond à un écart par rapport à la moyenne de 1.7, et essayons de calculer $P[Y(\underline{x}) \geq 220 | \underline{x} \in V_y]$, où V_y est un voisinage centré sur y. On estime cette probabilité pour chaque point y dans l'image et on obtient ainsi une carte de probabilité qui peut être elle-même montrée sous forme d'image. Nous avons calculé cette image pour les deux estimateurs de type disjonctif et pour l'espérance conditionnelle en utilisant les formules développées aux sections 6.4.2 et 6.5. On les compare à la carte de probabilité réelle ainsi qu'à un estimateur simple obtenu à partir de l'image filtrée par krigeage ordinaire en appliquant un seuillage à 220 et en appliquant ensuite une moyenne mobile.

Les résultats sont présentés aux figures 6.15 à 6.19 pour un voisinage V_y de 5×5 et pour les cas suivants : réel, krigeage ordinaire, krigeage disjonctif gaussien, krigeage disjonctif binomial négatif et espérance conditionnelle respectivement. Les résultats numériques sont résumés au Tableau 6.3.

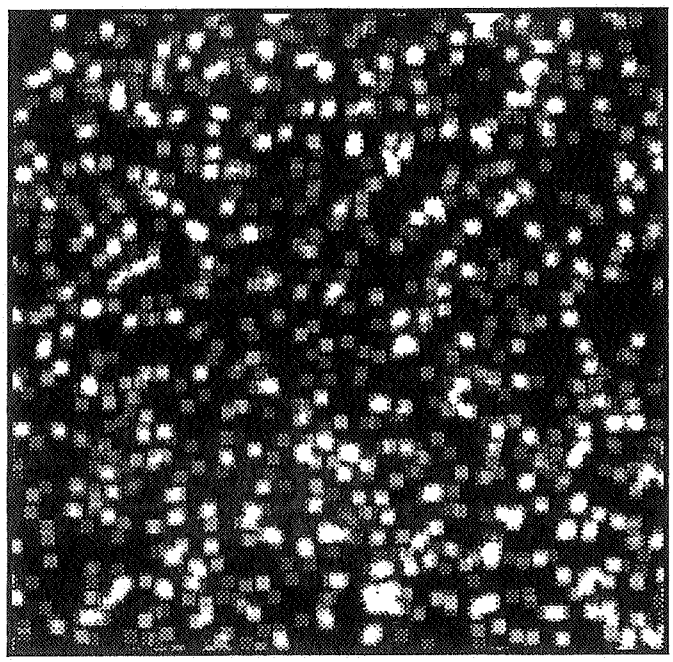


Figure 6.15 Histogramme local à 220. Voisinage = 5 x 5.

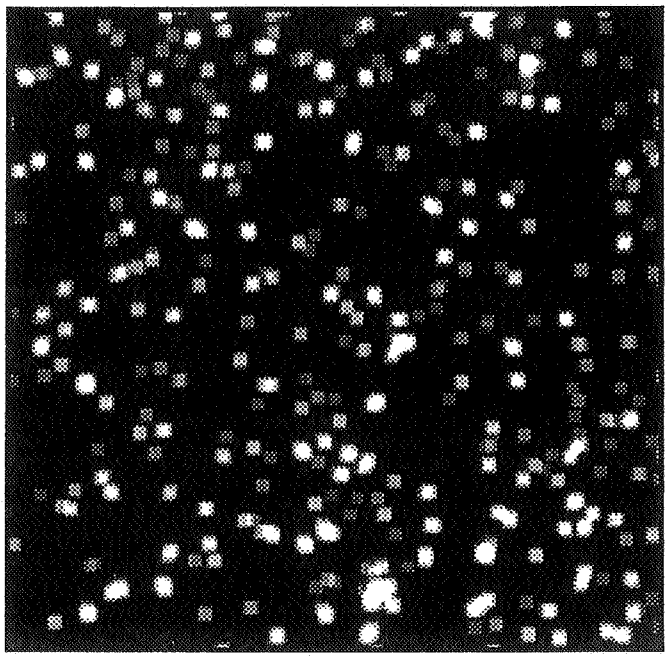


Figure 6.16 Histogramme local, krigeage linéaire.

On voit que le krigeage ordinaire est fortement biaisé et présente l'erreur d'estimation la plus importante. Les deux estimateurs du krigeage disjonctif sont moins biaisés, particulièrement dans le cas gaussien, et ils ont une variance d'erreur plus faible. L'espérance conditionnelle donne les meilleurs résultats, mais on note que le biais conditionnel est plus important que pour le krigeage disjonctif.

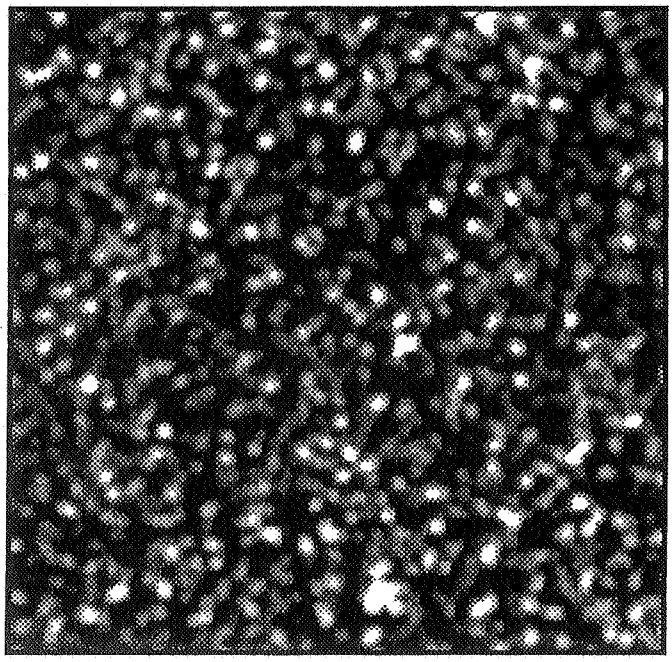


Figure 6.17 Histogramme local, krigeage disjonctif — Gaussien.

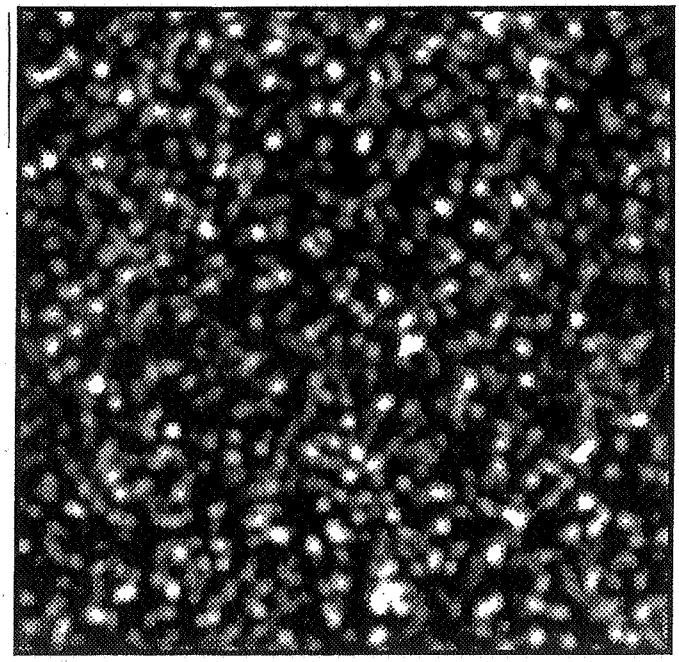


Figure 6.18 Histogramme local, krigeage disjonctif — Binomial négatif / Gamma

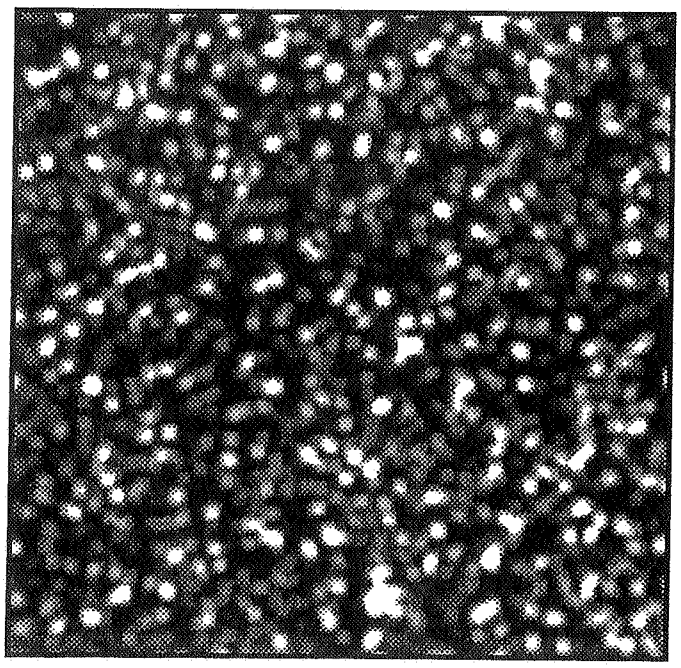


Figure 6.19 Histogramme local, espérance conditionnelle.

Nous avons calculé l'histogramme local pour la même valeur de 220, mais avec un voisinage de 11 × 11. Le krigeage disjonctif est maintenant presque aussi bon que l'espérance conditionnelle et deux fois meilleur que l'estimateur du krigeage ordinaire. Les résultats sont résumés au Tableau 6.3. Pour les voisinages de 5 × 5 ou 11 × 11, on présente deux "distances" entre la probabilité locale réelle et chaque estimateur, c'est-à-dire le biais ou erreur moyenne et la variance d'erreur. Les figures sont présentées de 6.20 à 6.24 respectivement.

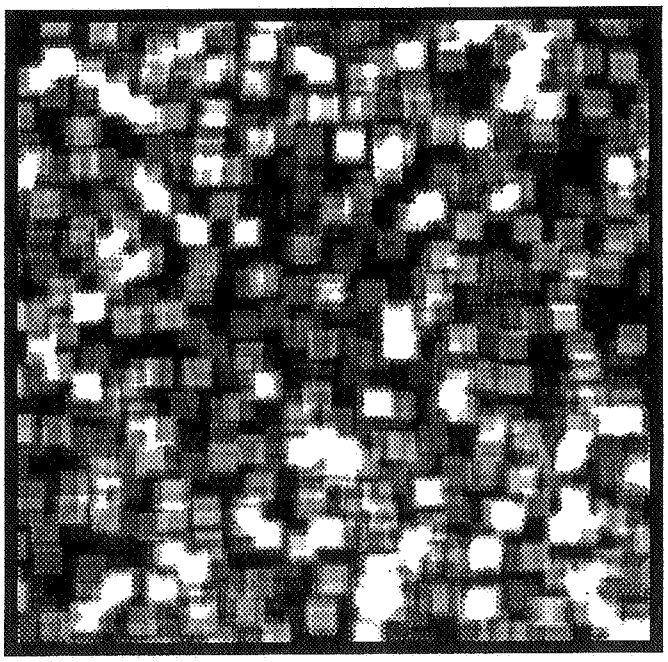


Figure 6.20 Histogramme local à 220. Voisinage = 11 x 11.

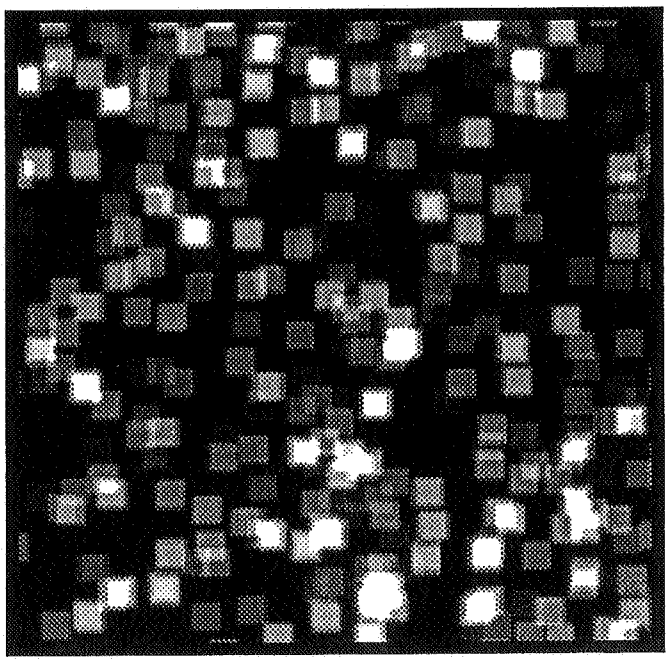


Figure 6.21 Histogramme local, krigeage linéaire.

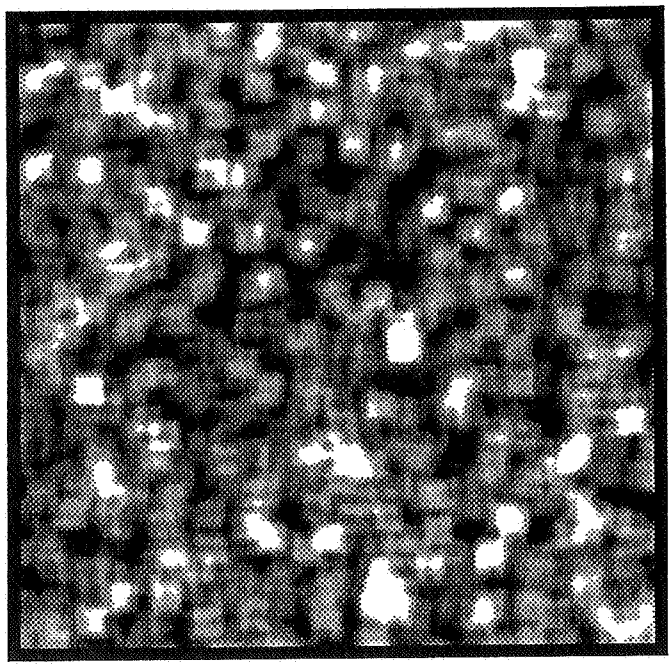


Figure 6.22 Histogramme local, krigeage disjonctif — Gaussien.

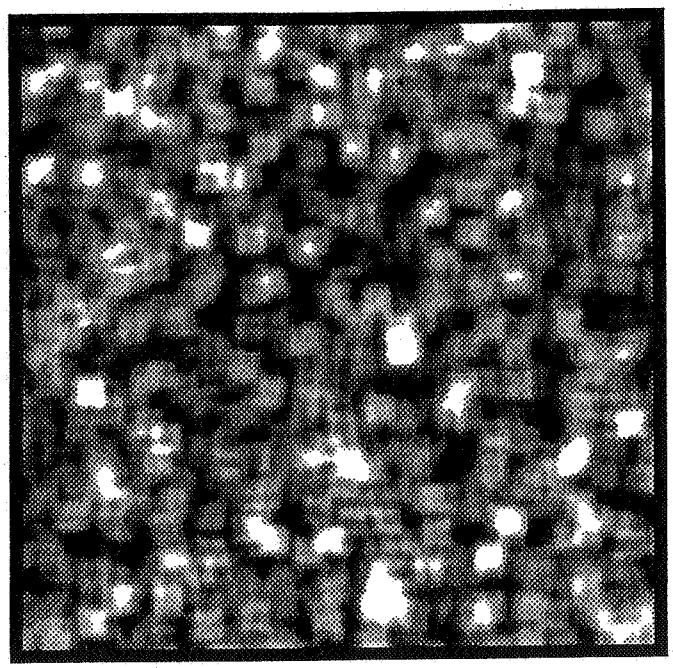


Figure 6.23 Histogramme local, krigeage disjonctif — Binomial négatif / Gamma

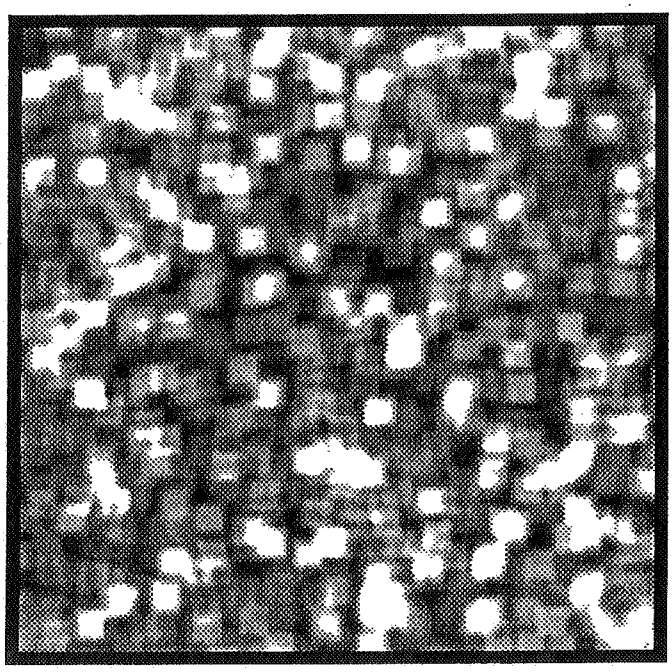


Figure 6.24 Histogramme local, espérance conditionnelle.

Tableau 6.3

	5 x 5		11 x 11	
Variance réelle	7.965×10^{-3}		2.147×10^{-3}	
	Erreur moyenne	Variance d'erreur	Erreur moyenne	Variance d'erreur
d (réel, krigeage linéaire)	0.0254	3.966×10^{-3}	0.0254	1.415×10^{-3}
d (réel, KD gaussien)	0.0048	3.156×10^{-3}	0.0047	0.726×10^{-3}
d (réel, KD binomial négatif)	0.0085	3.280×10^{-3}	0.0086	0.791×10^{-3}
d (réel, espérance conditionnelle)	-0.0114	2.368×10^{-3}	-0.0116	0.632×10^{-3}

2. — Dans notre deuxième exemple, nous restons avec une simulation de fonction aléatoire gaussienne d'une moyenne de 200 et d'une variance de 143, mais dont la portée est maintenant de 20 pixels. L'image vraie est représentée sur la figure 6.25 et l'image bruitée sur la figure 6.26. Comme pour l'exemple précédent, les quatre estimateurs donnent tous le même résultat à 0.5 % près, et l'erreur réelle est de 18.48 pour tous les quatre. Pour cette raison, nous ne présentons qu'une image filtrée, celle du krigeage ordinaire (fig. 6.27). L'histogramme local a été estimé avec un seuil de 220 et un voisinage de 15 × 15. Les résultats sont résumés au Tableau 6.4 et les images sont présentées sur les figures 6.28 à 6.32 respectivement.

Tableau 6.4

Variance réelle	4.7005×10^{-3}	
	Erreur moyenne	Variance d'erreur
d (réel, krigeage linéaire)	0.0131	6.253×10^{-4}
d (réel, KD gaussien)	0.0034	5.662×10^{-4}
d (réel, KD binomial négatif)	0.0095	6.689×10^{-4}
d (réel, espérance conditionnelle)	-0.0038	3.377×10^{-4}

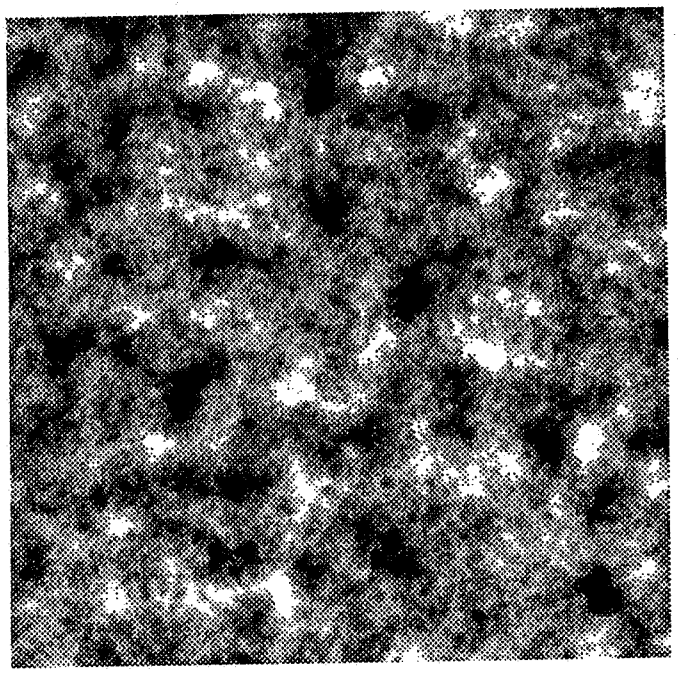


Figure 6.25 Image réelle

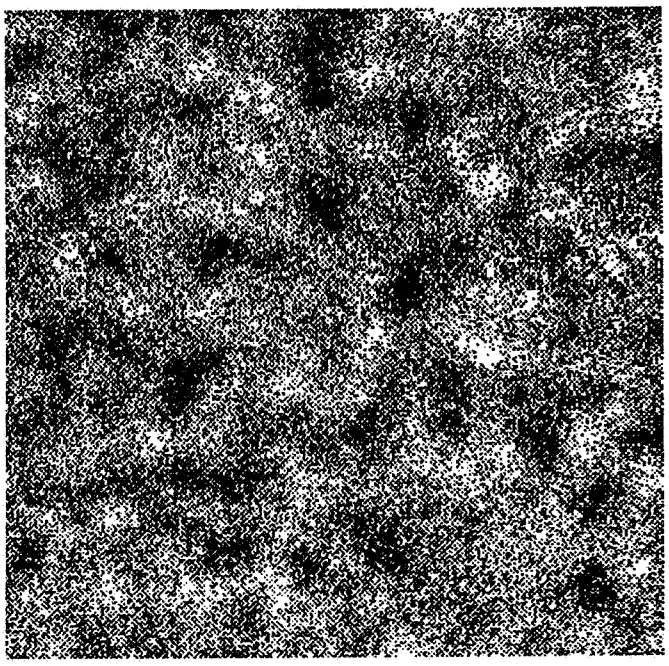


Figure 6.26 Image bruitée

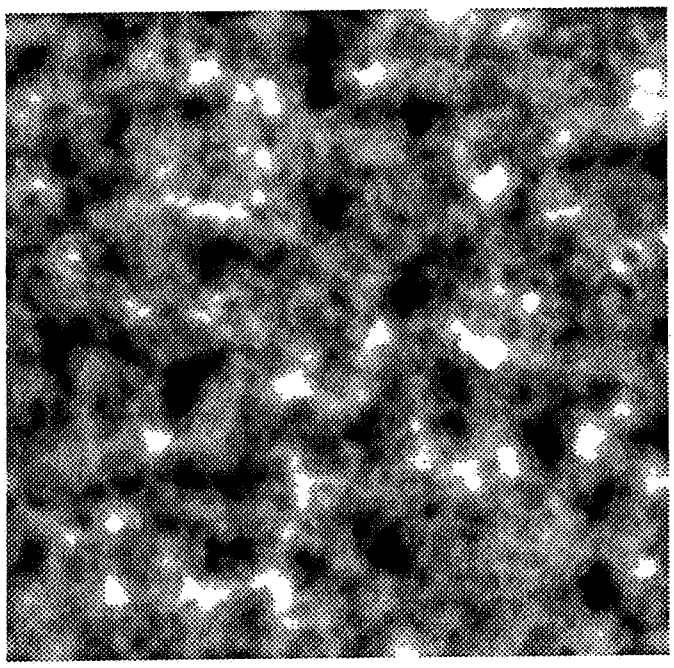


Figure 6.27 Image filtrée,
krigeage linéaire.

On voit que le krigeage linéaire est relativement moins biaisé que dans l'exemple précédent. En effet, quand la portée croît, les pondérateurs de krigeage se répartissent sur un ensemble de valeurs plus grandes et le filtrage est alors plus efficace avec un biais conditionnel plus faible. On obtient par conséquent de meilleurs estimateurs de l'histogramme local. Néanmoins, il n'est nullement évident que le biais soit suffisamment réduit pour que l'estimateur de krigeage linéaire soit un estimateur de l'histogramme local suffisamment bon.

Notons que dans les deux exemples que nous venons de traiter, le krigeage disjonctif binomial négatif/gamma donne essentiellement les mêmes images que le krigeage disjonctif gaussien, alors que les résultats numériques sont légèrement plus mauvais. Ceci peut être dû au fait qu'il n'existe pas de vraie fonction d'anamorphose pour le krigeage binomial négatif. Il semble donc que ce modèle doive être utilisé uniquement lorqu'on peut modéliser les données directement sans passer par une anamorphose. Il sera particulièrement intéressant lorque l'histogramme aura un atome à l'origine, et par conséquent dans les cas où on n'effectue pas d'anamorphose gaussienne.

Dans les exemples suivants, nous ne présentons qu'un krigeage disjonctif, le cas gaussien.

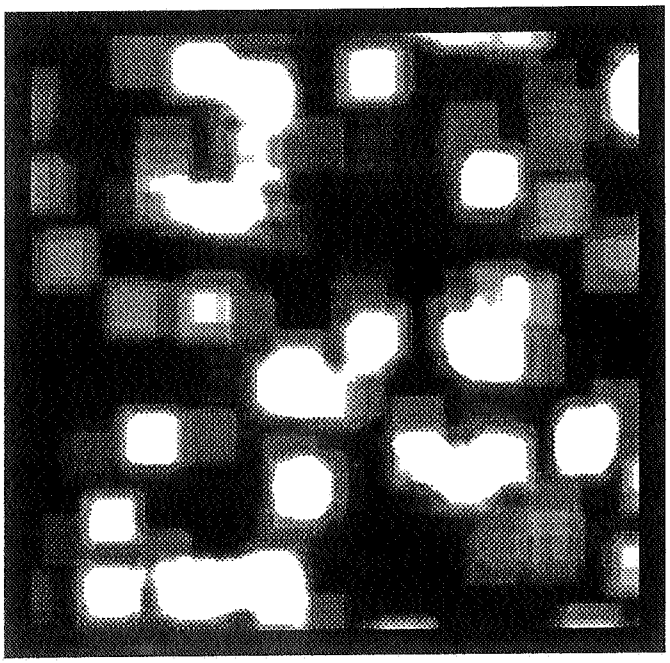


Figure 6.28 Histogramme local à 220. Voisinage = 15 x 15.

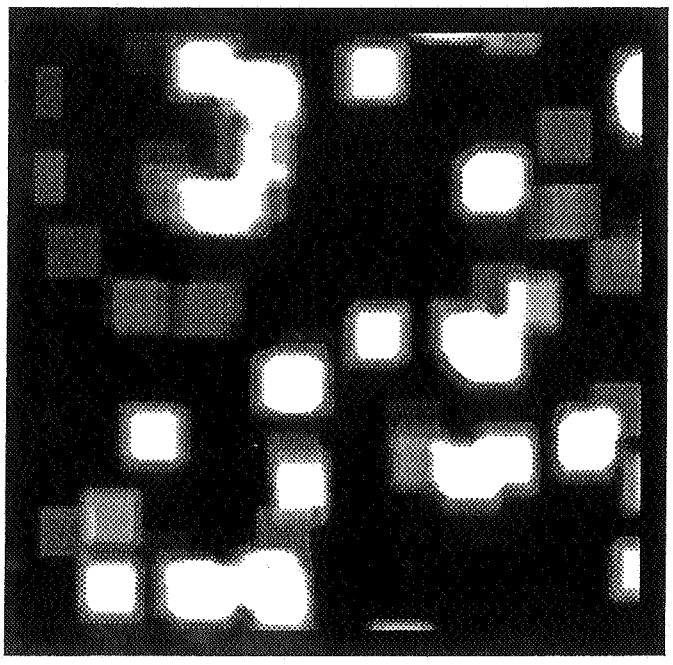


Figure 6.29 Histogramme local, krigeage linéaire.

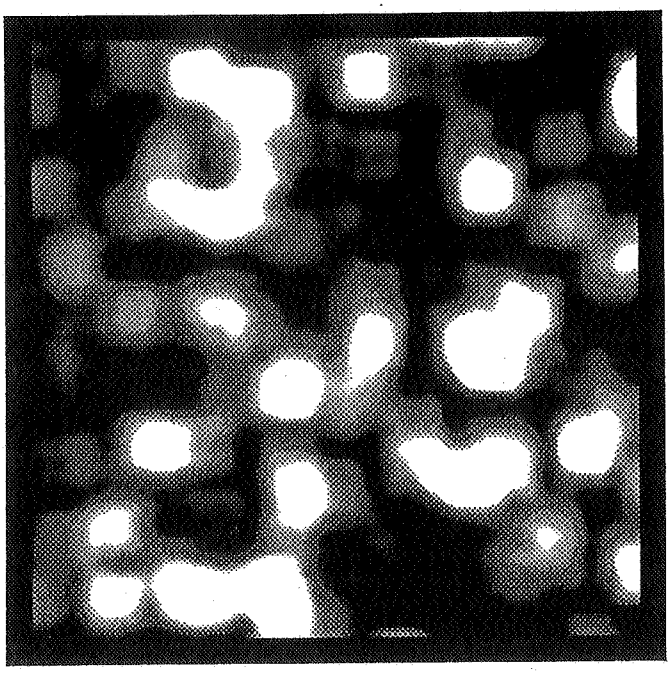


Figure 6.30 Histogramme local, krigeage disjonctif — Gaussien.

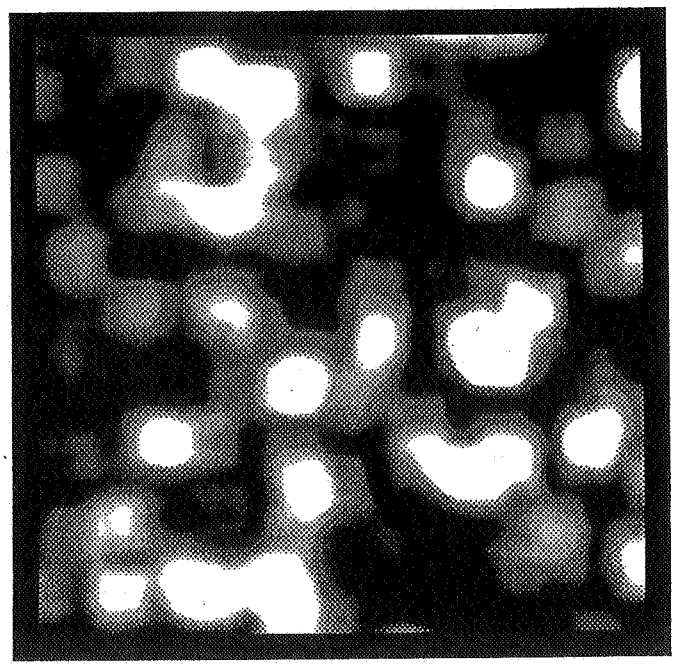


Figure 6.31 Histogramme local, krigeage disjonctif — Binomial négatif / Gamma

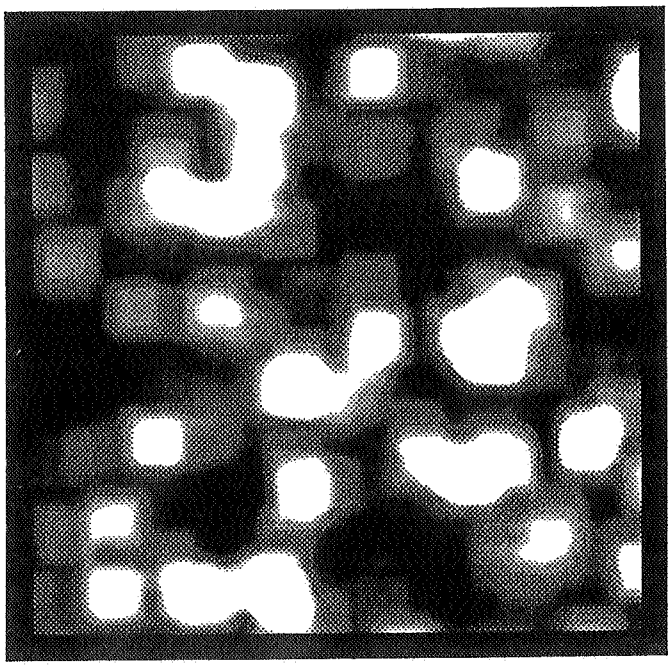


Figure 6.32 Histogramme local, espérance conditionnelle.

3. — Pour cet exemple, on reprend l'image de l'exemple 1 mais en appliquant une anamorphose avant d'effectuer la poissonisation. Les figures 6.33 et 6.34 représentent l'image réelle et sa poissonisée respectivement.

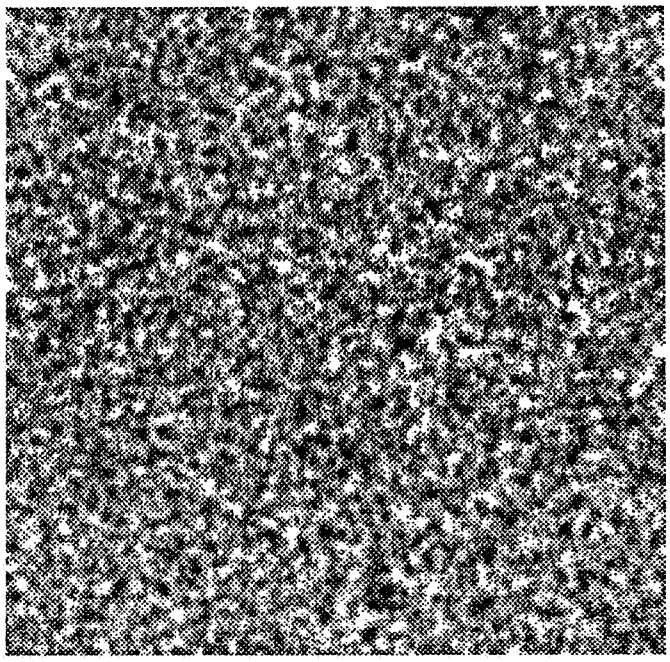


Figure 6.33 Image réelle

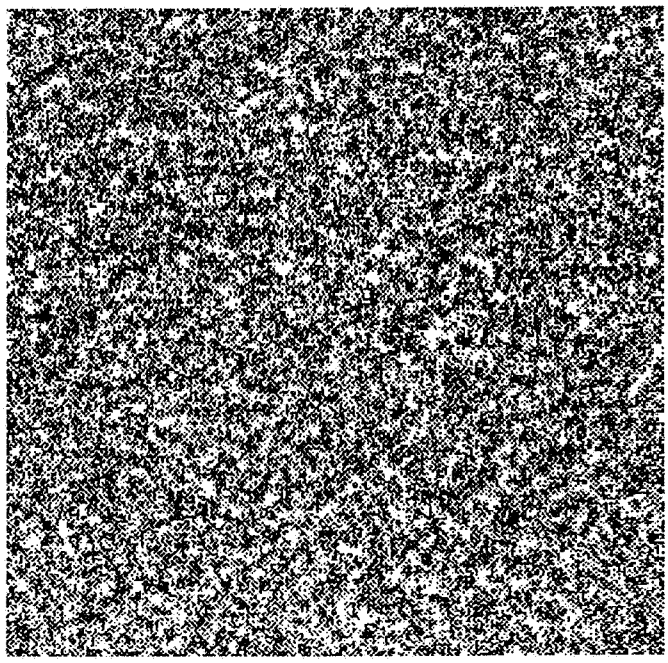


Figure 6.34 Image bruitée

Les histogrammes réel et poissonisé sont présentés à la figure 6.35 et l'estimateur par régularisation est montré à la figure 6.36.

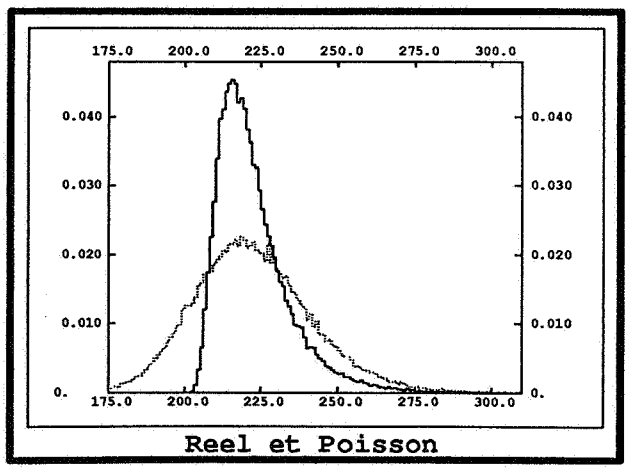


Figure 6.35 Réel ... trait continu
Poisson ... pointillé

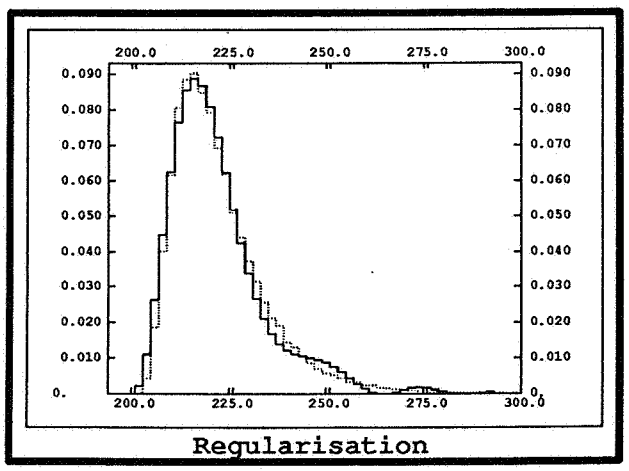


Figure 6.36 Régularisation ... trait continu
Réel ... pointillé

La moyenne de l'image sous-jacente est de 222.97 et sa variance de 147.73. Le filtrage linéaire donne un estimateur de la variance d'erreur de 48.329 et l'erreur réelle est de 52.157. L'image obtenue par filtrage linéaire est représentée à la figure 6.37.

Le krigeage disjonctif gaussien (fig. 6.38) a une variance de krigeage de 58.83 mais l'erreur réelle n'est que de 49.78. L'estimateur de la variance d'erreur est donc moins fiable que pour les exemples précédents. On a déjà une première indication de ce problème lorsqu'on calcule le coefficient de corrélation ϱ par les deux méthodes. La méthode de l'effet de pépite donne $\varrho = 0.6$ et celle de la covariance croisée donne $\varrho = 0.65$. On voit donc que le modèle commence à être un peu moins adapté pour cet exemple que pour les deux précédents. Mais on voit que le filtrage lui-même est resté très robuste, et c'est uniquement au niveau de l'estimation de la variance d'erreur qu'on a un mauvais estimateur.

NOTA. – Quand la corrélation estimée par les deux méthodes donne des résultats différents, on utilise celle obtenue par la méthode de la covariance croisée parce qu'elle est plus fidèle au problème d'estimation de $\bar{Y}$. (L'autre méthode n'utilise pas explicitement la loi de $\bar{Y}$).

L'expérience conditionnelle (fig. 6.39) a une variance d'erreur réelle de 51.92.

On voit maintenant que les estimateurs se ressemblent beaucoup, mais ne sont toutefois plus identiques. En particulier, on observe des "taches" de faible concentration sur l'estimateur de krigeage linéaire qui ne correspondent pas aux structures réelles. Ce point sera discuté dans un exemple suivant.

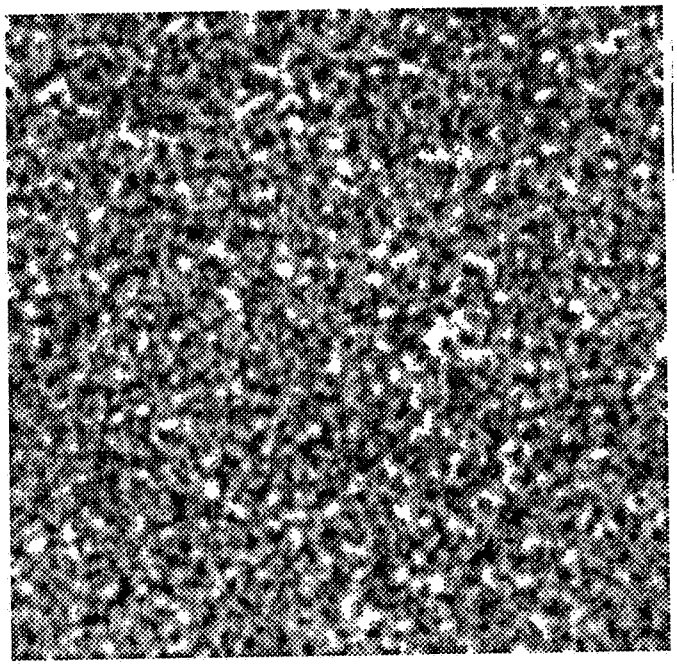


Figure 6.37 Krigeage linéaire

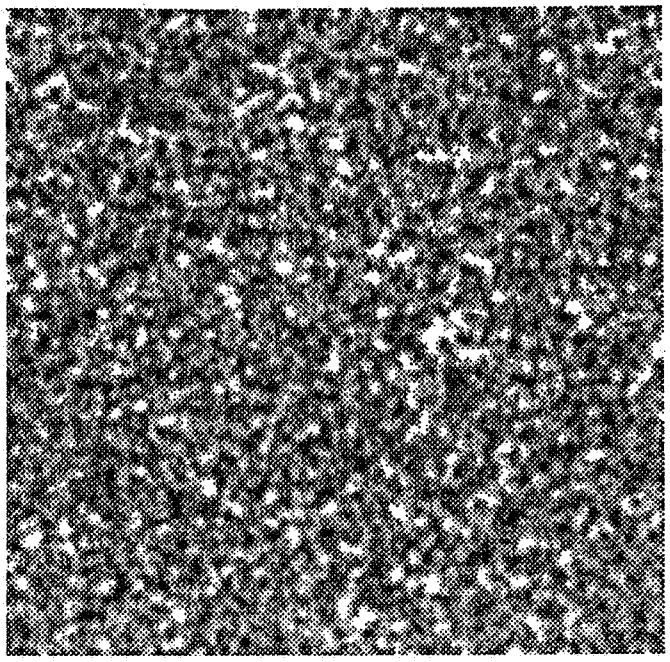


Figure 6.38 Krigeage disjonctif

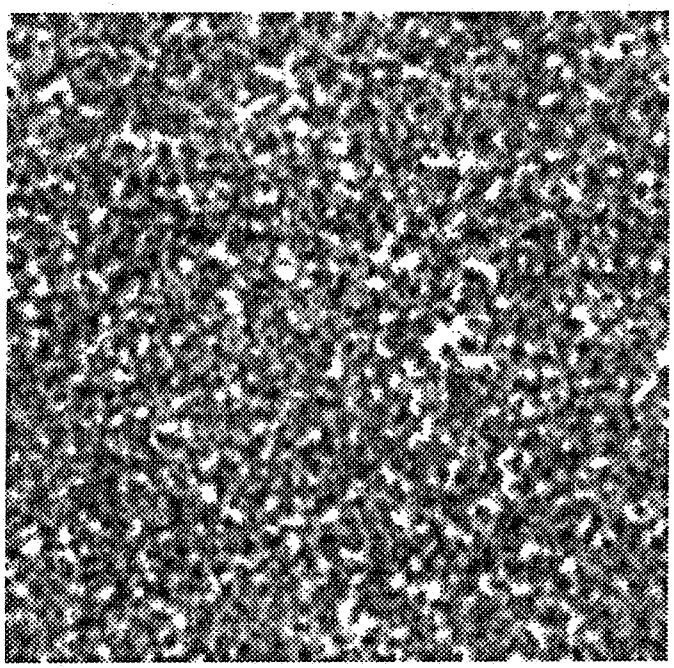


Figure 6.39 Espérance conditionnelle

Nous avons estimé l'histogramme local pour un seuil à 250 et pour des voisinages de 5 × 5 et 11 × 11. Les résultats sont résumés au Tableau 6.5 et les images pour un voisinage de 11 × 11 sont présentées sur les figures 6.40 à 6.43

Tableau 6.5

	5 x 5		11 x 11	
Variance réelle	5.746×10^{-3}		1.937×10^{-3}	
	Erreur moyenne	Variance d'erreur	Erreur moyenne	Variance d'erreur
d (réel, krigeage linéaire)	0.02200	2.710×10^{-3}	0.02196	9.836×10^{-4}
d (réel, KD gaussien)	0.00188	1.548×10^{-3}	0.00176	3.616×10^{-4}
d (réel, espérance conditionnelle)	-0.01866	1.895×10^{-3}	-0.01881	7.968×10^{-4}

On voit ici que l'estimation obtenue par krigeage disjonctif reste plus ou moins non biaisée et donne un estimateur largement meilleur que le cas linéaire ou l'espérance conditionnelle. Il semble donc que l'estimateur obtenu avec le krigeage disjonctif soit plus robuste que celui obtenu avec l'espérance conditionnelle.

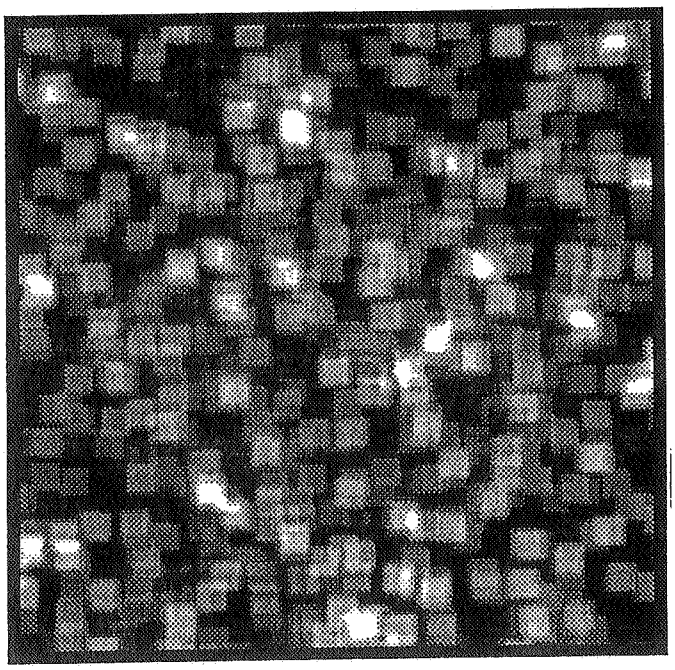


Figure 6.40 Histogramme local à 250. Voisinage = 11 x 11.

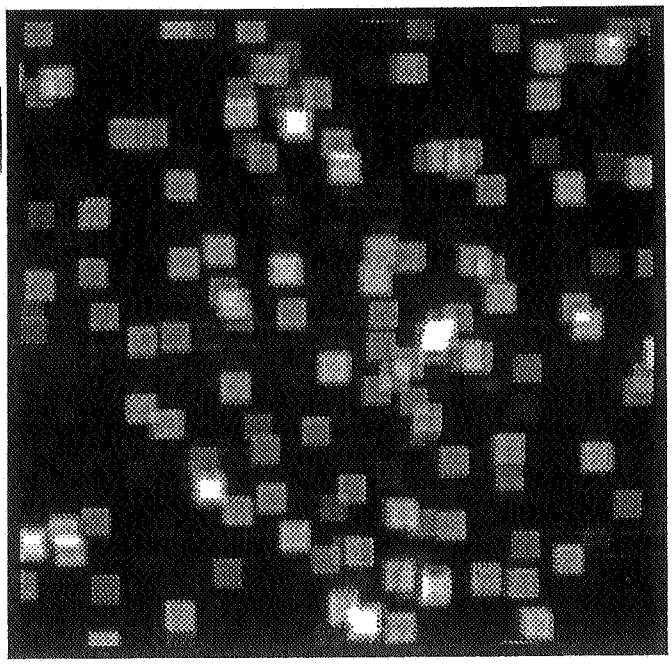


Figure 6.41 Histogramme local, krigeage linéaire.

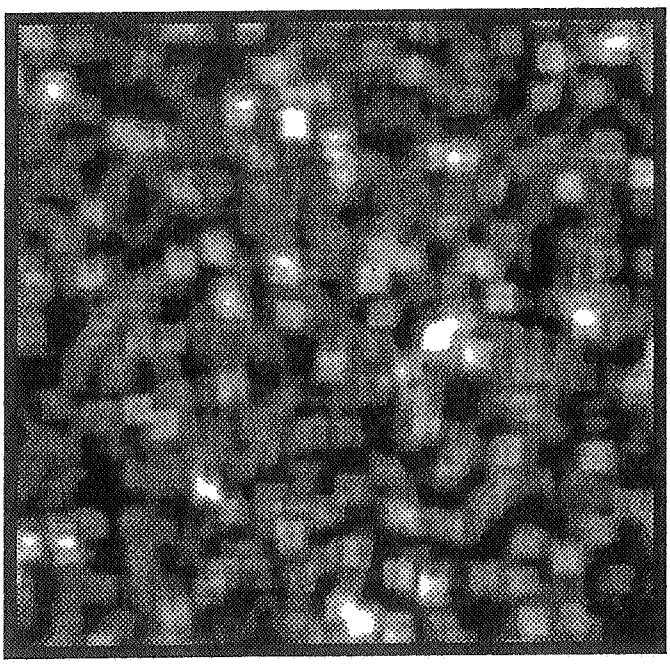


Figure 6.42 Histogramme local, krigeage disjonctif.

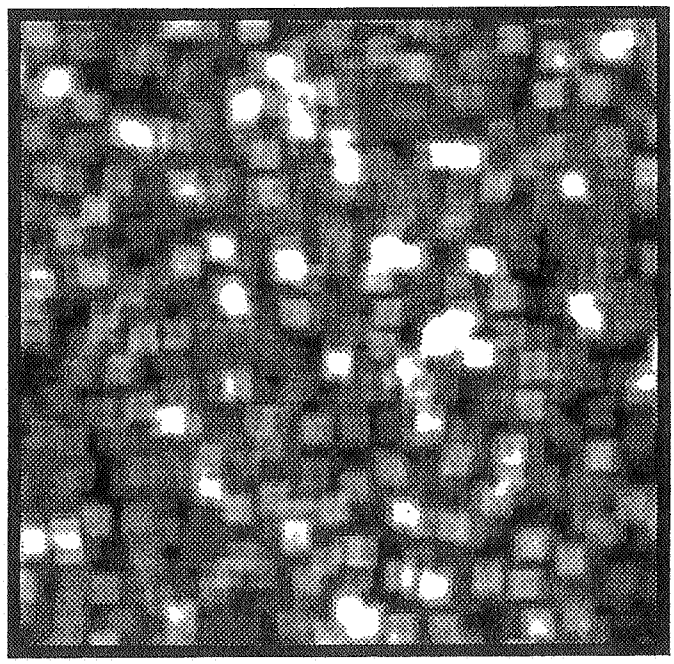


Figure 6.43 Histogramme local, espérance conditionnelle.

4. — Nous considérons maintenant un exemple réel. Il s'agit d'une étude du centre de la plaque obtenue en coulée continue (il s'agit d'une sous ensemble - la partie inférieure sur la gauche de l'image déjà montrée à la figure 2.10 du chapitre II). La moyenne est de 1053 et la variance de 7352. L'image observée est montrée à la figure 6.44. Le krigeage linéaire (fig. 6.45) a une variance d'estimation de 421.8. On montre aussi l'image obtenue par krigeage disjonctif (fig. 6.46), et l'image obtenue par l'espérance conditionnelle (fig. 6.47).

On présente l'estimation de l'histogramme sous-jacent et son ajustement en termes de polynômes d'Hermite (fig. 6.48). Si l'on considère l'estimateur de l'histogramme sous-jacent comme l'histogramme réel, on peut comparer cet histogramme vrai avec l'histogramme des images filtrées. La figure 6.49 montre cet histogramme "réel" avec la loi poissonisée. La loi de l'image obtenue par krigeage linéaire, par krigeage disjonctif et par espérance conditionnelle est présentée aux figure 6.50, 6.51 et 6.52 respectivement. Si l'histogramme "réel" - dont on rappelle qu'il est obtenu par régularisation, peut être considéré comme un bon estimateur de la loi réelle, alors on voit que le krigeage linéaire n'a pas suffisamment lissé les points ayant des valeurs entre 950 et 1070, ce qui est cohérent avec les remarques déjà faites pour les exemples du chapitre III. Ceci peut expliquer le "taches" de faible concentration qu'on peut voir sur l'image du krigeage linéaire, et qui sont absentes des images obtenues avec les deux autres estimateurs. Nous confirmerons cette hypothèse avec l'exemple suivant.

Figure 6.44 Image observée

Figure 6.45 Filtrage linéaire

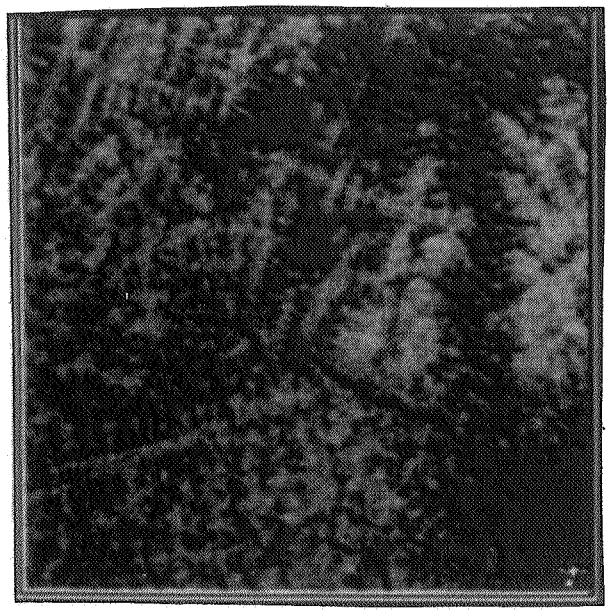


Figure 6.46 Krigeage disjonctif

Figure 6.47 Espérance conditionnelle

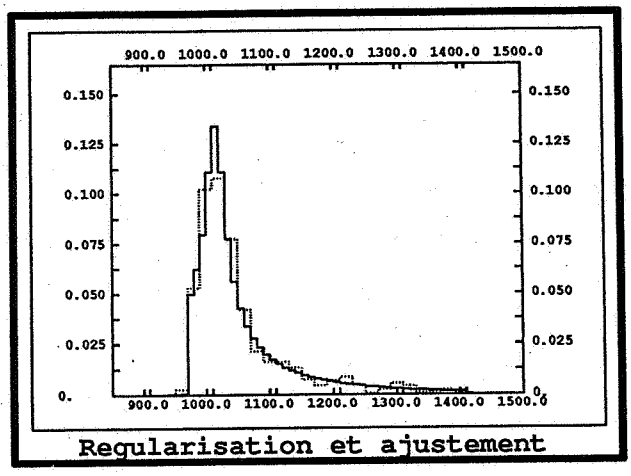


Figure 6.48 Estimation par régularisation en trait pointillé et son ajustement avec 10 polynômes d'Hermite (trait continu)

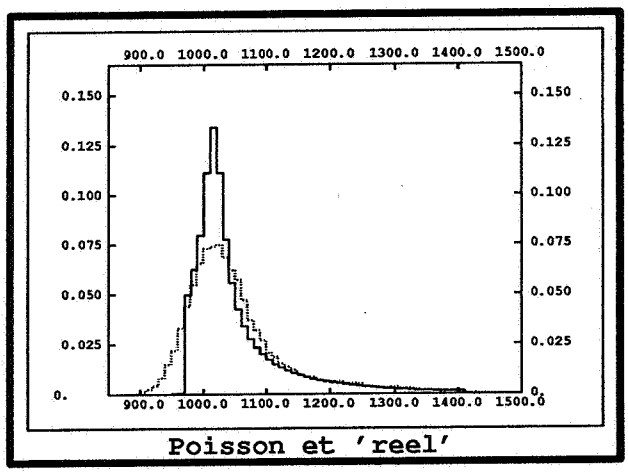


Figure 6.49 Régularisation ... trait continu
Poisson ... pointillé

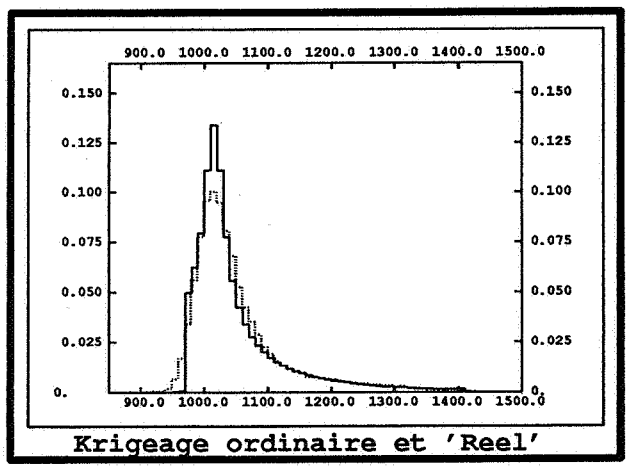


Figure 6.50 Régularisation ... trait continu
Krigeage linéaire ... pointillé

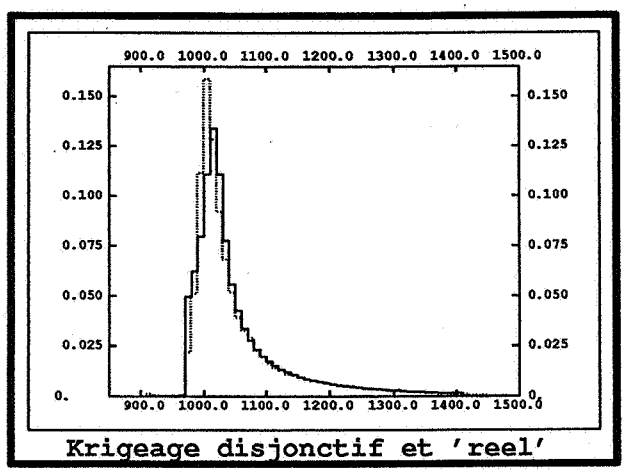


Figure 6.51 Régularisation ... trait continu
Krigeage disjonctif ... pointillé

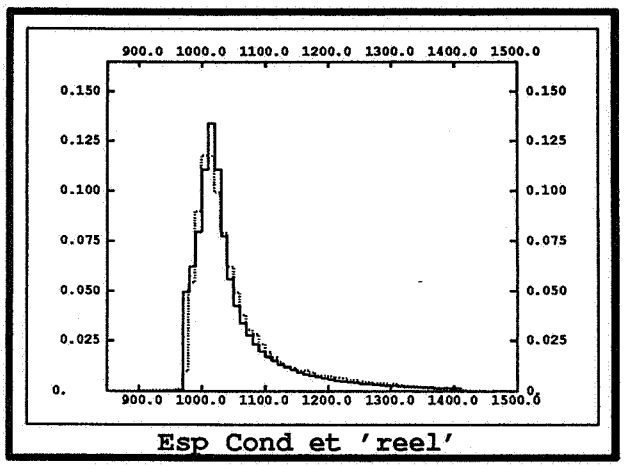


Figure 6.52 Régularisation ... trait continu
Espérance conditionnelle ... pointillé

Auparavant, estimons l'histogramme local pour un voisinage de 5×5 et un seuil de 1070. Les images obtenues sont présentées sur les figures 6.53 à 6.56, pour lesquelles on a également calculé l'histogramme local sur l'image bruitée. Ceci a été possible avec cet exemple parce que le bruit était très faible dans cette image. Néanmoins, on s'attend à ce que l'image bruitée surestime la probabilité locale de manière assez importante. Le fait que l'espérance conditionnelle ait des valeurs aussi fortes suggère, bien que cela soit impossible à vérifier, qu'il existe un biais important sur l'estimateur de l'espérance conditionnelle.

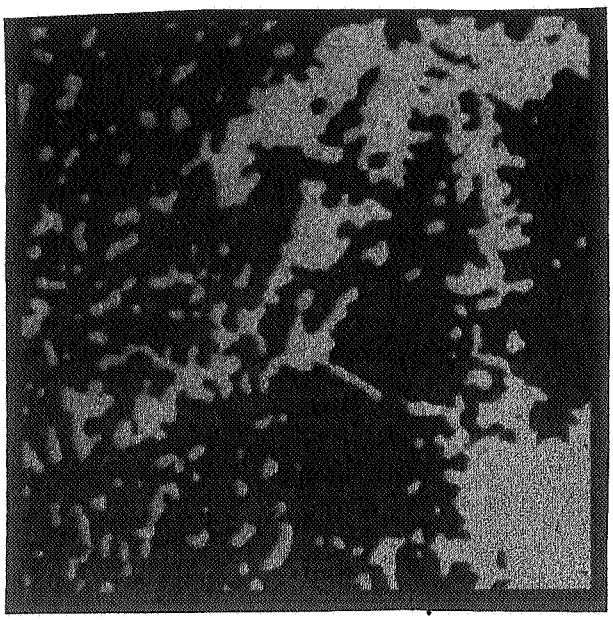


Figure 6.53 Histogramme local à 1070. Voisinage = 5 x 5, krigeage linéaire.

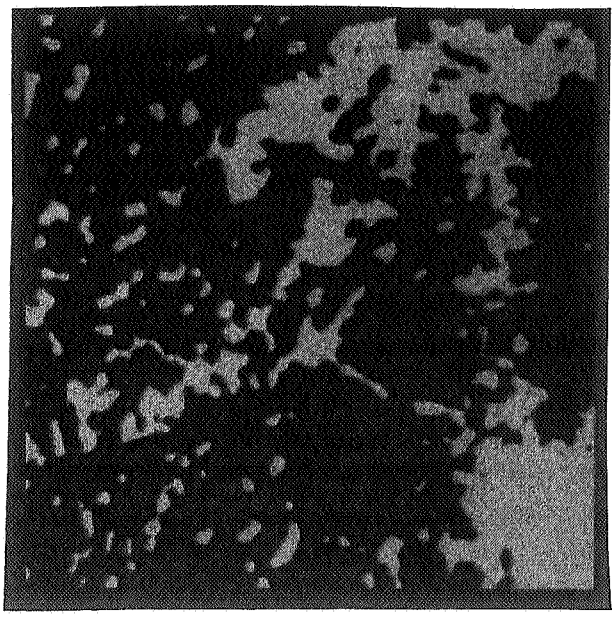


Figure 6.54 Histogramme local, krigeage disjonctif.

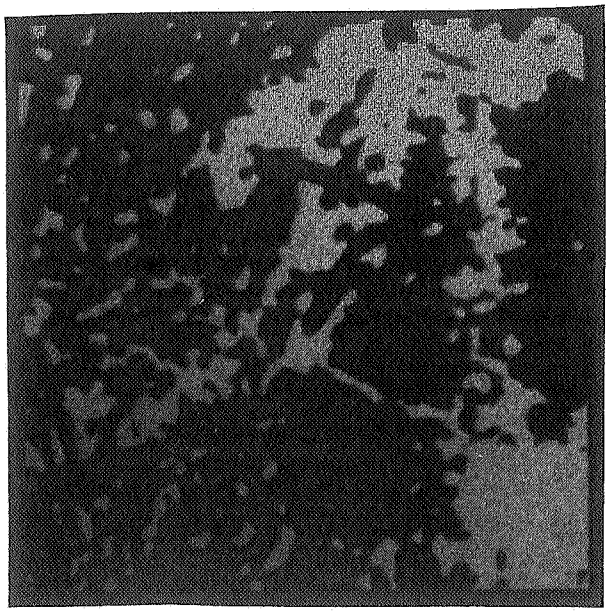


Figure 6.55 Histogramme local, espérance conditionnelle.

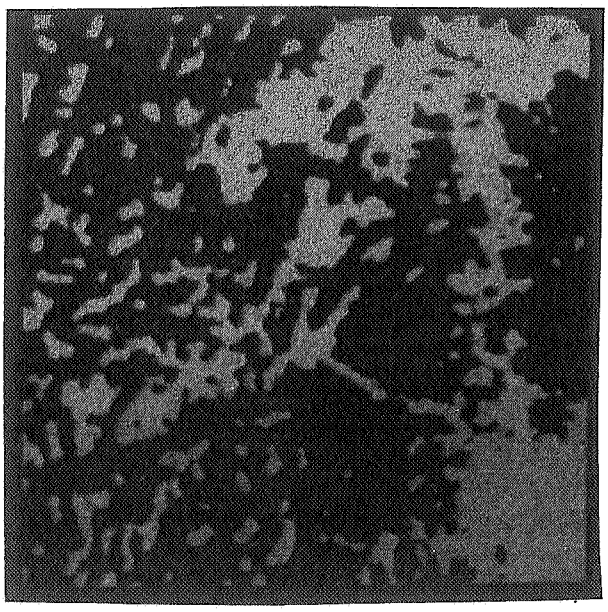


Figure 6.56 Histogramme local calculé sur l'image bruitée.

5. — Pour mieux comprendre l'exemple précédent, on reprend l'image simulée de portée 5 et on applique une anamorphose qui donne une loi équivalente à celle de l'image réelle de l'exemple précédent. Notre image simulée est présentée à la figure 6.57 et l'image bruitée à la figure 6.58. La variance sous-jacente est de 5993.5 et la moyenne de 1052.7. Les histogrammes réels et poissoniens sont montrés à la figure 6.59.

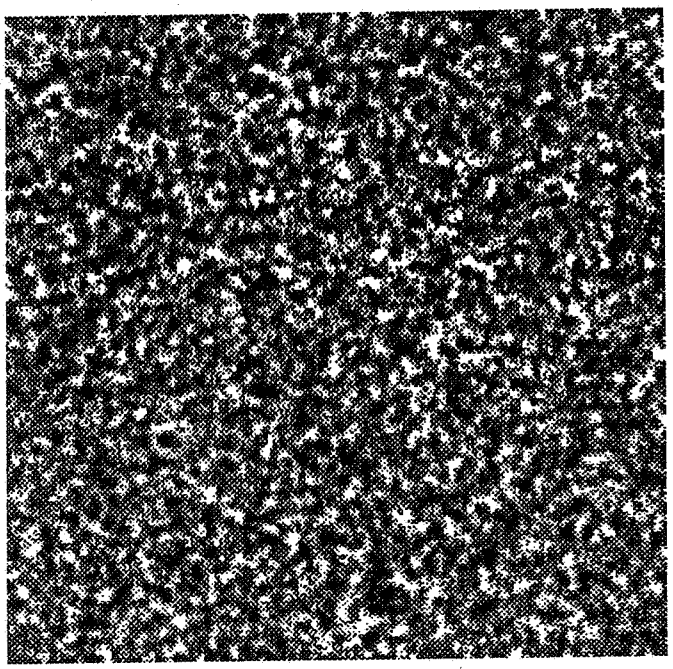


Figure 6.57 Image réelle

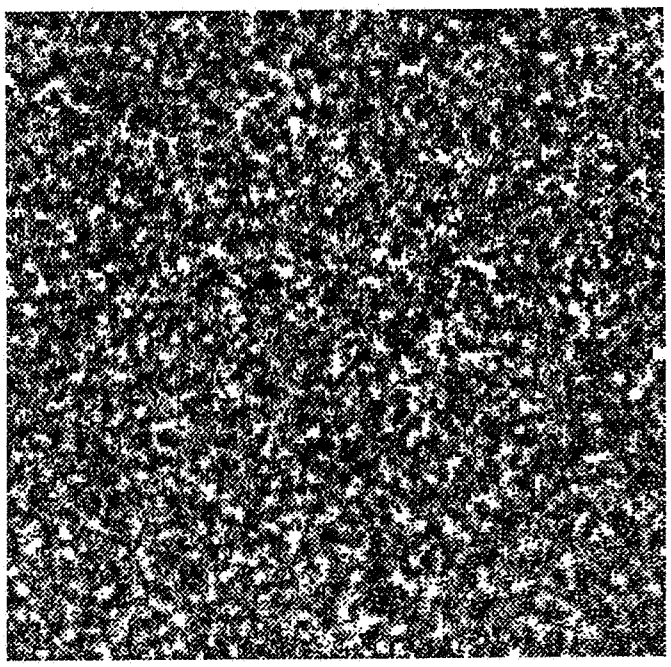


Figure 6.58 Image bruitée

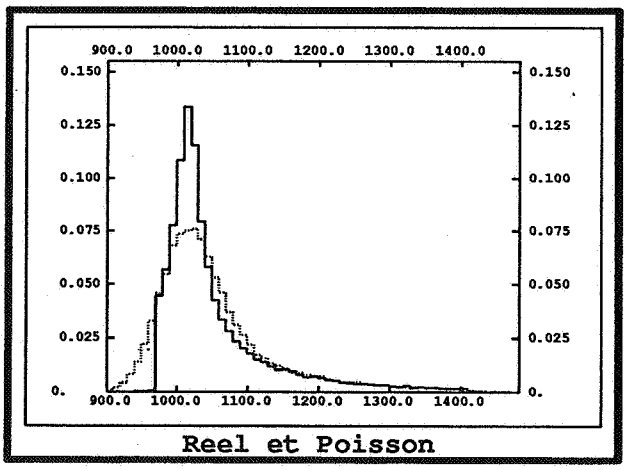


Figure 6.59 L'histogramme réel est en trait continu. L'histogramme bruité est en trait pointillé.

On a fait un filtrage linéaire et l'image qui en résulte est à la figure 6.60, et l'histogramme de l'image obtenue est montré à la figure 6.61. Comme on peut le voir, le krigeage n'a pas suffisamment filtré les pixels ayant une valeur proche du mode, ce qui confirme notre intuition dans l'exemple réel. L'image obtenue par krigeage disjonctif est présentée à la figure 6.62, et l'histogramme de cette image à la figure 6.63. On observe une fois encore les mêmes types de résultats que pour l'exemple réel. Pour finir, on trouve l'image obtenue par espérance conditionnelle à la figure 6.64 et l'histogramme correspondant à la figure 6.65. Une fois de plus, on voit qu'il existe un très bon accord entre les valeurs ainsi obtenues et l'exemple précédent. La variance d'estimation suggérée par le krigeage linéaire est de 812.83 et l'erreur réelle est de 780.21. Pour le krigeage disjonctif, l'erreur d'estimation est de 752.4 et l'erreur réelle de 609.39. L'erreur réelle pour l'espérance conditionnelle est de 647.71.

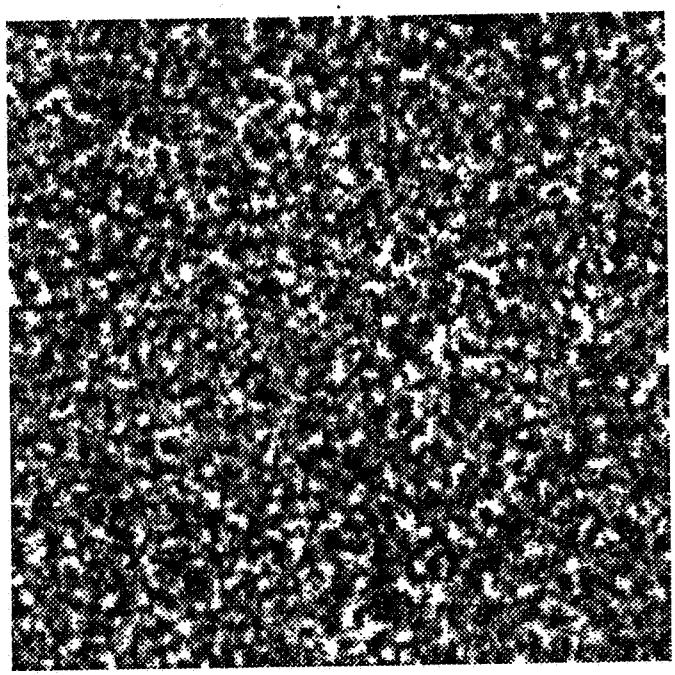


Figure 6.60 Krigeage linéaire

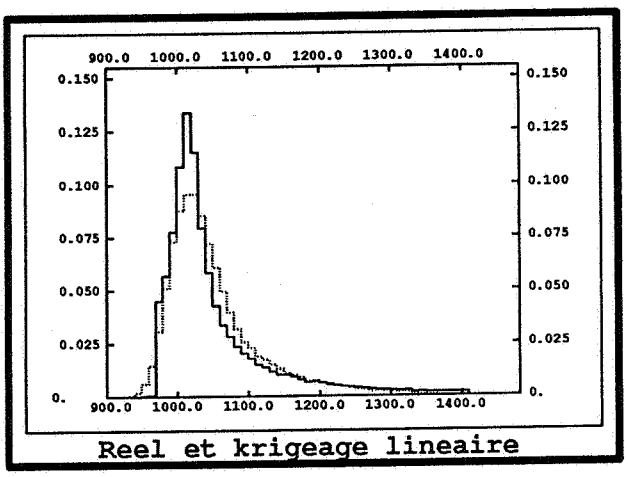


Figure 6.61 L'histogramme réel est en trait continu. L'histogramme obtenu par krigeage linéaire est en trait pointillé.

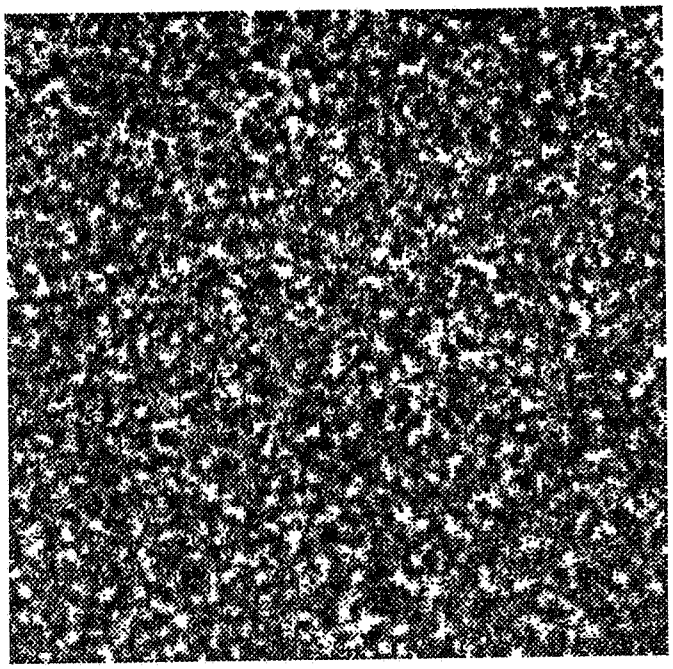


Figure 6.62 Krigeage disjonctif

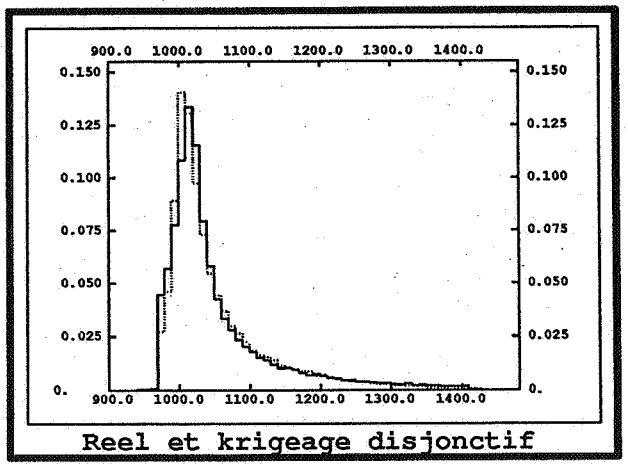


Figure 6.63 L'histogramme réel est en trait continu. L'histogramme obtenu par krigeage disjonctif est en trait pointillé.

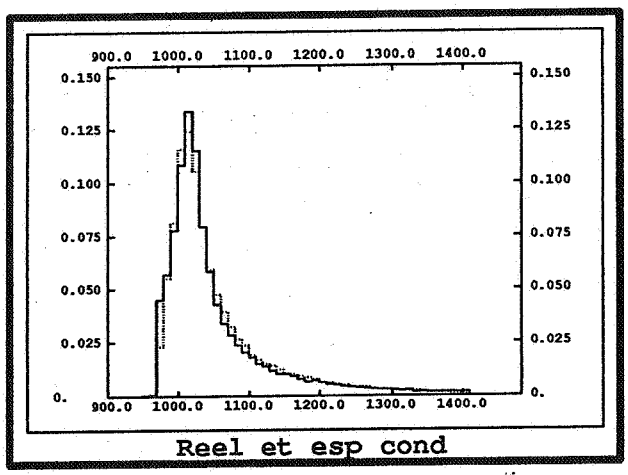


Figure 6.64 L'histogramme réel est en trait continu. L'histogramme obtenu par l'espérance conditionnelle est en trait pointillé.

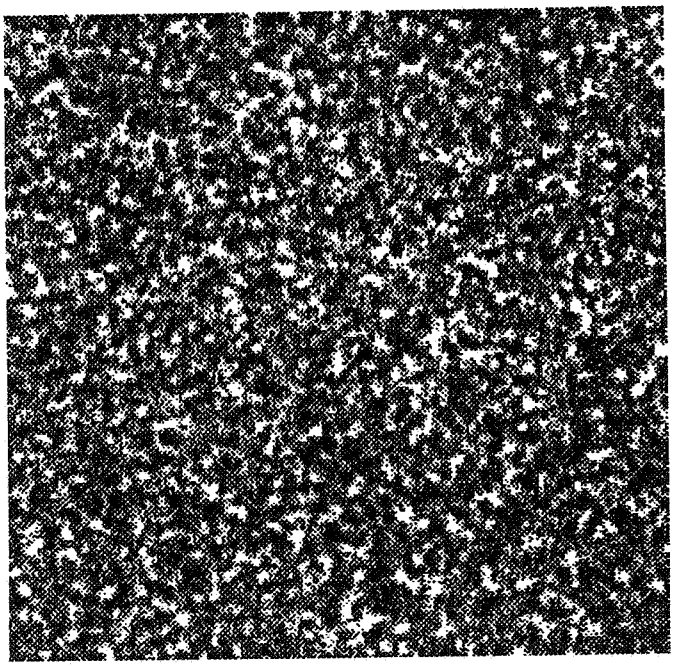


Figure 6.65 Espérance conditionnelle

Passons au calcul des histogrammes locaux. Comme précédemment, on estime l'histogramme à un seuil de 1070 pour les méthodes linéaires, le krigeage disjonctif, l'espérance conditionnelle et également l'image brute. Les résultats sont résumés au Tableau 6.6. Les images sont montrées aux figures 6.66 à 6.70 pour le cas 11 × 11.

Tableau 6.6

	5 x 5		11 x 11	
Variance réelle	3.7200×10^{-1}		1.2458×10^{-1}	
	Erreur moyenne	Variance d'erreur	Erreur moyenne	Variance d'erreur
d (réel, krigeage linéaire)	-0.0288	5.791×10^{-3}	-0.0289	2.072×10^{-3}
d (réel, KD gaussien)	0.0021	3.027×10^{-3}	0.0020	0.638×10^{-3}
d (réel, espérance conditionnelle)	-0.0034	2.946×10^{-3}	-0.0035	0.637×10^{-3}
d (réel, image bruitée)	-0.0398	5.728×10^{-3}	-0.0399	2.476×10^{-3}

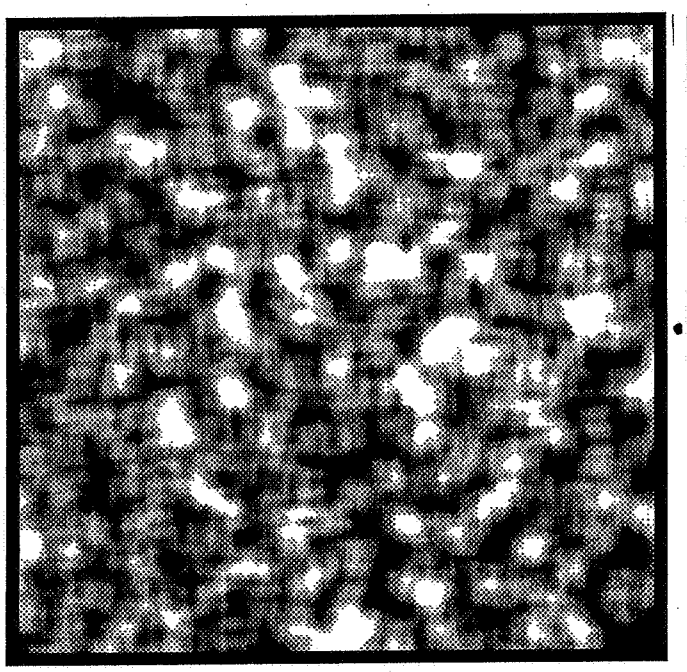


Figure 6.66 Histogramme local à 1070. Voisinage 11 x 11

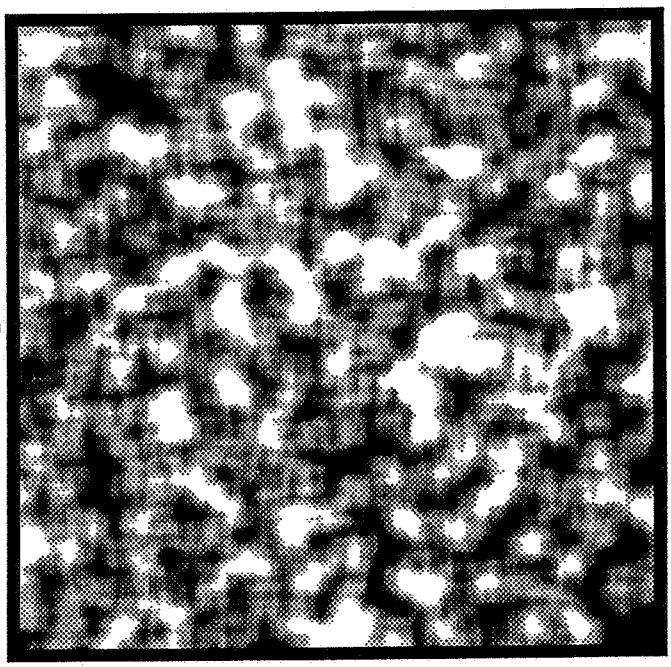


Figure 6.67 Histogramme local, krigeage linéaire.

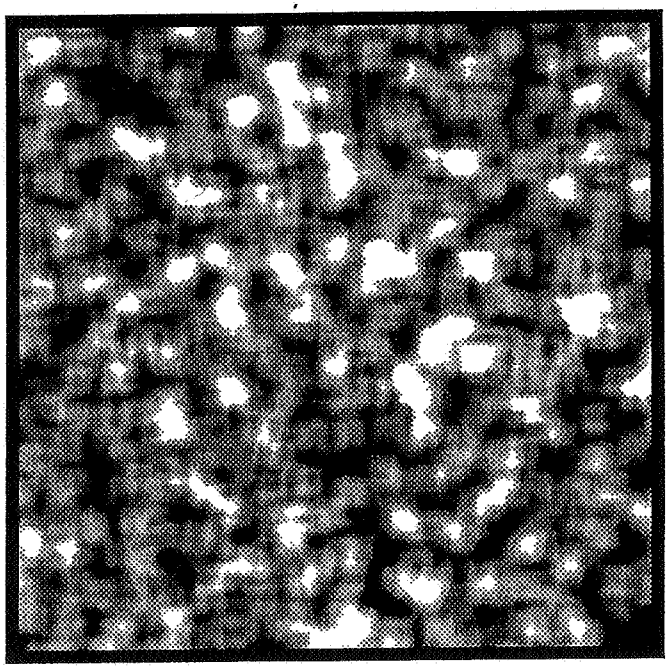


Figure 6.68 Histogramme local, krigeage disjonctif

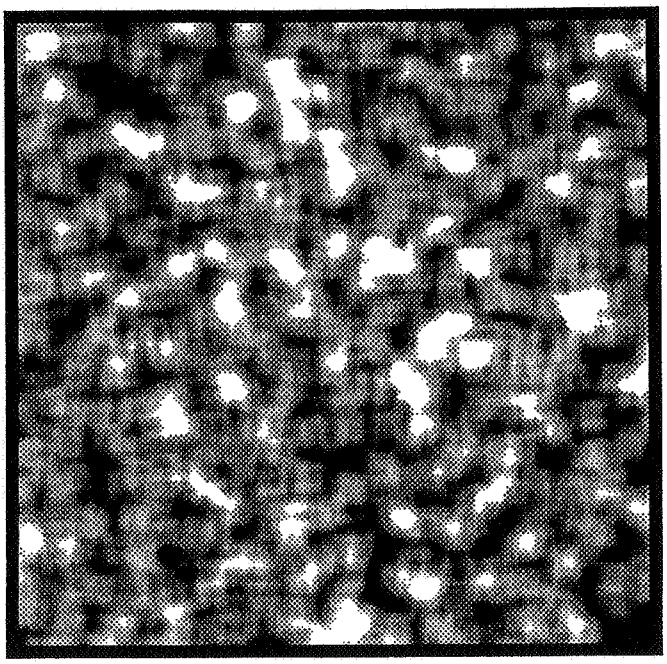


Figure 6.69 Histogramme local, espérance conditionnelle.

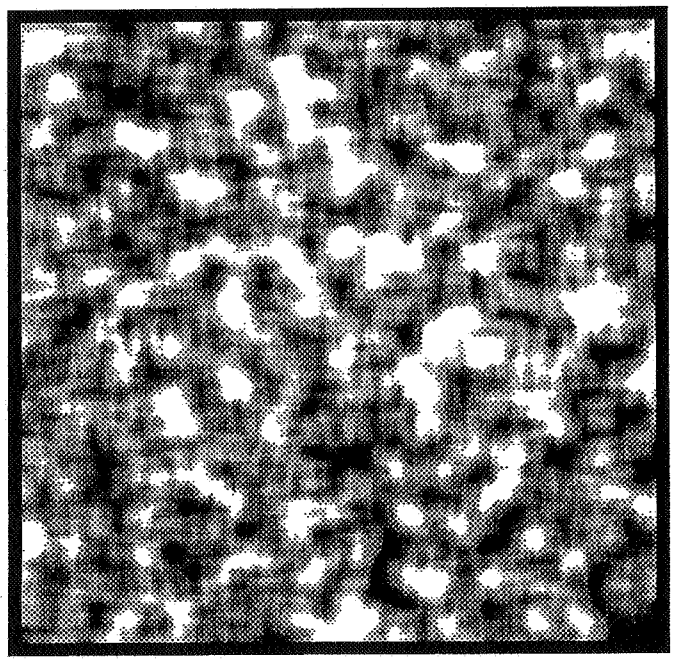


Figure 6.70 Histogramme local calculé sur l'image bruitée

Dans cet exemple, on trouve que le filtre du krigeage disjonctif (et, à un moindre degré, l'espérance conditionnelle) donne un meilleur estimateur que le krigeage linéaire, mais on constate que l'estimateur de la variance d'erreur n'est toujours pas très bon. Pour l'estimation de la probabilité locale, le krigeage disjonctif est nettement meilleur que le cas linéaire et que l'image obtenue à partir de l'image brute. L'espérance conditionnelle donne un aussi bon résultat que le krigeage disjonctif, mais on a vu dans les exemples précédents qu'il donnait des estimateurs de qualité assez variable, même lorsque par construction les images sous-jacentes étaient vraiment les anamorphoses de fonctions aléatoires de loi multigaussienne. Il semble donc plus prudent d'utiliser le krigeage disjonctif dont on a vu qu'il avait le mérite d'être beaucoup plus stable et robuste par rapport aux déviations du cas idéal.

6. — Comme dernier exemple, nous prenons un cas où le krigeage disjonctif n'est pas vraiment réalisable. Commençons avec notre simulation de portée 5 et effectuons une anamorphose afin d'obtenir l'image de la figure 6.71. L'image bruitée est représentée à la figure 6.72.

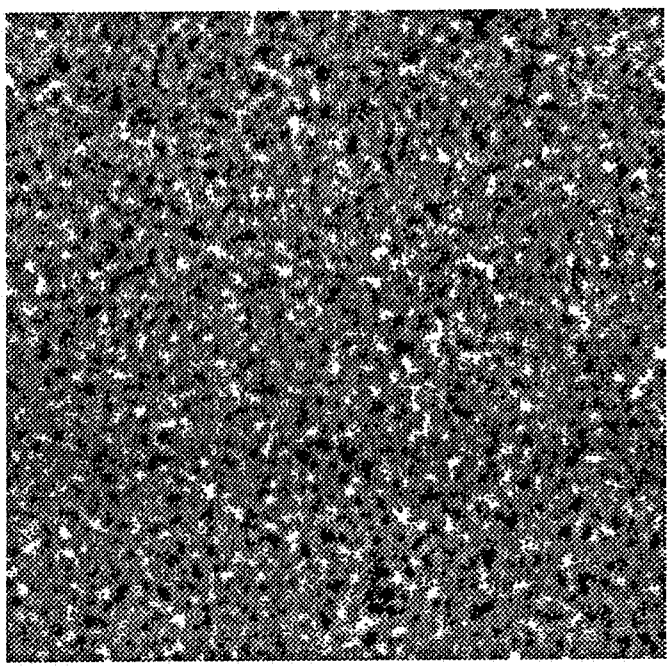


Figure 6.71 Image réelle

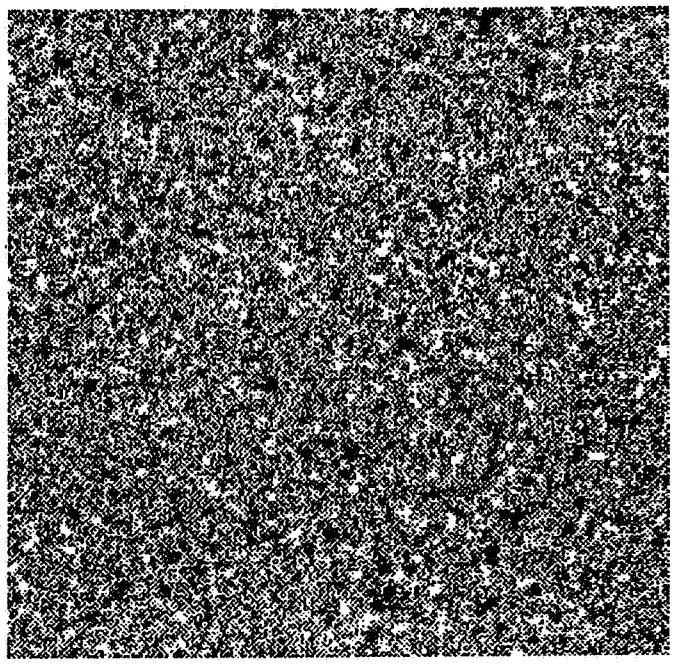


Figure 6.72 Image bruitée

La moyenne de l'image sous-jacente est de 700.69 et la variance de 1404.48. La variance de l'image bruitée est de 2100.97. On a construit un estimateur par régularisation et on a ensuite trouvé une anamorphose en polynômes d'Hermite pour cet estimateur. On voit le résultat avec la loi réelle à la figure 6.73. Bien que l'ajustement semble bon visuellement, la variance de la loi obtenue par régularisation n'est en fait que de 1077.5, et son ajustement par polynômes d'Hermite n'explique que 1042.2. Nous avons donc une diminution de variance de presque 40 % sur la variance réelle. Dans tous les cas étudiés jusqu'à présent, la perte de variance ne dépassait jamais 5 %, même pour l'étude réelle. Avec un estimateur de la loi sous-jacente de mauvaise qualité comme celui-ci, on peut s'attendre à des résultats inférieurs pour le krigeage disjonctif.

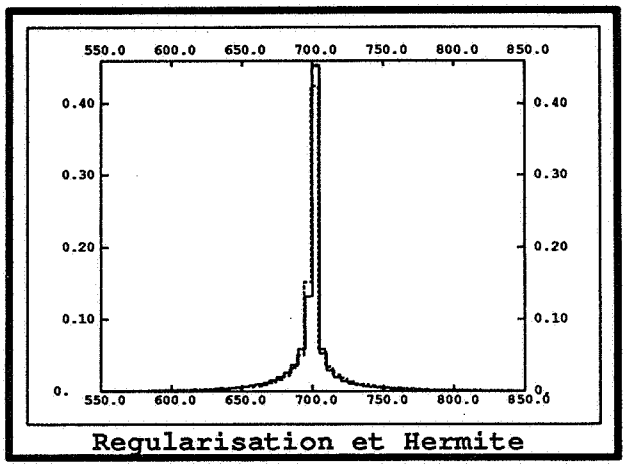


Figure 6.73 Régularisation ajustée par 10 polynômes d'Hermite (trait continu) et l'histogramme réel.

La variance d'estimation pour le filtrage linéaire est de 347.8 et l'erreur réelle de 351.48. On voit donc que, même dans le cas extrêmement non gaussien, la variance d'estimation du krigeage linéaire reste très robuste. L'image obtenue par filtrage linéaire est montrée à la figure 6.74. Le krigeage disjonctif (fig. 6.75) a une variance d'estimation de 428.3 et une erreur réelle de 349.85, et l'espérance conditionnelle (fig. 6.76) une erreur réelle de 306.19. On constate donc que le krigeage disjonctif ne donne pas de plus mauvais résultats que le krigeage linéaire, et de plus il donne une image qui ressemble davantage à la réalité, en ce sens qu'il la lisse beaucoup moins. Les histogrammes obtenus par les filtrages différents sont montrés à la figure 6.77

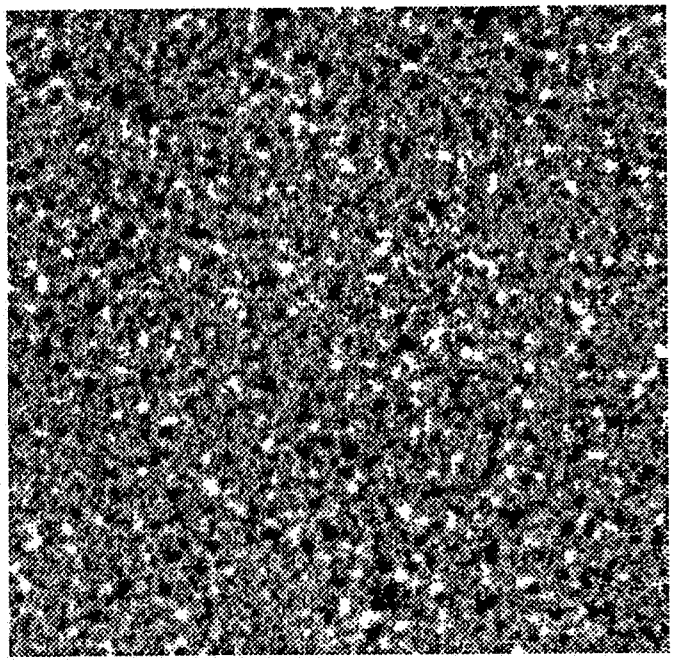


Figure 6.74 Krigeage linéaire

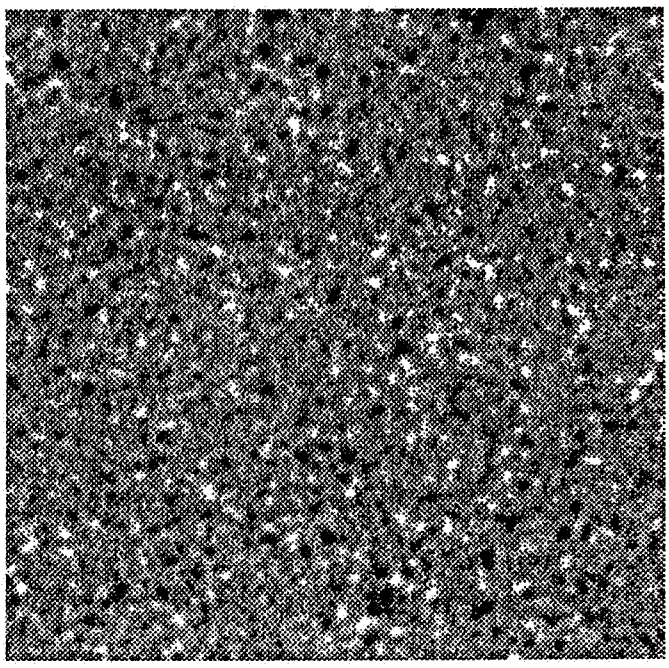


Figure 6.75 Krigeage disjonctif

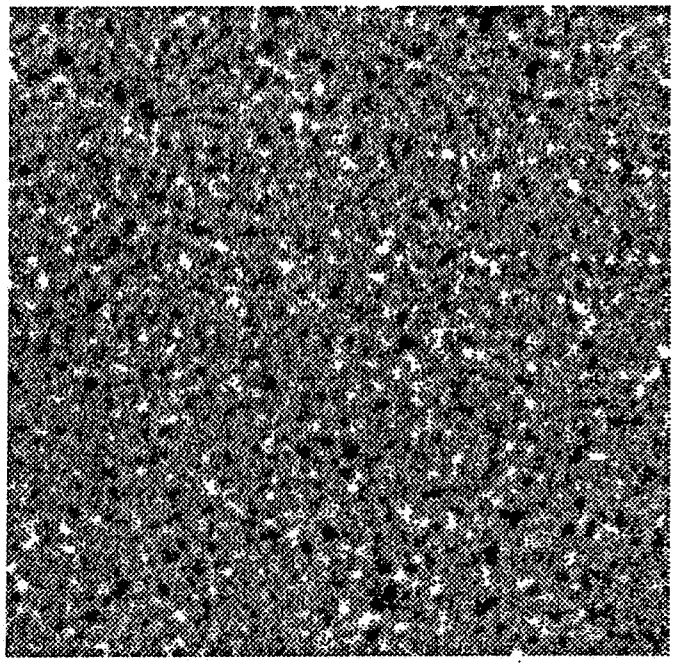


Figure 6.76 Espérance conditionnelle

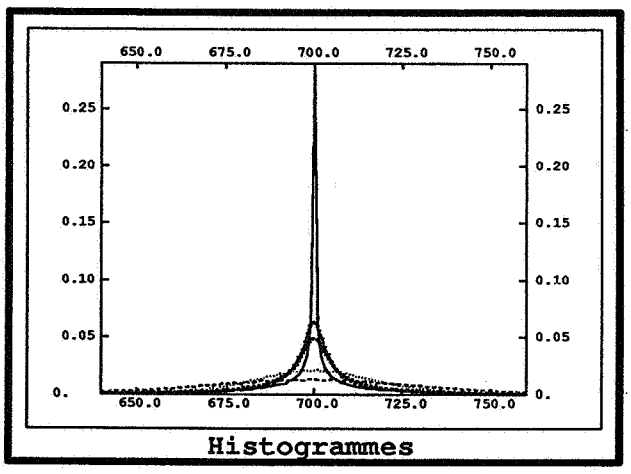


Figure 6.77 De bas en haut à 700 : bruité, krigeage ordinaire, espérance conditionnelle, krigeage disjonctif, réel.

Nous avons estimé la probabilité locale à un seuil de 760. Il s'agit d'un choix relativement avantageux pour le krigeage linéaire qui sera par contre le plus biaisé et le plus mauvais pour un seuil compris entre 715 et 725. Les images obtenues pour un voisinage de 11 × 11 sont présentées aux figures 6.78 à 6.81 et les résultats sont résumés au Tableau 6.7. On constate que même pour ce seuil pourtant favorable au krigeage linéaire, on obtient d'aussi bons résultats avec le krigeage disjonctif.

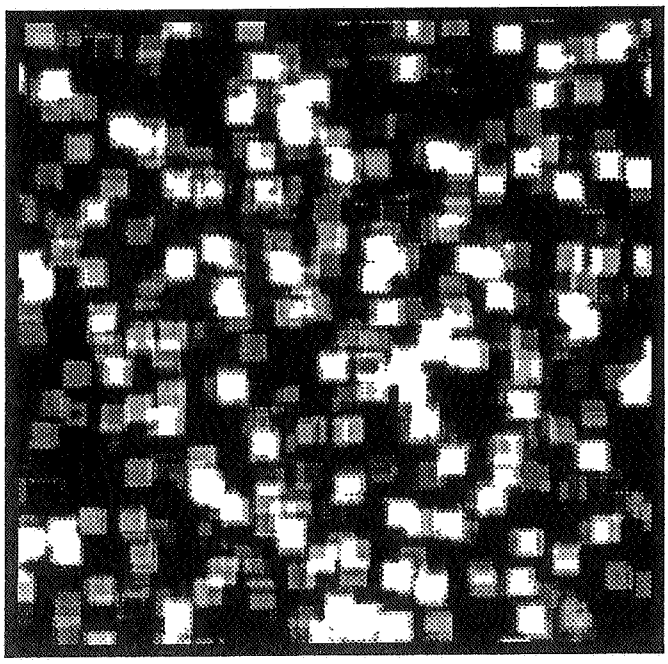


Figure 6.78 Histogramme local à 760, voisinage 11 x 11.

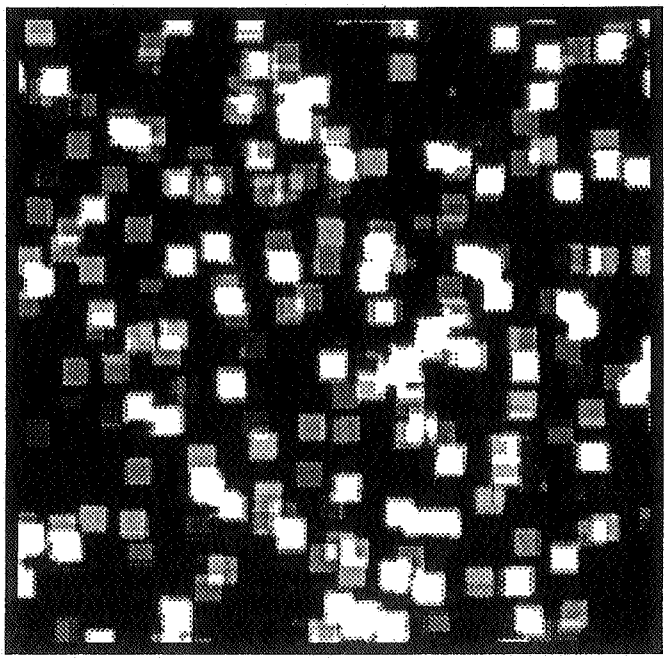


Figure 6.79 Histogramme local, krigeage linéaire.

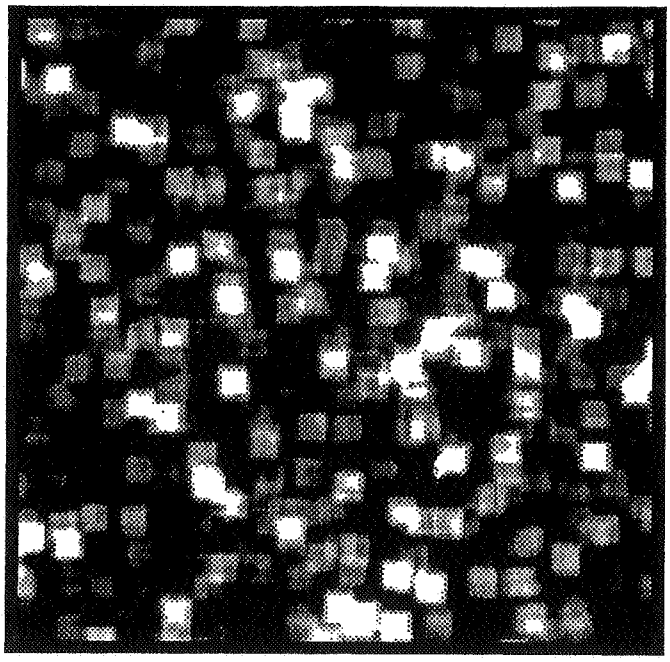


Figure 6.80 Histogramme local, krigeage disjonctif.

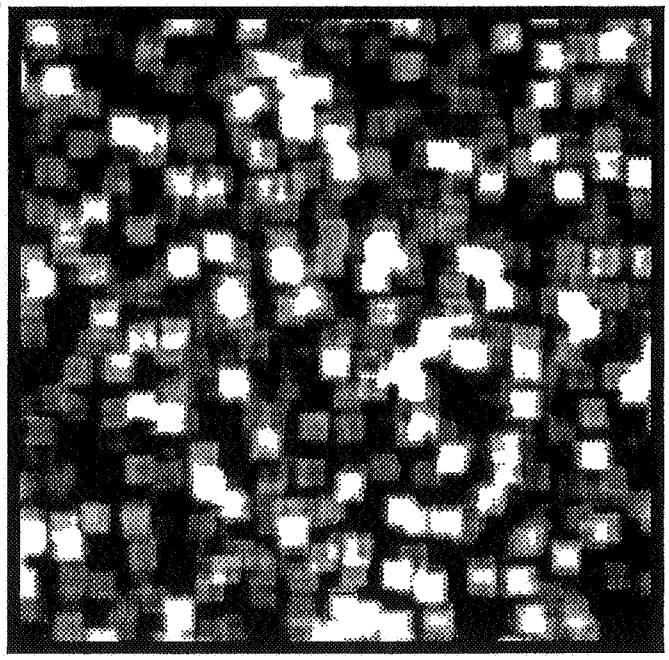


Figure 6.81 Histogramme local, espérance conditionnelle.

Tableau 6.7

	5 x 5		11 x 11	
Variance réelle	55.845×10^{-4}		14.800×10^{-4}	
	Erreur moyenne	Variance d'erreur	Erreur moyenne	Variance d'erreur
d (réel, krigeage linéaire)	0.00673	9.332×10^{-4}	0.00672	2.360×10^{-4}
d (réel, KD gaussien)	0.00533	9.927×10^{-4}	0.00531	2.340×10^{-4}
d (réel, espérance conditionnelle)	-0.00758	8.407×10^{-4}	-0.00758	2.250×10^{-4}

Il n'est pas difficile de trouver la cause des problèmes qu'on rencontre avec le krigeage disjonctif dans cet exemple. Il s'agit de la mauvaise qualité de l'estimation de la loi sous-jacente par régularisation, et plus particulièrement la sous-estimation de la variable sous-jacente. Si l'on applique le krigeage disjonctif avec la vraie loi au lieu de celle obtenue par régularisation (fig. 6.82), on réduit la variance d'erreur réelle à 280.23. Avec l'espérance conditionnelle (fig. 6.83) la variance d'erreur réelle est de 306.19. L'estimation de la probabilité locale pour un voisinage de 11×11 par krigeage disjonctif (fig. 6.84) a une erreur réelle de 1.81×10^{-4} avec une erreur moyenne de seulement 0.00086. Pour l'estimateur de l'espérance conditionnelle (fig. 6.85) l'erreur réelle est de 2.399×10^{-4} et l'erreur moyenne est de -0.00758. On voit donc que les problèmes du krigeage disjonctif proviennent d'une mauvaise estimation de la loi sous-jacente et non d'un manque de robustesse de la méthode de filtrage. Il faut donc veiller à ce que la variance de l'estimateur de la loi sous-jacente obtenue par régularisation ne diffère pas trop de la variance de l'image sous-jacente avant de l'utiliser dans un krigeage disjonctif.

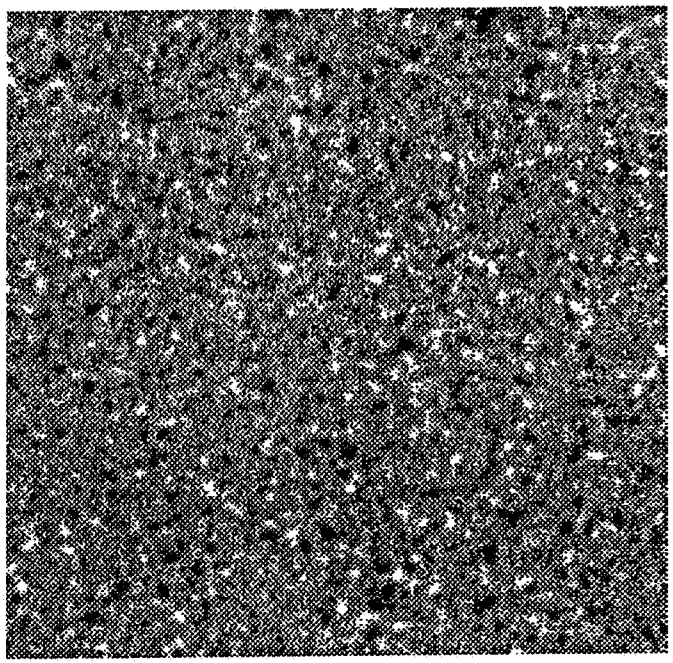


Figure 6.82 Krigeage disjonctif avec la vraie loi.

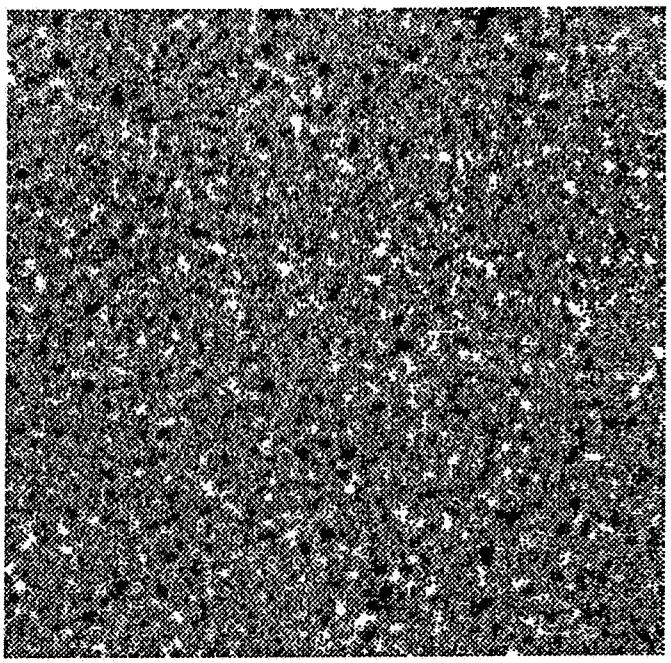


Figure 6.83 Espérance conditionnelle avec la vraie loi.

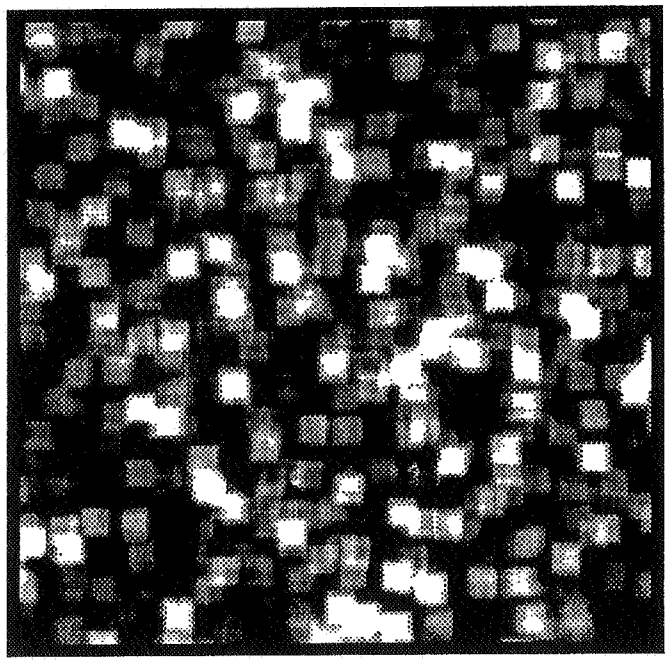


Figure 6.84 Histogramme local, krigeage disjonctif avec la vraie loi.

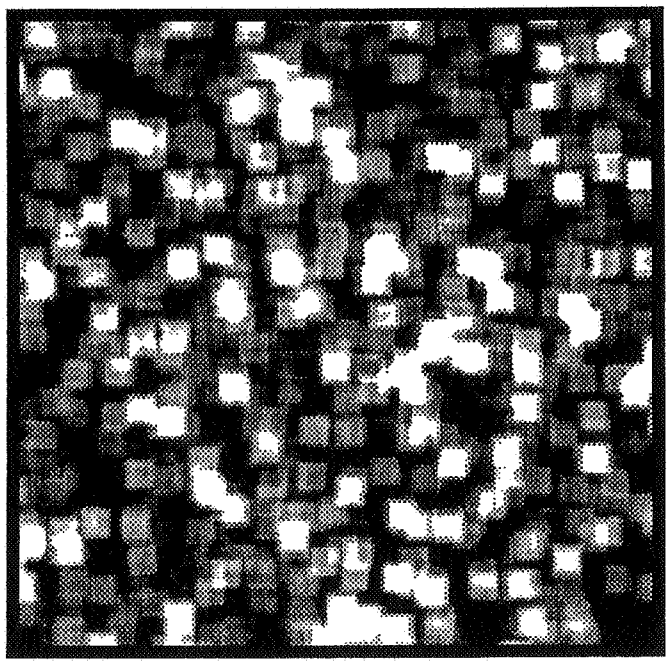


Figure 6.85 Histogramme local, espérance conditionnelle avec la vraie loi.

7. — Nous avons dit que l'estimateur de la probabilité locale peut être utilisé pour obtenir des définitions de taux de ségrégation. Ici nous considérons une mesure possible. Cette définition a été choisie plus pour sa simplicité que pour des raisons physiques, et elle doit être étudiée davantage et comparée avec d'autres définitions avant de décider s'il s'agit d'un bon choix.

Nous avons déjà obtenu un estimateur de la probabilité locale $p_z(y) = \mathrm{P}[Y_x > z | x \in v_y]$. Par conséquent la probabilité qu'un point x choisi au hasard dans v dépasse z est connue. Maintenant nous admettons l'hypothèse que chaque point $x \in v_y$ dépasse z avec la probabilité $p_z(y)$ indépendamment des autres membres de v_y. Cette hypothèse est, en toute rigueur, incohérente avec un modèle de covariance non-pépitique, et de plus elle utilise la loi multivariable dans v_y au lieu des seules lois bivariables, mais puisque l'hypothèse est admise seulement après le calcul de la loi locale, le risque d'erreur important doit être assez faible, et l'on peut admettre cette hypothèse pour les applications. Une vérification plus complète reste à faire.

Avec cette hypothèse, connaissant $p_{z_s}(y)$, on peut maintenant trouver la valeur de $z_s(y)$ telle que

$$P[\bigcup_{x \in v} \{ Y_x > z_s(y) \} \mid x \in v_y] = .95$$

c'est à dire, la probabilité que pour au moins un $x \in v_y$ nous avons $Z_x > z_s(y)$. On trouve que $z_s(y)$ doit être choisi tel que

$$p_{z_s}(y) = 1 - \exp\left(\frac{-\log 20}{N}\right)$$

où N est le nombre de points dans v_y.

De la même manière, on trouve la valeur $p_{z_i}(y)$ telle que

$$P[\bigcup_{x \in v} \{ Y_x \leq z_i(y) \} \mid x \in v_y] = .95$$

et l'on trouve $z_i(y)$ par

$$p_{z_i}(y) = \exp\left(\frac{-\log 20}{N}\right)$$

On peut donc définir un taux de ségrégation locale par

$$TS(y) = \frac{z_s(y) - z_i(y)}{m_Y}$$

où m_Y est la moyenne globale de $Y(x)$.

Comme exemple, nous reprenons les images simulées gaussiennes de portées 5 et 20 respectivement déjà étudiées dans les exemples 1) et 2). On présente les deux images à nouveau sur les figures 6.86 et 6.87.

Les cartes des taux de ségrégation locale ont été calculées à l'aide de la probabilité locale pour un nombre de seuils différents $\{ z_j \}$ et ensuite en appliquant un test à chaque point pour voir si sous l'hypothèse d'indépendance il y aura au moins un point dans le voisinage qui dépasse le seuil z_j avec la probabilité 0.95.

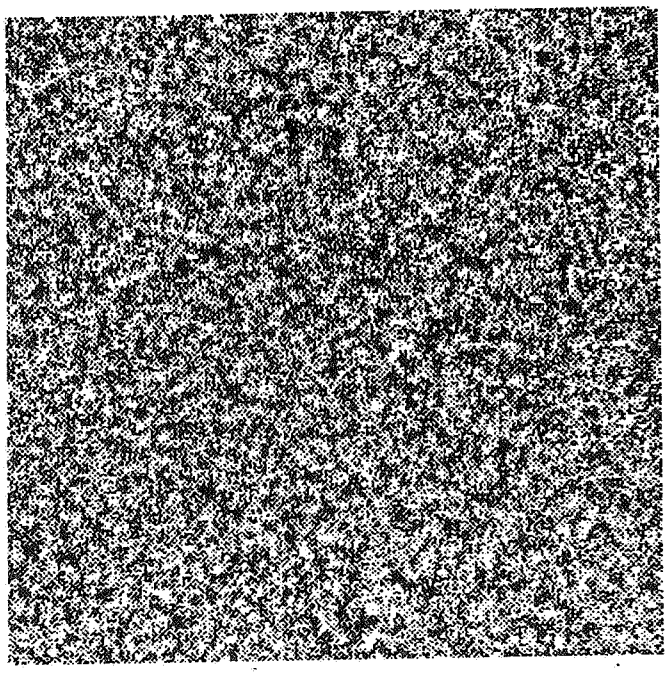


Figure 6.86 Image bruitée, portée 5

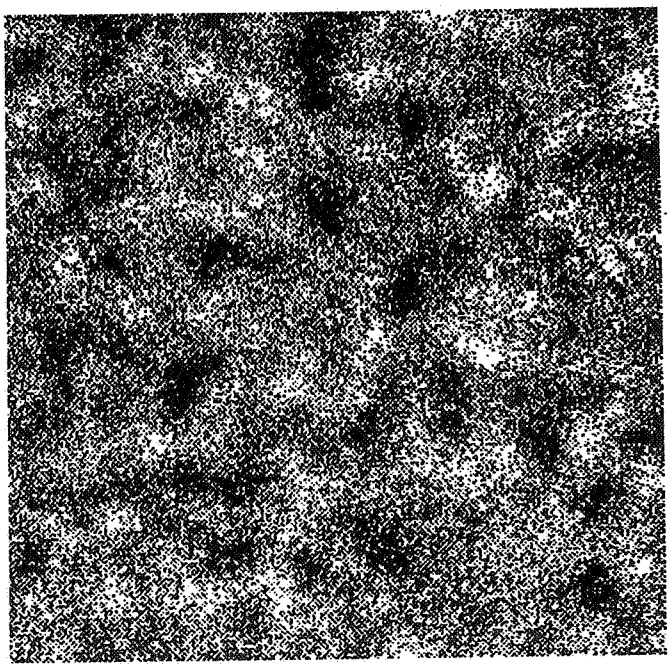


Figure 6.87 Image bruitée, portée 20.

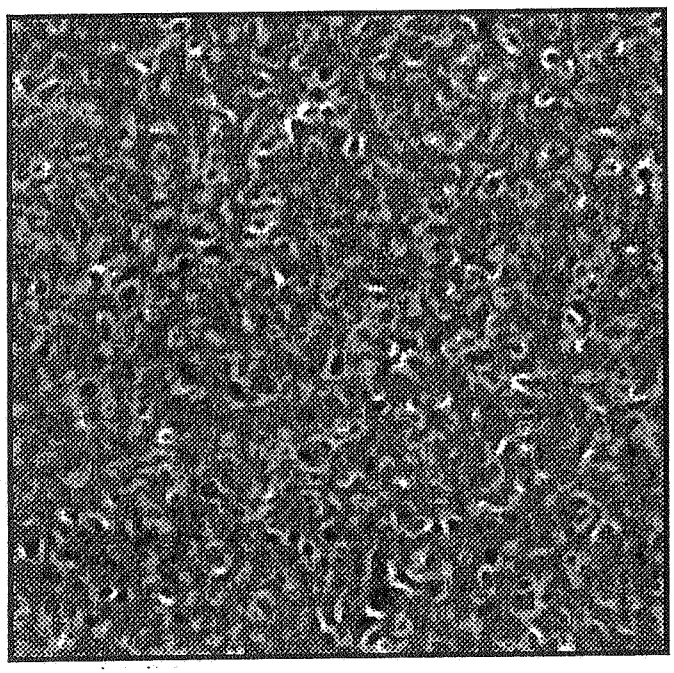


Figure 6.88 Ségrégation locale, portée 5.

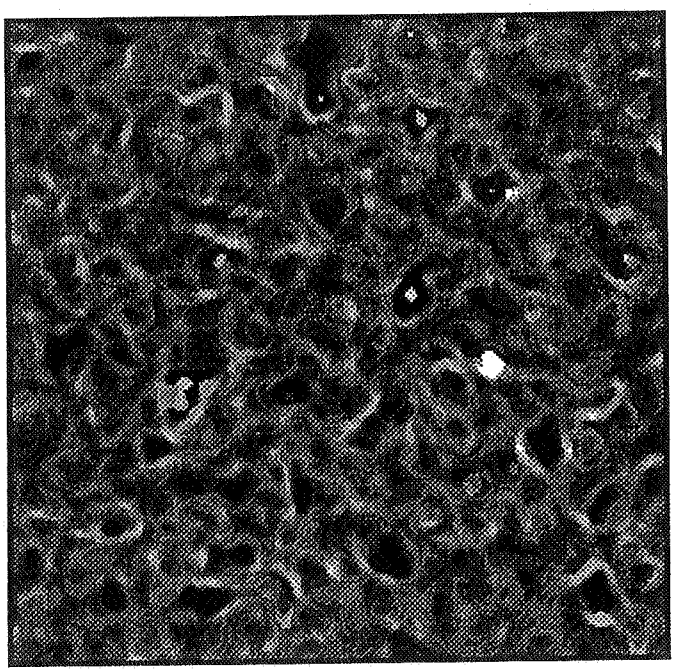


Figure 6.89 Ségrégation locale, portée 20.

A cause d'une contrainte de programmation, il n'a pas été possible dans cette première étude d'avoir des seuils $\{ z_j \}$ équi-repartis, ($|z_{j+1} - z_j|$ = constante), mais on a choisi des écarts ayant la même probabilité. Il était difficile donc de calculer le taux de ségrégation pour des valeurs très fortes ou très faibles. Pour cette raison on voit des "taches" noires sur les deux cartes de ségrégation locale (figures 6.88 et 6.89) où le taux de ségrégation, qui doit être assez élevé, est mis égal à zero. Nous rappelons encore que ce problème est seulement algorithmique, lié au mauvais choix des unités dans le logiciel.

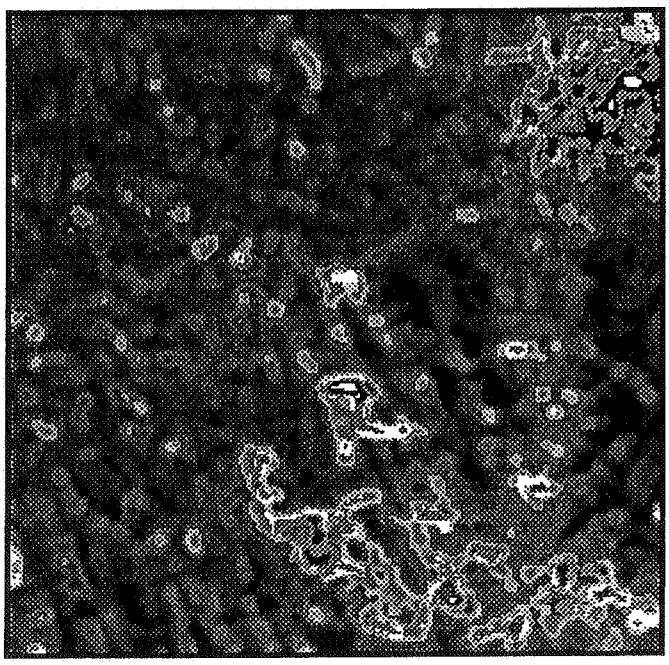


Figure 6.90 Taux de ségrégation calculé par krigeage disjonctif.

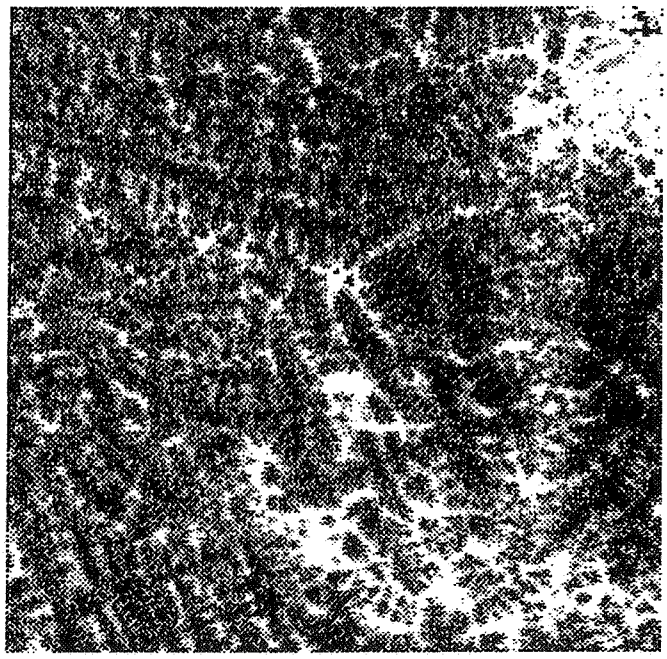


Figure 6.91 Image observée

Sur la figure 6.90 on trouve une application du calcul du taux de ségrégation à l'image réelle déjà étudiée dans l'exemple 4. L'image initiale est montrée sur la figure 6.91. Ici on compare cette définition du taux de ségrégation avec les définitions plus simples $TS_1(y) = \dfrac{\sigma_Y^*(y)}{m_Y^*}$ ou $TS_2(y) = \dfrac{\sigma_Z(y)}{m_Z}$, où σ_Y^* est l'estimateur de l'écart-type local sous-jacent obtenu sur l'image krigée et σ_Z est l'écart-type de l'image bruitée [17]. Les estimations obtenues par TS_1 et TS_2 sont montrées sur les figures 6.92 et 6.93.

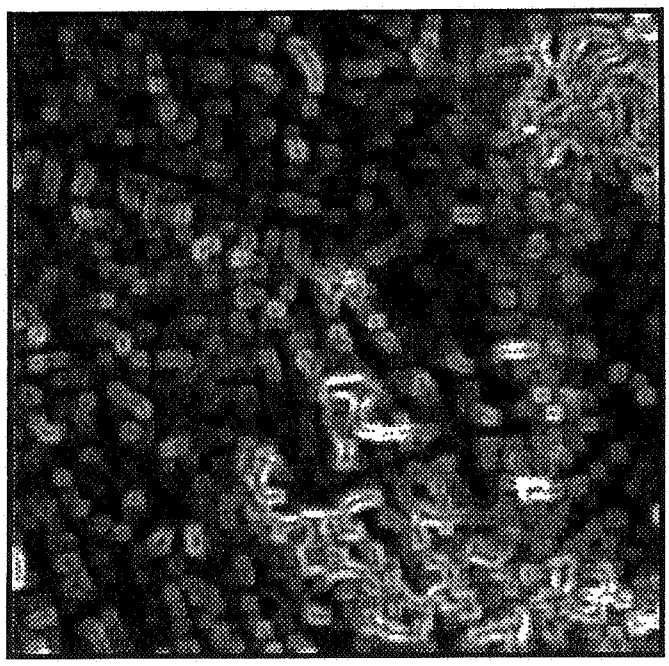


Figure 6.92 Taux de ségrégation calculé sur l'image filtrée.

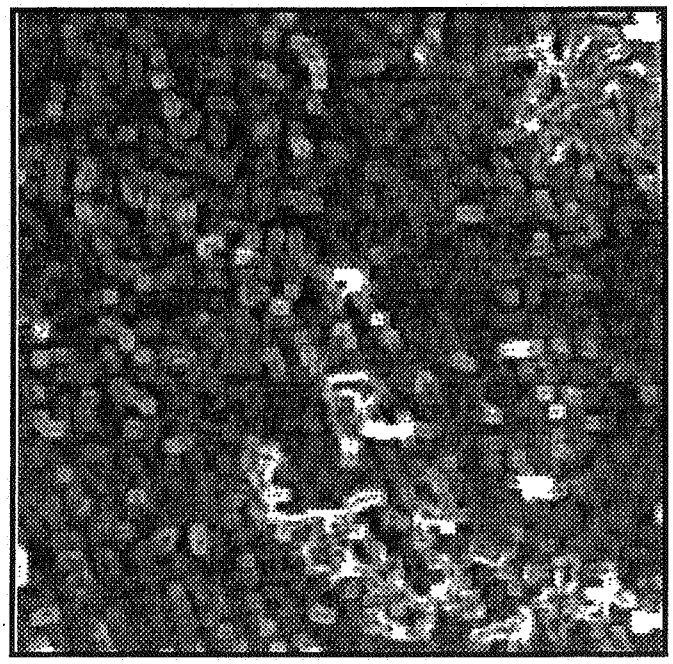


Figure 6.93 Taux de ségrégation calculé sur l'image bruitée.

Malgré un bruit faible pour cette image on voit que ces estimateurs donnent des valeurs très différentes. Plus généralement il semble clair donc qu'on ne peut pas comparer les cartes de taux de ségrégation locales pour des images avec des niveaux de bruit différents avec les mesures TS_1 et TS_2. Par contre on a vu que le krigeage disjonctif donne des estimateurs de la probabilité locale avec beaucoup moins de biais, et par conséquent des estimateurs basés sur le krigeage disjonctif, tel que celui étudié dans cet exemple semblent plus adaptés à l'estimation d'une carte de ségrégation locale.

CHAPITRE VII

CONCLUSION

Au cours de ce travail, nous avons étudié trois problèmes de filtrage. A partir d'une image bruitée, nous avons estimé la variable sous-jacente $Y(x)$, la loi de probabilité globale, et la loi de probabilité locale.

Nous avons appliqué en premier la méthode du krigeage linéaire pour estimer la variable sous-jacente. Nous avons trouvé que cette méthode donnait de bons estimateurs pour une grande variété d'images, même quand la loi sous-jacente était assez éloignée d'une loi gaussienne. Dans tous les cas simulés que nous avons considérés, la variance d'estimation donnée par le krigeage linéaire était en bon accord avec l'erreur réelle. Le seul cas où le krigeage linéaire ait donné de mauvais résultats a été celui où l'image sous-jacente était très éloignée d'une loi gaussienne (exemple 6, Chapitre VI), et même alors, les estimateurs non linéaires n'ont pas sensiblement amélioré l'estimation de la variable elle-même, à cause de la mauvaise estimation de la loi sous-jacente dans un cas aussi extrême. Le filtre linéaire se révèle donc être un estimateur robuste et fiable qui peut s'appliquer loin de son "territoire naturel", le cas gaussien stationnaire. Nous avons également étudié d'autres applications du filtrage linéaire, telles que l'estimation des composantes principales associées à une structure particulière du variogramme ou l'estimation d'un élément non observé à partir des éléments observés, tout en notant que, du moins en principe, ces variables peuvent aussi être estimées par des méthodes non linéaires.

Le problème de l'estimation de l'histogramme sous-jacent d'un histogramme obtenu par randomisation est souvent difficile à résoudre en pratique. Dans beaucoup de cas il n'existe pas de solution unique, et même quand celle-ci existe, le problème est souvent mal posé. Tel était le cas du problème de poissonisation traité dans cette étude. Nous avons utilisé la technique de régularisation pour l'estimer. Un problème de régularisation sans contrainte admet une interprétation duale en termes de problème de krigeage. Bien que nous n'ayons pas pu exploiter pleinement cette dualité d'abord parce que notre problème présentait des contraintes de positivité, et ensuite parce que nous avons travaillé en voisinage unique et que par conséquent les procédures d'inférence des covariances généralisées ne s'appliquaient pas, l'équivalence est intéressante dans la mesure

où elle nous permet, au moins qualitativement, de choisir le type de fonction de régularisation. Par exemple, si on ne dispose que de peu de données, l'histogramme sous-jacent est lui-même assez bruité et un effet de pépite semble être le meilleur choix de fonction de régularisation. L'équivalence avec le krigeage nous permet aussi d'obtenir une justification heuristique du choix d'une norme de type χ^2 pour les erreurs. La méthode de régularisation donne souvent de bons résultats, mais elle n'est pas vraiment capable de traiter le cas où la fonction de répartition sous-jacente est approximativement un mélange de lois discrètes et continues. L'estimation dans ce cas semble assez difficile à faire si l'on ne dispose pas de quelques informations *a priori* sur la forme de la loi sous-jacente.

La technique du krigeage disjonctif a été appliquée ici pour l'estimation des lois de probabilités locales. Il semble que dans le cas général où l'on a une loi sous-jacente et une forme de randomisation données, en toute rigueur seules les équations générales du krigeage disjonctif peuvent être utilisées pour résoudre les problèmes de filtrage. Nous n'avons étudié qu'un seul exemple simple dans ce travail. L'utilisation des modèles isofactoriels, qui sont théoriquement inadaptés sauf dans quelques cas bien précis; semble être suffisamment robuste pour donner de bons estimateurs non biaisés de la loi de probabilité locale. L'espérance conditionnelle apparaît quant à elle plus sensible aux petits perturbations que le krigeage disjonctif.

RÉFÉRENCES

Chapitre 0

[1] KOLMOGOROV A.N. (1939), *Sur l'interpolation et l'extrapolation des suites stationnaires*, C.R. Acad. Sci. Paris, 208, 2043–2045.

[2] WIENER N. (1949), *Extrapolation, interpolation and smoothing of stationary time series*, MIT Technology Press and John Wiley, N.Y.

[3] MATHERON G. (1989), *Estimating and choosing*, Springer Verlag, Berlin.

[4] GEMAN S. and GEMAN D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell.*, 6, 721–741.

Chapitre I

[1] WIENER N. (1949), *Extrapolation, interpolation and smoothing of stationary time series*, MIT Technology Press and John Wiley, N.Y.

[2] DALY C., LAJAUNIE Ch., MUIR K. R., OLIVER M.A., WEBSTER R. (1992), Detecting spatial structure in rare human disease: a case study of childhood cancer in the West Midlands of England, *à paraître.*

[3] KURZ W. and FISHER D.J. (1989), *Fundamentals of solidification* (3ème édition), Trans. Tech. Publications, Switzerland.

[4] FELLER W. (1966), *An introduction to probability theory and its applications*, Vol. II, John Wiley & Sons, N.Y.

[5] MATHERON G. (1970), *La théorie des variables régionalisées et ses applications*, Fasc. 5, Cahiers du Centre de Morphologie Mathématique, ENSMP, Paris.

[6] MATHERON, G. (1989), *Estimating and choosing*, Springer Verlag, Berlin.

[7] DELFINER P. (1982), *The intrinsic model of order k*, Ecole d'Eté 1982, Centre de Géostatistique, Fontainebleau.

[8] MATHERON G. (1973), The intrinsic random functions and their applications, *Adv. App. Prob.*, 5, pp. 439–468.

[9] PARIENTI V., MASSOL J., BIENVENU Y. (1988), Influences des conditions de solidification sur la ségrégation et la morphologie de l'interface solide–liquide dans un superalliage base nickel de fonderie, *Mem. Et. Sci. Rev. Mét,* p 514.

Chapitre II

[1] SERRA J. (1982), *Image analysis and mathematical morphology*, Vol. 1, Acad. Press, London.

[2] GEMAN S. and GEMAN D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell.*, 6, 721–741.

[3] BESAG J. (1986), On the statistical analysis of dirty pictures, *JR Statist. Soc.* B48, pp. 259–302.

[4] HASLETT J., HORGAN G. (1987), Linear models in spatial discriminant analysis, Proc. NATO ASI *Pattern Recognition Theory and Applications*, Springer–Verlag, Berlin.

[5] MATHERON G. (1970), *La théorie des variables régionalisées et ses applications*, Fasc. 5, Cahiers du Centre de Morphologie Mathématique, ENSMP, Paris.

[6] MATHERON G. (1982), Pour une analyse krigeante des données régionalisées, Centre de Géostatistique, ENSMP, Fontainebleau.

[7] WACKERNAGEL H. (1985), *L'inférence d'un modèle linéaire en géostatistique*, Thèse de Doct.-Ing., ENSMP, Paris.

[8] BÉHAXETÉGUY J.-P., LAJAUNIE Ch. (1989), Elaboration d'un programme d'ajustement semi-automatique d'un modèle de corégionalisation, Centre de Géostatistique, ENSMP, Fontainebleau.

[9] GOULARD M. (1989), Inference in a coregionalization model, in Proc. 3rd Intern. Geostatistics Congress, Sept. 5–9, 1988, Avignon, *Geostatistics*, ed. by M. Armstrong, Vol. 1, Kluwer Academic Publ., pp. 397–408.

[10] DALY C., LAJAUNIE Ch., JEULIN D. (1989), Application of multivariate kriging to the processing of noisy images, D° pp. 749–760.

[11] DALY C. (1990), Amélioration d'images X par filtrage multivariable, Chapitre E, *Traitement d'images en microscopie à balayage et en microanalyse par sonde électronique*, ANRT, Paris.

[12] HASLETT J. (1989), Geostatistical neighbourhoods and subset selection, in Proc. 3rd Intern. Geostatistics Congress, Sept. 5–9, 1988, Avignon, *Geostatistics*, ed. by M. Armstrong, Vol. 2 , Kluwer Academic Publ., pp. 569–578.

[13] JEULIN D., KURDY M.B., DALY C., (1990), Etude de l'influence de la rugosité sur l'aspect de surface de tôles peintes, Centre de Morphologie Mathématique, ENSMP, Fontainebleau, *résultats non-publiés*.

[14] DALY C., JEULIN D., KURDY M.B., Monteil G., Etude de l'influence de la rugosité sur l'aspect de surface de tôles peintes, *publication en préparation*.

Chapitre III

[1] TIKHONOV A. et ARSÉNINE V. (1976), *Méthodes de résolution de problèmes mal posés*, Ed. Mir, Moscou.

[2] GALLI A., MEUNIER G. (1987), Study of a gas reservoir using the external drift method, in *Geostatistical Case Studies*, ed. by M. Armstrong and G. Matheron, Reidel Publ. Co., pp. 105–119.

[3] O'SULLIVAN F. (1986), Ill posed inverse problems, *Statistical Science*, Vol. 1, n° 4, pp. 502–519.

[4] COOK D.R. and WEISBERG S. (1982), *Residuals and influence in regression*, Chapman and Hall, N.Y.

[5] CRAVEN P. and WAHBA G. (1979), Smoothing noisy data with spline functions, *Numer. Math.* 31, pp. 377–403.

[6] HUTCHINSON M.F. and DEHOOG F.R. (1985), Smoothing noisy data with spline functions, *Numer. Math.* 47, pp. 99–106.

[7] DELFINER P. (1982), The intrinsic model of order k, Ecole d'Eté 1982, Centre de Géostatistique, Fontainebleau.

[8] JEULIN D. and RENARD D. (1991), Practical limitations of deconvolution kriging, *à paraître.*

Chapitre IV

[1] BRÉZIS H. (1983), *Analyse fonctionnelle*, Masson, Paris.

[2] ALHEIZER N.I. et Glazman I.M. (1966), *Theory of linear operators in Hilbert space*, Vol. 1, Frederick Ungar Publish. Co., N.Y.

[3] FELLER W. (1966), *An introduction to probability theory and its applications*, Vol. II, John Wiley & Sons, N.Y.

[4] CONWAY J.B. (1973), *Functions of one complex variable*, Springer Verlag, N.Y.

[5] RUDIN W. (1966), *Real and complex analysis*, McGraw-Hill, N.Y.

[6] MENDELSOHN J. and RICE J. (1982), Deconvolution of microfluorometric histograms with B splines, *JASA* Vol. 77, no. 380, pp. 748–753.

[7] De BOOR C. (1978), *A practical guide to splines*, Springer–Verlag, N.Y.

[8] JEULIN D. and RENARD D. (1991), Practical limitations of deconvolution kriging, *à paraître.*

Chapitre V

[1] TIKHONOV A. et ARSÉNINE V. (1976), *Méthodes de résolution de problèmes mal posés*, Ed. Mir, Moscou.

[2] BRÉZIS H. (1983), *Analyse fonctionnelle*, Masson, Paris.

[3] YOSIDA K. (1968), *Functional analysis*, Springer-Verlag, N.Y.

[4] DIEUDONNÉ J. (1969), *Eléments d'analyse*, Vol. 1, Gauthier-Villars, Paris.

[5] De BOOR C. and LYNCH R.E. (1966), On splines and their minimum properties, *J. Math. Mech.* 15, 953-969.

[6] ATTEIA M. (1966), *Etude de certains noyaux et théorie des fonctions splines en analyse numérique*, Publications de l'Institut de Mathématiques Appliquées de Grenoble.

[7] KIMMELDORF G.S. et WAHBA G. (1971), Some results on Tchebycheffian spline functions, *Math. Anal. Appl.* 33, 82-95.

[8] DUCHON J. (1976), Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces, *RAIRO Analyse Numérique*, 10, 12, pp. 5-12.

[9] MATHERON G. (1981), Splines and kriging, their formal equivalence, in *Down to Earth Statistics Solutions Looking for Geological Problems*, Syracuse University Geology Contributions.

[10] MATHERON G. (1973), The intrinsic random functions and their applications, *Adv. Appl. Prob.*, 5, pp. 439-468.

[11] GALLI A., MEUNIER G. (1987), Study of a gas reservoir using the external drift method, in *Geostatistical Case Studies*, ed. by M. Armstrong and G. Matheron, Reidel Publ. Co., pp. 105-119.

[12] LAURENT P. (1972), *Approximation et optimisation*, Hermann, Paris.

[13] DUBRULE O. et KOSTOV C. (1985), An interpolation method taking into acount inequality constraints, *Math. Geol.*, Vol. 18, N° 1.

Chapitre VI

[1] MATHERON G. (1973), Le krigeage disjonctif, Centre de Géostatistique, ENSMP, Fontainebleau.

[2] JEULIN D. (1991), *Modèles morphologiques de structures aléatoires et de changement d'échelle*, Thèse de Docteur ès Sciences Physiques.

[3] SERRA J. (1982), *Image analysis and mathematical morphology*, Vol. 1, Acad. Press, London.

[4] SERRA J. (1988), *Image analysis and mathematical morphology*, Vol. 2, Acad. Press, London.

[5] MATHERON G. (1982), La déstructuratuion des hautes teneurs, Centre de Géostatistique, ENSMP, Fontainebleau.

[6] MATHERON G. (1976), A simple substitute for conditional expectation : the disjunctive kriging, in Proc. NATO ASI, *Advanced Geostatistics in the Mining Industry*, M. Guarascio et al. Ed., D. Reidel Publ. Co., Dordrecht, Holland, pp. 221–236.

[7] MATHERON G. (1984), Isofactorial models and change of support, in Proc. 2nd NATO ASI, *Geostatistics for Natural Resources Characterization*, G. Verly et al. eds., Reidel Publ. Co., Dordrecht, Holland, Part 1, pp. 449–467.

[8] MATHERON G. (1984), Une méthodologie générale pour les modèles isofactoriels discrets, *Sciences de la Terre*, Nancy, n° 21, pp. 1–64.

[9] LAJAUNIE Ch. and LANTUÉJOUL Ch. (1989), Setting up the general methodology for discrete isofactorial models, in Proc. 3rd Intern. Geostatistics Congress, Sept. 5–9, 1988, Avignon, *Geostatistics*, M. Armstrong ed., Vol. 1, Kluwer Academic Publ., pp. 323–334.

[10] MATHERON G. (1984), Changement de support en modèle mosaïque, *Sciences de la Terre*, n° 18, pp. 71–123.

[11] RIVOIRARD J. (1988), Modèles à résidus d'indicatrices autokrigeables, *Sciences de la terre*, Série Inf. No 28.

[12] RIVOIRARD J. (1989), Models with orthogonal indicators, in Proc. 3rd Intern. Geostatistics Congress, Sept. 5–9, 1988, Avignon, *Geostatistics*, M. Armstrong ed., Vol. 1, Kluwer Academic Publ., pp. 91–108.

[13] MATHERON G. (1975), Compléments sur les modèles isofactoriels, Centre de Géostatistique, ENSMP, Fontainebleau.

[14] HU L.Y. (1988), *Mise en oeuvre du modèle gamma pour l'estimation des distributions spatiales*, Thèse de Doct.-Ing. en Géostatistique, ENSMP, Paris.

[15] LANTUÉJOUL Ch. et RIVOIRARD J. (1984), Une méthode de détermination d'anamorphose, Centre de Géostatistique, ENSMP, Fontainebleau.

[16] DEMANGE C., LAJAUNIE Ch., LANTUÉJOUL Ch., RIVOIRARD J. (1987), Global recoverable reserves : testing various change of support models in uranium data, in *Geostatistical Case Studies*, ed. by M. Arsmtrong and G. Matheron, Reidel Publ. Co., Dordrecht, Holland.

[17] DALY C., JEULIN D., BENOIT D., AUCLAIR G. (1990), Application of Multivariate Geostatistics to Macroprobe Mappings in Steels, *ISIJ International*, 30, No. 7, pp 529–534.

[18] BESAG J., YORK J., MOLLIÉ A. (1991), Bayesian image restoration, with two applications in spatial statistics, *Ann. Inst. Statist. Math.*, Vol 43, No 1, 1–59.

ENSMP
1991