De l'analyse structurale à la modélisation des ressources naturelles : contributions géostatistiques

Jacques Rivoirard

▶ **To cite this version:**

Jacques Rivoirard. De l'analyse structurale à la modélisation des ressources naturelles : contributions géostatistiques. Statistiques [math.ST]. Université Pierre et Marie Curie - Paris VI, 2012. tel-00664238

HAL Id: tel-00664238

https://theses.hal.science/tel-00664238

Submitted on 30 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie – Paris 6

De l'analyse structurale à la modélisation des ressources naturelles: contributions géostatistiques

Jacques Rivoirard

Mines-ParisTech, Centre de Géosciences

Habilitation à Diriger des Recherches

Spécialité Sciences de la Terre et de l'Univers

11 Janvier 2012

Jean-Noël Bacro	Université Montpellier 2	Rapporteur
Jean-Paul Chilès	Mines-ParisTech	Examinateur
Denis Marcotte	Ecole Polytechnique, Montréal	Rapporteur
Ghislain de Marsily	Université Pierre et Marie Curie	Examinateur
Pascal Monestiez	INRA, Avignon	Rapporteur
Pierre Petitgas	IFREMER, Nantes	Examinateur

De l'analyse structurale à la modélisation des ressources naturelles: contributions géostatistiques

Résumé

Mes travaux de recherche s'inscrivent dans le développement d'outils géostatistiques pour mieux décrire, comprendre et modéliser la distribution spatiale de ressources naturelles : ressources minières, pétrolières, halieutiques. J'ai parcouru, chemin faisant, les grands secteurs de la géostatistique que sont l'analyse structurale, l'estimation linéaire par krigeage ou cokrigeage, les méthodes non-linéaires, et enfin les simulations.

Toute étude géostatistique commence par l'analyse structurale, dans laquelle on cherche à mettre en évidence la structure spatiale des variables en jeu, à l'aide d'outils structuraux tel le variogramme. J'ai étudié de façon empirique l'influence déterminante que pouvait avoir le choix du support de travail (surface ou volume élémentaire sur lequel est mesurée la variable), ainsi que l'émergence difficile d'une structure variographique. En halieutique, l'approche transitive (basée sur un échantillonnage à maille régulière ignorant les frontières), et certaines statistiques apparentées, se révèlent appropriées pour décrire des populations spatiales possédant quelques valeurs très fortes et des frontières diffuses.

L'analyse structurale est naturellement cruciale lorsqu'il s'agit de mettre en évidence des liens structuraux entre variables, notamment hiérarchiques. Ainsi la taille des harengs plutôt que leur age, au Nord de l'Ecosse, ou la proportion de minerai plutôt que le métal, dans un gisement d'uranium en petites veines, apparaissent-elles comme pré-éminentes. Le croisement de variables, et les méthodes d'estimation associées (cokrigeage, analyse krigeante), se révèlent particulièrement adaptés pour filtrer des séries de données, sismiques par exemple.

Les techniques de cartographie que sont krigeage et cokrigeage posent en pratique un problème majeur, celui du choix du voisinage utilisé pour faire l'estimation, c'est-à-dire le sous-ensemble des données qui est utilisé pour estimer la valeur inconnue en un point. Un voisinage trop petit dans l'estimation de blocs miniers s'accompagne d'un biais conditionnel, responsable d'une surestimation des blocs considérés comme riches. J'ai proposé des outils permettant de mieux choisir le voisinage. Par ailleurs, l'utilisation d'un voisinage glissant peut générer des discontinuités indésirables : une solution générale a été proposée. Enfin je me suis intéressé aux conditions dans lesquelles certaines versions simplifiées de cokrigeage, par exemple le cokrigeage collocalisé, ne s'accompagnent pas de perte d'information.

Les problèmes de dépassement de seuils nécessitent, pour le moins, le recours à la géostatistique non-linéaire, basée sur des transformations de la variable en gaussienne ou indicatrices, par exemple. Dans certaines circonstances, le cokrigeage d'indicatrices s'obtient par krigeage de résidus d'indicatrices. Un modèle d'écrêtage permet une estimation adaptée à la présence de valeurs extrêmes, comme dans les gisements d'or. Par ailleurs, une méthode d'ajustement d'histogramme a été développée, permettant en particulier de combler de manière controlée les queues de distribution insuffisamment informées. En mine, la technique dite de conditionnement uniforme permet de prédire la distribution des valeurs de blocs à l'intérieur d'un panneau minier à partir de la seule teneur de ce panneau : la méthode a été étendue au cas multivariable.

Les simulations géostatistiques permettent de représenter des phénomènes dans leur variabilité spatiale, et donc d'estimer des quantités complexes. Ainsi a-t-on pu évaluer les ressources récupérables en nodules polymétalliques dans le pacifique Nord, compte tenu de contraintes de pente. De même, l'incertitude sur l'estimation de l'abondance du hareng au Nord de l'Ecosse a pu être chiffrée en combinant différentes sources d'incertitude. Il a fallu pour cela adapter la classique méthode de simulation transformée gaussienne aux problèmes de valeurs nulles en grand nombre. Quant au modèle résultant du seuillage d'une gaussienne, il permet la simulation de faciès de réservoirs hétérogènes, conditionnée par des données de faciès aux puits. Une technique a été imaginée afin de rendre une telle simulation cohérente avec les fluides observés aux puits. Avec le modèle booléen, les corps sableux sont considérés comme des objets implantés aléatoirement et indépendamment : contraignant dans le cas stationnaire, le modèle est beaucoup plus souple en présence de non-stationnarités, comme c'est généralement le cas des réservoirs hétérogènes. Cependant les simulations génétiques, simulant la genèse des formations géologiques, sont plus adaptées lorsqu'on cherche à reproduire par exemple la géométrie et l'agencement des corps sédimentaires complexes liés aux réservoirs chenalisés méandriformes. Des méthodes de conditionnement ad hoc et l'usage de formules heuristiques s'avèrent alors fort utiles.

Remerciements

Tout d'abord je tiens à remercier Denis Marcotte, Jean-Noël Bacro et Pascal Monestiez, pour avoir accepté bien volontiers la tâche de rapporteurs de cette HDR. Je remercie également Pierre Petitgas, Jean-Paul Chilès et Ghislain de Marsily pour leur participation au jury au titre d'examinateurs : Pierre Petitgas, qui m'a introduit au domaine de l'halieutique ; Jean-Paul Chilès, auteur avec Pierre Delfiner de l'ouvrage de référence en géostatistique, et auquel j'ai l'honneur de succéder à la tête de l'équipe géostatistique de Fontainebleau ; et enfin Ghislain de Marsily, président de ce jury et grâce à qui j'ai entrepris la rédaction de ce mémoire. Des années après être venu à Fontainebleau inciter les chercheurs à préparer une HDR, il a reçu de ma part un mémoire auquel il a bien voulu prêter une attention immédiate. Je remercie au passage Alain Tabaggh, Dominique Vassiliadis, et Laura, pour leur discrète efficacité dans le bon déroulement des opérations.

Résumer une trentaine d'années de recherche est un travail intéressant, qui donne l'occasion de revenir sur ses pas, et de saluer les rencontres ayant jalonné ce parcours. Au vu de mon CV (études d'ingénieur à l'Ecole des Mines de Paris, et membre d'un centre de recherche de cette même école depuis plus de 30 ans…), on pourrait croire à un parcours linéaire et bien tracé. C'est pourtant par un chemin détourné que je me suis mis à la géostatistique, et encore plus à la recherche. Alors à l'Ecole, saturé de mathématiques, j'avais choisi l'option « sciences de la terre ». Je me suis ensuite retrouvé sur le terrain, en Nouvelle-Calédonie, pour une durée déterminée au BRGM. C'est là, à l'occasion des visites du géostatisticien du BRGM, que j'ai entrevu ce qu'était la géostatistique, et que j'ai décidé de donner suite à la proposition que m'avait faite Pierre Chauvet, de venir faire de la géostatistique à Fontainebleau.

J'avais eu Pierre Chauvet à l'Ecole en petites classes de probabilités, et il faut dire que l'enseignement des probabilités à l'Ecole était extraordinaire. L'amphi de Philippe Formery était toujours plein, et les petites classes qui suivaient nous donnaient l'occasion d'exercer la finesse et la souplesse des probabilités pour résoudre des problèmes et des paradoxes plus divertissants les uns que les autres. La géostatistique se situait dans la lignée de ces probabilités.

Je me suis donc retrouvé au Centre de Géostatistique de l'Ecole des Mines à Fontainebleau, dirigé par Georges Matheron, fondateur de la discipline. Après un apprentissage accéléré de la géostatistique en deux semaines d'école d'été, je me suis retrouvé à faire des études de gisements miniers, d'uranium en particulier. Alors même que la géostatistique vise à représenter la variabilité spatiale des variables étudiées, j'étais émerveillé de voir apparaitre des structures variographiques sur des données qui me semblaient sans espoir. Ainsi, pendant quelques années d'un travail de bureau d'études, ai-je réellement appris, pour ainsi dire sur le tas, la géostatistique. Cependant, les questions ne manquaient pas, et c'est en y réfléchissant et en essayant d'apporter des solutions, que peu à peu, je me suis mis à faire de la recherche. La géostatistique, mélange de probabilités et de notions de physique, est une discipline difficile, pleine de *tours de main* selon l'expression de Georges Matheron, et encore aujourd'hui, je suis frappé par la présence de sujets de recherche à peine cachés derrière les applications géostatistiques les plus courantes. De mes premières années à Fontainebleau, je conserve le vif souvenir de la liberté dont nous disposions, ainsi que des contacts riches et stimulants que j'ai pu nouer sur le long terme avec nos

partenaires, notamment avec les industriels de Cogéma/Aréva : Henri Sans, Henri Sanguinetti, Claude Demange…

La vie dans un centre de recherche peut être passionnante, et je remercie ici tous mes collègues, anciens comme nouveaux, pour les échanges et les liens que nous avons tissés, tant professionnels qu'amicaux. C'est en particulier grâce à mon ancien collègue Alain Galli que je me suis intéressé au domaine du pétrole. C'est aussi à la faveur de discussions avec Pierre Petitgas, biologiste alors en thèse à Fontainebleau, que j'ai découvert le domaine de l'halieutique. Si la géostatistique est une, chaque domaine possède ses propres spécificités, auxquelles il est intéressant et essentiel de s'adapter. Aujourd'hui comme hier, l'équipe de géostatistique de Fontainebleau travaille dans plusieurs domaines, et c'est cette position si particulière, à la croisée de l'académique et de l'applicatif, qui fait, me semble-t-il, l'originalité des recherches qui y sont menées.

Table des matières

1 Introduction 8

2 L'analyse structurale 12

2.1 L'influence du support de la donnée, exemple de l'acoustique en halieutique 12

2.2 Le délicat calcul des variogrammes de concentrations 18

2.3 L'approche transitive pour le suivi des populations marines 23

2.4 L'analyse structurale multivariable 26

2.4.1 Un modèle global pour le hareng écossais 27

2.4.2 Des modèles adaptés au filtrage de données sismiques 30

2.4.3 Le modèle à résidu, exemple d'une hiérarchie entre minerai et métal 33

3 Krigeage et cokrigeage 37

3.1 Repousser les frontières du voisinage de krigeage ? 37

3.2 Eliminer les discontinuités de cartes krigées 40

3.3 Cokrigeage et autokrigeabilité 42

3.4 Le voisinage de cokrigeage, dans la prise en compte de variables auxiliaires 44

4 Les modèles non-linéaires 48

4.1 L'estimation au support ponctuel et les dépassements de seuils 48

4.2 L'écrêtage dans les mines d'or 51

4.3 Modélisation de l'histogramme : combler les queues de distribution 54

4.4 Le changement de support : du bloc de sélection minier au panneau 55

5 Simulations 58

5.1 Simuler des abondances par les modèles gaussiens 58

5.2 Simuler des corps sédimentaires par gaussiennes seuillées 62

5.3 Les corps sédimentaires comme objets de simulations boolénnes 64

5.4 Modèles génétiques de réservoirs chenalisés méandriformes 66

6 Perspectives 71
6.1 Outils structuraux plus souples, plus généraux 71
6.2 Voisinage de krigeage et cokrigeage 72
6.3 Krigeage Disjonctif empirique 72
7 Références 75
Annexe: publications choisies 84

1 Introduction

Pour l'essentiel, les recherches en géostatistique auxquelles j'ai contribué ont trait à la modélisation des ressources naturelles, qu'il s'agisse de gisements miniers, pétroliers, ou d'abondances de poissons. Un certain nombre d'entre elles émanent de problèmes rencontrés à l'occasion d'une activité que l'on pourrait qualifier de bureau d'études. D'autres s'inscrivent dans de véritables projets de recherche dans un domaine d'application bien précis. Enfin certaines recherches, plus spontanées, sont des avancées méthodologiques, éventuellement oublieuses du domaine d'origine qui a conduit à leur développement.

Beaucoup de mes recherches sont parties du domaine minier, et en particulier de l'évaluation des gisements. Ainsi pour certains problèmes que j'ai rencontrés à mes débuts, lorsque j'ai commencé à faire des études géostatistiques de gisements miniers, d'uranium en particulier. La recherche de structures variographiques sur des données à histogramme fortement dissymétrique se révélait peu sûre, et j'ai finalement été amené à étudier de manière empirique l'émergence de telles structures. De même j'ai été confronté à la manière d'arrêter un voisinage de krigeage, alors que quel que soit celui-ci, les poids des données les plus extérieures du voisinage dominent sur leurs voisins, et alors même que je disposais d'une structure variographique stationnaire avec un beau palier. De mes observations sur les poids de krigeage, j'ai tiré le sujet de ma thèse de docteur-ingénieur, ainsi que des recommandations pour le choix des voisinages. Ce sujet d'intérêt me conduira, 25 ans plus tard, à proposer une solution simple et générale pour éviter les discontinuités du krigeage en voisinage glissant (mais ceci ne concerne déjà plus la mine, où on krige surtout des blocs ; en revanche, c'est important pour les applications en hydrologie, par exemple).

Un sujet de recherche important sur lequel j'ai travaillé est celui de l'évaluation des ressources de gisements d'uranium en petites veines. Adoptant une approche par variables utiles (variables additives minerai et métal), j'ai été conduit à identifier, par l'analyse des variogrammes, ce que j'ai appelé un modèle à résidu : modèle hiérarchique, dans lequel l'une des variables est subordonnée à l'autre. Cette dernière, la variable maîtresse, bénéficie d'heureuses propriétés d'autokrigeabilité : en bref, son estimation par cokrigeage ne fait pas appel à l'autre variable et coïncide donc avec son krigeage propre. Une quinzaine d'années plus tard, ce même type de modèle me permettra de faire la lumière sur les conditions d'optimalité du très populaire cokrigeage colocalisé (dans le milieu pétrolier), ainsi qu'à identifier un certains nombre de modèles multivariables, dans lesquels le cokrigeage peut se simplifier.

Dans la lignée du modèle à résidu, j'ai développé le modèle à résidus d'indicatrices, permettant la prévision des réserves récupérables au-dessus de coupures (ou la cartographie de dépassements de seuils dans d'autres domaines d'application), lorsque les ensembles au-dessus de différentes coupures s'agencent les uns dans les autres sans effets de bord. L'existence ou non de tels effets de bord peut se vérifier sur les variogrammes. De tels outils m'ont permis plus tard d'examiner le problème de l'estimation des gisements d'or, dans lesquels une pratique courante consiste à écrêter plus ou moins arbitrairement les valeurs les plus fortes. Une analyse fine de plusieurs jeux de données m'a alors conduit à proposer un modèle cohérent d'écrêtage, apportant par là même une solution au problème des valeurs fortes, fréquent dans beaucoup de domaines.

Enfin, en mine, je me suis intéressé à la prévision de réserves récupérables locales, et ai étendu au cas de plusieurs éléments la technique connue sous le nom de conditionnement uniforme.

Mes recherches dans le domaine halieutique (ou peut-être devrais-je dire celui des populations marines) se sont essentiellement inscrites, soit dans la problématique de l'évaluation d'une abondance spatiale et de son incertitude, soit dans le suivi de populations, mais toujours à partir de données de campagnes scientifiques.

En ce qui concerne le premier aspect, le domaine halieutique a quelques points en commun avec le domaine minier. En particulier, comme beaucoup de teneurs minières, les densités ou concentrations d'individus présentent des histogrammes très dissymétriques, avec quelques valeurs fortes et beaucoup de valeurs faibles, sinon nulles (effet zéro). Ceci pose problème en modélisation, dans l'un ou l'autre domaine, ce qui m'a conduit à proposer certaines améliorations pour la modélisation des histogrammes, ainsi qu'une manière de traiter l'effet zéro en simulation. Par ailleurs, pour comprendre les contrastes de variabilité et les différentes composantes structurales de densités d'individus en halieutique, j'ai été amené à m'intéresser de près à l'influence du support de la variable, tellement crucial également en mine. Enfin, ce sont les difficultés de l'analyse structurale en halieutique qui m'ont conduit à proposer de pondérer les variogrammes pour améliorer l'inférence des structures.

Cependant, l'étude de la répartition spatiale de beaucoup d'espèces pose un problème de modélisation multivariable particulier : il s'agit en effet de comprendre, de proposer un modèle apte à représenter la distribution des individus dans leur environnement. Un modèle multivariable spécifique a pu être ainsi élaboré pour le hareng écossais, liant la taille des individus à la profondeur du fond à travers les années, et permettant de décliner l'abondance totale et en âge selon les années. La simulation d'un tel modèle conditionnellement aux données de campagnes, permet d'appréhender l'incertitude due à l'échantillonnage spatial.

Du point de vue purement méthodologique, l'aspect le plus spécifique des recherches en halieutique auxquelles j'ai participé a été le recours à la théorie transitive et à d'autres statistiques spatiales de population apparentées, essentiellement en vue de la description et du suivi de populations spatiales. La théorie transitive permet l'estimation d'une abondance et de son incertitude, pourvu que l'on dispose d'un échantillonnage assez régulier. Mais plus généralement, le covariogramme transitif se révèle être un outil structural beaucoup plus robuste que le variogramme, que ce soit au niveau des valeurs fortes de densité, ou des valeurs basses. Les frontières de populations spatiales sont souvent diffuses, et il est alors difficile d'arrêter un champ par la présence de valeurs nulles ou presque, surtout quand en plus ce champ est susceptible de varier selon les âges et les années, comme en suivi de populations. L'approche transitive, qui ne nécessite pas de délimiter un champ, est particulièrement adaptée à de telles situations. Une donnée de densité nulle a une contribution nulle au covariogramme. Dans la même lignée, nous avons développé ou sélectionné un certain nombre d'indices statistiques possédant cette avantageuse propriété, et aptes à résumer et à suivre une population spatiale (conjugués avec des indices biologiques comme la longueur à maturité, ces indices peuvent servir à alerter sur l'état de la population).

En ce qui concerne le domaine pétrolier, j'ai un peu travaillé sur le traitement géostatistique de données sismiques, et en particulier sur la cartographie en présence de croisements problématiques entre profils sismiques. Un modèle simple, issu d'une variographie fine, a permis

d’apporter une solution appropriée. Mais j’ai surtout travaillé en caractérisation de réservoirs, plus précisément sur certains aspects propres à l’usage de tel ou tel modèle de réservoir : modèle gaussien seuillé, modèle booléen, modèle génétique.

Le problème majeur que j’ai rencontré à l’occasion de la modélisation d’un champ de gaz en contexte fluvio-deltaïque par gaussienne seuillée fut le respect des données de fluide. Le modèle gaussien seuillé fournissait une simulation conditionnée par les faciès aux puits, mais non par les fluides aux puits. Une solution ad hoc a été développée pour informer la simulation en fluide en évitant les incohérences (comme par exemple de l’eau au-dessus du gaz). En ce qui concerne le modèle booléen pour représenter par exemple des lentilles sableuses, je me suis intéressé à la validation possible du modèle à partir de données supposées connues le long de puits. Plusieurs statistiques uni-dimensionnelles ont été testées sur simulations booléennes ou non, dont certaines se sont révélées assez efficaces, du moins dans un cas stationnaire. J’ai enfin contribué au développement de modèles génétiques aléatoires de réservoirs chenalisés méandriformes, notamment sur des aspects de conditionnement aux puits et d’outils de prévision facilitant le choix des paramètres.

En dehors de ces problèmes spécifiques liés aux modèles de réservoirs, je m’aperçois que mon approche de la géostatistique, ciblée sur la modélisation, est en pratique très souvent guidée par deux points :

- comment, lors d’études de données réelles, utiliser ou adapter les outils structuraux de la géostatistique de façon à extraire les structures spatiales d’une ou plusieurs variables ;

- comment interpréter les résultats observés de la manière, non pas nécessairement la plus précise, mais la plus simple conceptuellement, et partant la plus féconde, en utilisant la souplesse des modèles probabilistes.

D’une certaine façon, il s’agit d’une recherche de structuration de la connaissance. La géostatistique est une science relativement récente, où méthodes éprouvées voisinent avec recettes de cuisine. Des opérations courantes comme la variographie, ou la recherche d’un voisinage, restent encore trop souvent un art, en même temps que de réels thèmes de recherche affleurent ou émergent dans nombre d’applications géostatistiques plutôt classiques (ainsi le traitement des valeurs fortes). Dans ce contexte, l’émergence ou la résurgence de concepts (effets de bordures, autokrigeabilité…) offre un éclairage nouveau à des pans entiers de la géostatistique.

Bien que les domaines dans lesquels j’ai exercé (mine, pêche, pétrole) soient bien différents les uns les autres, ainsi que le rôle qu’y joue la géostatistique, il existe une unité de la méthodologie géostatistique. Dans la suite de ce mémoire, il m’a paru plus simple de décliner mes activités de recherche, non pas selon le domaine d’application ou l’objectif suivi, mais selon le type de méthodologie. Ainsi la même méthodologie de krigeage peut servir à cartographier une abondance de harengs ou à estimer les teneurs de blocs miniers, qu’il s’agisse d’ailleurs de gros blocs à l’étape de l’exploration minière, ou de petits blocs au moment de leur exploitation. Une amélioration de la méthode, réalisée dans un certain contexte, peut profiter à d’autres situations. Je parcourrai donc successivement ce qui relève de l’analyse structurale (au sens géostatistique), du krigeage ou cokrigeage, des modèles non-linéaires, et des simulations, avant d’envisager quelques perspectives.

Chacun de ces chapitres sera lui-même divisé en sections, chacune des sections correspondant à un sujet particulier. Dans chaque section, je résumerai mes contributions avec un rappel du contexte, en commençant par indiquer les références de celles-ci, et en terminant par les références complémentaires propres à la section. L'ensemble des références du mémoire sera regroupée dans un chapitre supplémentaire. Enfin, le lecteur non spécialisé en géostatistique trouvera l'essentiel des informations voulues dans l'ouvrage de référence :

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

2 L’analyse structurale

L’analyse structurale est une étape clé de toute étude géostatistique. Il s’agit en effet d’extraire la structure spatiale des variables d’intérêt ou de caractériser les relations spatiales entre variables, structure ou relations qui devront être ensuite modélisées pour les étapes suivantes de l’étude géostatistique (par exemple, calcul de variances d’estimation, estimation par krigeage ou cokrigeage, simulation géostatistique). Mais l’analyse structurale est souvent une étape très délicate. Dans le cas monovariable, l’irrégularité ou certaines caractéristiques de l’échantillonnage, la dissymétrie de l’histogramme des données et notamment la présence de valeurs extrêmes, la présence de non-stationnarité peuvent poser problème. Dans le cas multivariable, l’hétérotopie (variables connues sur des ensembles de points différents, même si pas forcément disjoints) et l’infinie richesse des relations structurales possibles rendent l’analyse souvent difficile.

2.1 L’influence du support de la donnée, exemple de l’acoustique en halieutique

Dans cette section, je résume certains résultats relatifs au support d’une variable extraits des sections 4.3 et 4.6 de :

Rivoirard, J., Simmonds, J., Foote, K.G., Fernandes, P., and Bez, N., 2000. Geostatistics for estimating fish abundance. Blackwell Science, Oxford. 206 p.

Un des atouts importants de la géostatistique est la prise en compte du support, c’est-à-dire la surface ou le volume générique (éventuellement réduit quasiment à un point) sur lequel est définie ou mesurée la variable d’intérêt. Le support peut avoir une importance considérable lorsque le support d’échantillonnage est beaucoup plus petit que le support d’intérêt : ainsi en exploitation minière sélective, où les blocs de sélection sont beaucoup plus gros que le support des échantillons (d’où la genèse du krigeage dans les années 50, pour corriger la variabilité excessive des échantillons lorsqu’on veut sélectionner le minerai riche). Mais l’influence du support peut s’avérer considérable également, lorsqu’on passe du support des échantillons initiaux au support d’échantillons composites dont l’usage sera plus pratique pour une estimation des ressources, par exemple.

Lorsque l’on passe d’un support donné à un support multiple, la variable se régularise. Sa variance diminue et le comportement du variogramme près de l’origine se fait plus régulier : l’effet de pépite diminue, et les composantes structurales les plus courtes ont tendance à se faire absorber par la régularisation au plus grand support. Depuis la création de la géostatistique linéaire, il existe d’ailleurs des formules théoriques rigoureuses, permettant par exemple de prévoir l’évolution de la variance, ou de passer d’un variogramme ponctuel à un variogramme régularisé sur un support plus important (Matheron, 1965 ; Journel et Huijbregts, 1978 ; Chilès et Delfiner, 1999).

Malgré tout, en pratique, les choses sont parfois surprenantes. Ainsi la variable peut-elle apparaître paradoxalement beaucoup mieux structurée sur petit support que sur grand support. Dans le domaine halieutique, en particulier, les contrastes d’échelle peuvent être très importants

et l'influence du support très grande. A l'occasion du projet Européen « Geostatistics for fish stock assessment », réalisé avec le Marine Lab d'Aberdeen (Ecosse) et l'Institute for Marine Research de Bergen (Norvège), j'ai pu étudier en détail l'influence du support. Ces résultats ont été documentés dans Rivoirard et al. (2000), section 4.3.

Dans l'exemple présent, les données proviennent d'une campagne acoustique réalisée par le Marine Lab d'Aberdeen sur le hareng écossais en mer du Nord (Simmonds et al., 1997). Le navire effectue des transects d'est en ouest et d'ouest en est, de façon à couvrir la zone à échantillonner (Fig. 1). L'écho acoustique indique la présence du hareng. L'analyse fine des données acoustiques enregistrées le long d'un transect montre les résultats suivants. Ces données ont un histogramme très dissymétrique et le tableau 1 montre les statistiques correspondant au support initial du « ping » enregistré tous les 3 mètres environ le long de la direction du transect parcouru par le navire, ainsi qu'aux supports de régularisation de 10, 100 et 1500 pings. La chute impressionnante de variance s'accompagne de l'apparition de structures incroyablement différentes selon les échelles (Fig. 2). La majeure partie de la variance des pings s'explique par une composante structurale d'environ 4 pings (12 m) de portée. Au support de 10 pings (30 m), la variance ne vaut plus que 27.6% de celle des pings, et la composante structurale précédente a pratiquement été absorbée par la régularisation, ne laissant plus qu'une structure d'allure largement pépitique. Au support de 100 pings (300 m), la variance ne vaut plus que 14% de la précédente, ou 3.9% de celle des pings, la composante pépitique précédente a été très largement consommée, et, outre l'effet de pépite, le variogramme présente une composante structurale de portée 700 pings (2000 m). Cette dernière structure, à son tour, est essentiellement absorbée par la régularisation au support de 1500 pings, soit 4500 pings ou 2.5 milles nautiques. La variance de ce support vaut moins de 1% de celle des pings, et la structure présente, outre un petit effet de pépite, une composante de portée 6000 pings, soit 18 km ou 10 milles nautiques. A noter que c'est ce support de 2.5 milles nautiques qui est utilisé en pratique pour l'évaluation de l'abondance à partir des transects (après transformation de l'acoustique en densité surfacique de poissons selon une formule dont on dira un mot plus loin, section 2.4.1). La variance d'estimation de l'abondance sera de 0.5% de celle du support de 2.5 milles nautiques, soit 0.003% de celle des pings, et conduira à un CV d'estimation d'abondance de 12% environ.

Il n'est peut-être pas inutile de remarquer qu'un tel *support élémentaire* de 2.5 milles nautiques sera considéré comme quasi-ponctuel lors du calcul de la variance d'estimation du champ (de 200 milles nautiques environ de côté) à partir des transects (parallèles, et espacés de 15 milles nautiques). Ce champ pourra être alors discrétisé tous les 2.5 milles nautiques. Cette discrétisation n'est pas qu'une approximation. En effet, un transect peut être (à des effets de bord près) représenté rigoureusement par un ensemble de points tous les 2.5 milles nautiques. Support et discrétisation étant ainsi liés, il sera inutile, au moins dans la direction des transects, de raffiner l'ajustement du modèle de variogramme entre 0 et le pas de 2.5 milles nautiques, comme le permettraient les formules théoriques de régularisation à partir d'un support plus fin. En particulier, l'effet de pépite n'a à ce stade pas à être considéré comme variant inversement proportionnellement à la mesure du support. (Je renvoie au passage le lecteur à une petite note de cours que j'avais rédigée après mes premières années de pratique géostatistique (Rivoirard, 1983), après avoir constaté les écarts entre la théorie géostatistique et sa pratique, au fond une sorte de théorie de la pratique.) Naturellement la discrétisation du champ dans la direction perpendiculaire, et la prise en compte ou non d'une anisotropie aux petites distances au support linéaire de 2.5 milles nautiques, sont affaires d'approximation.

Dans certains cas, le *support élémentaire* est imposé ou presque (ainsi la hauteur – ou un sous-multiple - des blocs de sélection ou des gradins d'exploitation d'un gisement exploité en carrière et reconnu par sondages verticaux). En halieutique, il est plus conventionnel. Cependant on s'attend à ce que la variance d'estimation que l'on en déduit pour l'abondance, par exemple, ne soit pas affectée par son choix. Des études menées sur le merlan bleu, prospecté sur les accores atlantiques par l'Institute of Marine Research de Bergen (Norvège) ont permis de comparer le choix de deux supports de travail pour la donnée acoustique (1 et 5 milles nautiques) (Rivoirard et al. (2000), section 4.6). Le support de 1 mille nautique apparaît mieux structuré, notamment par la présence d'une forte composante de 4 milles nautiques, bien visible au pas de 1 mille, laquelle, régularisée au support de 5 milles, apparaît essentiellement sous forme de pépite au pas de calcul de 5 milles alors considéré, Fig. 3. (Déjà, si l'on prend, pour le support de 1 mille, un schéma sphérique de portée 5 milles et qu'on le régularise aussi à 5 milles, on perd la moitié de la variance initiale, et le variogramme au pas de 5 milles atteint 87% de son palier, Fig. 4). Avec des ajustements directs des variogrammes pour les deux supports, sans même viser leur cohérence théorique, on trouve des CV d'estimation de l'abondance très voisins. Mais alors que ceux-ci valent par exemple 18%, on trouverait respectivement 8 et 15% avec ces deux supports si l'on négligeait la structure spatiale (variance en $1/(n-1) \sim 1/n$, correspondant à des valeurs supposées indépendantes) : on serait alors beaucoup trop optimiste, surtout avec le support de 1 mille nautique. A noter que négliger ainsi la structure spatiale ne va pas toujours dans le sens de l'optimisme, car outre la corrélation entre données, intervient aussi celle entre données et champ (Petitgas et Rivoirard, 1993).

Ainsi, la géostatistique permet de prendre en compte de façon cohérente l'effet du support de la variable. Cependant, la mise en évidence des différentes composantes structurales peut dépendre beaucoup du support utilisé. Du fait de la chute de variance, la régularité de la structure apparente n'augmente pas nécessairement avec le support.

Références

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Journel, A. G., and Huijbregts, Ch. J., 1978, Mining geostatistics, Academic Press, London. x + 600 p.

Matheron, G., 1965, Les variables régionalisées et leur estimation. Masson, Paris. 306 p.

Rivoirard, J., 1983, Remarques pratiques à propos des variances et du variogramme, Note de cours C-99, Centre de Géostatistique, Fontainebleau, France, 11 p.

Petitgas, P., and Rivoirard, J., 1993. Global estimation : s^2/n and the geostatistical estimation variance, in "Report of the Workshop on the Applicability of Spatial Statistical Methods to Acoustic Survey Data", Copenhagen: ICES Cooperative Research Report 195, p.42-60.

Simmonds, E. J., Bailey, M., Toresen, R., Couperus, B., Pedersen, J., Reid, D. G., Fernandes, P. G. & Hammer, C. 1997. 1996 ICES coordinated acoustic survey of ICES Divisions IIIa, IVa, IVb, and VIa. ICES C.M. 1997 /H:11, 13 p.

Tableau 1 : Densité acoustique : variabilité statistique fonction du support (1 ping vaut environ 3 m).

Support	Variance	Variance ratio (%)	CV
1 ping	298000	100.0	17.40
10 pings	82300	27.6	9.15
100 pings	11500	3.9	3.42
1500 pings	2190	0.7	1.46

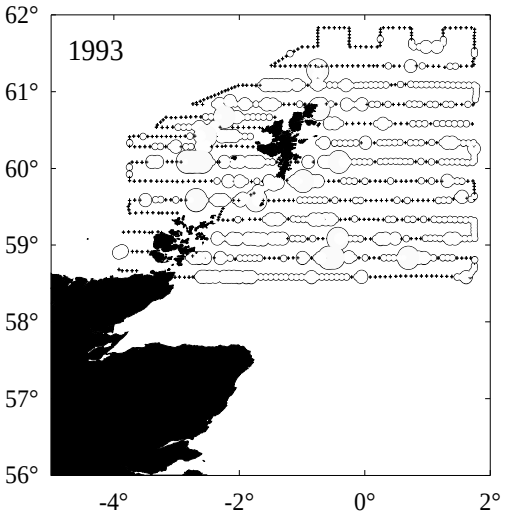
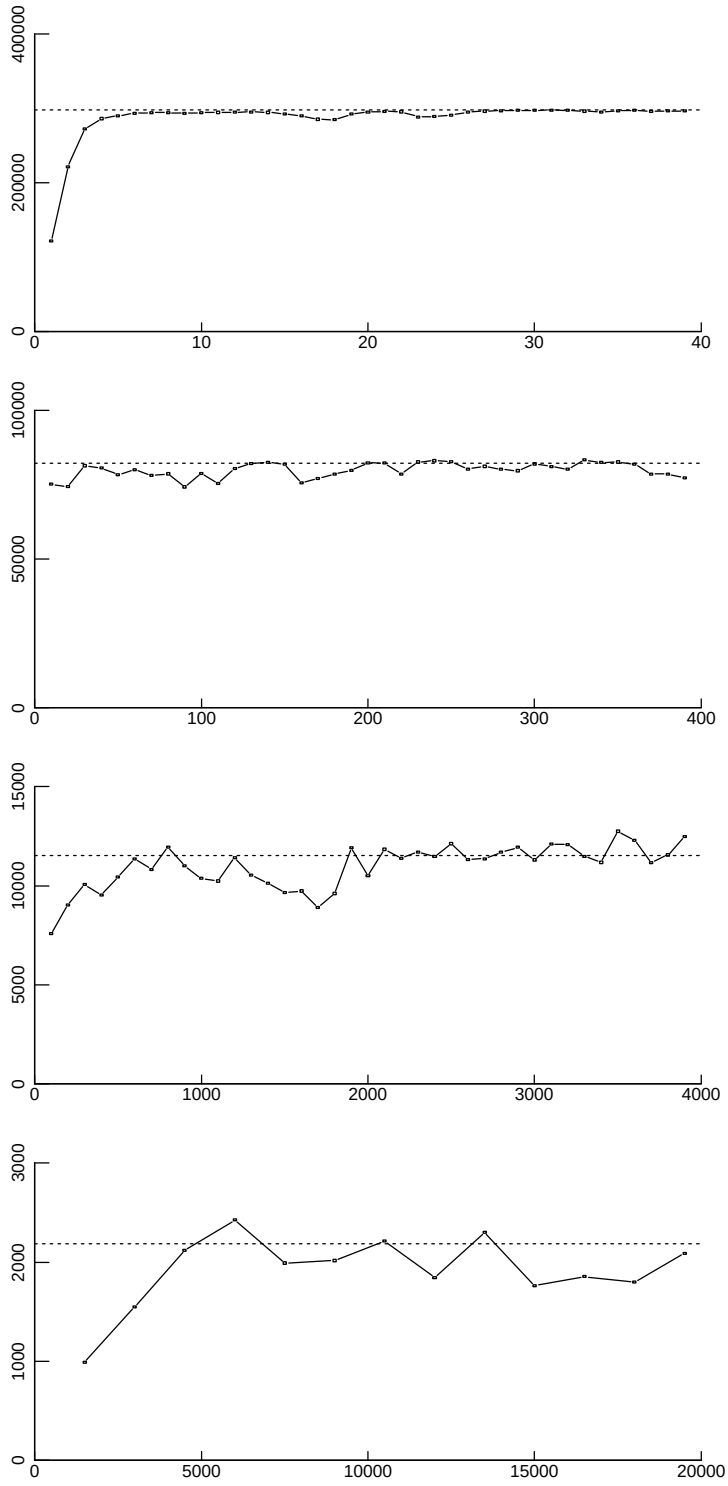
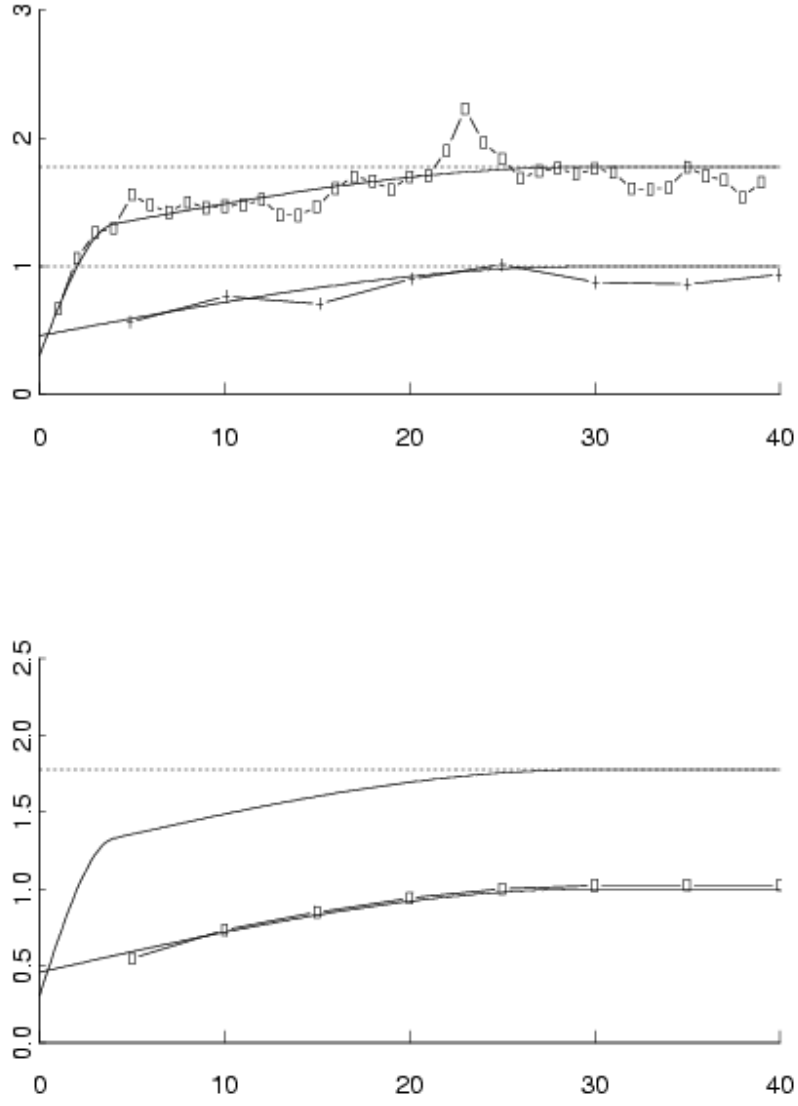


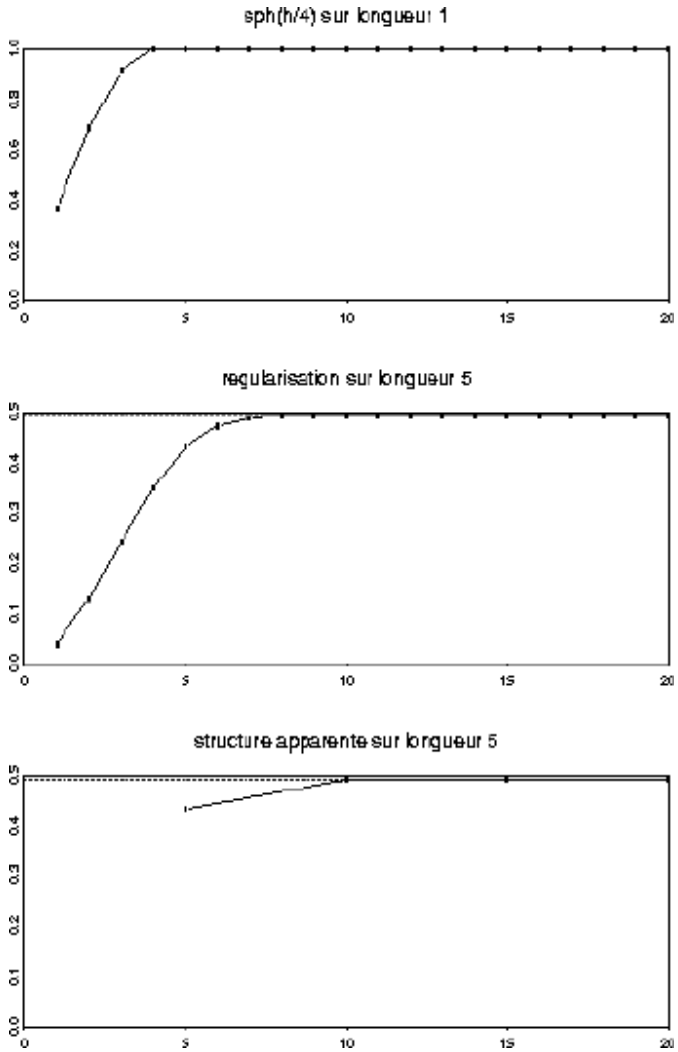
Figure 1. Densité acoustique le long de transects Est-Ouest autour des Shetland. Campagne acoustique sur le hareng de la Mer du Nord. (Extrait de Rivoirard et al., 2000.)



Figures 2. Variogrammes de la densité acoustique au support de (de haut en bas) 1, 10, 100, et 1500 pings (distances en pings, 1 ping vaut environ 3 m). (Extrait de Rivoirard et al., 2000.)



Figures 3. Variogramme de la densité acoustique du merlan bleu. En haut: variogrammes au support de 1 (carrés) et 5 (croix) milles nautiques, ajustés indépendamment. En bas, les deux modèles, ainsi que le régularisée à 5 m.n. (carrés) du modèle à 1 m.n. (Extrait de Rivoirard et al., 2000.)



Figures 4. De haut en bas : variogramme sphérique de portée 4 m.n. pour le support de 1 m.n. ; son régularisée au support de 5 m.n. ; le même aux pas multiples de 5 m.n..

2.2 Le délicat calcul des variogrammes de concentrations

Dans cette section, je donne un résumé des papiers suivants :

Guiblin, P., Rivoirard J., et Simmonds, E. J., 1995, Analyse structurale de données à répartition dissymétrique : exemple du hareng écossais. Cahiers de Géostatistique, 5, ENSMP, 137-159.

Rivoirard, J., 1987 a, Computing variograms on uranium data. In G. Matheron et al. (eds.), Geostatistical case studies. Dordrecht : Reidel, p 1-22.

Rivoirard, J., 2000. Weighted variograms. In: Kleingeld, W.J., Krige, D.G. (Eds.), Geostatistics 2000 Cape Town. Geostatistical Association of Southern Africa, Cape Town, 145–155.

La modélisation d'un variogramme résulte de son interprétation : ainsi dans un cas favorable, on fera passer un modèle théorique de variogramme, croissant sur les premières distances, se stabilisant sur un palier ensuite, à travers les points constituant le variogramme expérimental, en considérant les écarts comme des fluctuations statistiques que l'on peut négliger pour le choix du modèle. Faire la part de ce qui est significatif dans un variogramme expérimental, et de ce qui ne l'est pas, est un exercice délicat, à la base de l'apprentissage de la géostatistique, mais également décisif pour la suite des opérations. La situation est en particulier délicate lorsqu'on travaille avec des données de type « concentration » à histogramme dissymétrique, qu'il s'agisse de teneurs en métal ou, pire, de densités de poisson.

Il arrive que l'on ne « voit » rien, autrement dit que l'on ne décèle aucune structure interprétable. Ce peut être parce qu'il n'y a rien à voir, ou plus exactement que le variogramme, à des fluctuations près, est plat (purement pépitique, proche de l'exemple du hareng Ecossais). En échantillonnage irrégulier, cependant, il peut être risqué de conclure à un effet de pépite pur, et ensuite de prendre la moyenne arithmétique comme estimateur de la moyenne du champ. D'une part le variogramme peut présenter une allure significative à des distances grandes (souvent négligées, probablement à tort, dans le calcul : la règle de la moitié du champ proposée par exemple dans Journel et Huijbregts (1978) s'appuie sur des calculs de Matheron (1965) concernant la variance de fluctuation du variogramme autour de son modèle, et non la variance d'estimation du variogramme régional recherché par le variogramme expérimental des données). Par ailleurs, une décroissance du variogramme à l'origine, considérée comme non significative d'une structure, peut être due à une densité d'échantillonnage plus élevée pour les données plus riches et plus variantes : choisir un effet de pépite pur et stationnaire aura pour conséquence une surestimation du champ. A noter qu'un lien entre valeur des données et densité d'échantillonnage peut être voulu : le mineur préférant par exemple réduire l'incertitude en privilégiant l'échantillonnage des zones plus riches et plus variantes. Mais il peut aussi être fortuit : exemple des campagnes halieutiques multispécifiques: telle espèce de poisson pourra présenter des valeurs plus élevées là où la densité d'échantillonnage est la plus forte, avec un variogramme expérimental décroissant à l'origine et le risque de choisir l'hypothèse pépitique faute de mieux ; on acceptera *a contrario* avec une facilité trompeuse la structure variographique d'une espèce qui apparaît croissante sur la base du même échantillonnage...

Naturellement il est important de comprendre, autant que faire se peut, ce qui, dans les données, leurs valeurs, leurs implantations, conduit ou peut conduire à des distorsions du variogramme. Et bien sûr d'essayer d'y remédier.

Ainsi l'influence des valeurs extrêmes sur le variogramme, par exemple les valeurs fortes de concentrations, peut être considérable. J'ai été confronté à ce problème lors de ma toute première analyse géostatistique, ce qui m'a conduit à m'intéresser de manière purement empirique à l'émergence de la structure variographique (Rivoirard, 1987 a). Il s'agissait d'un gisement d'uranium de type stockwerk reconnu par sondages verticaux.

Il s'avérait que la structure verticale moyenne, obtenue par moyenne des variogrammes verticaux des différents sondages, était fortement influencée par les quelques sondages de loin les plus variants, et donc assez mal connue. Pourtant la situation était bien pire pour beaucoup de sondages pris individuellement. J'avais pu décrire les chutes brutales et les croissances hyperboliques du variogramme de certains sondages, en liaison avec la distance entre la position de la valeur extrême du sondage et le début ou la fin du sondage, à conjuguer avec le trou

provoqué par deux valeurs extrêmes pour leur inter-distance. Malgré une telle situation, l'émergence d'une structure verticale se faisait assez nettement sentir en moyennant les variogrammes de plusieurs sondages, révélant une structuration de l'ordre de la dizaine de mètres, avec une probable composante vers 4 mètres. Par contre les autres directions apparaissaient trop mal connues pour indiquer une anisotropie quelconque.

Cependant le variogramme classique d'ordre 2, basé sur des carrés d'accroissements, est peu robuste. Une panoplie d'outils alternatifs ont été proposés, et continuent de l'être, dans le but de mieux accéder à la structure variographique. Ainsi la covariance non-centrée (Journel et Huijbregts, 1978) ; le variogramme robuste estimé à partir du variogramme d'ordre ½ sous hypothèse gaussienne (Cressie and Hawkins, 1980) ; le variogramme d'ordre 1, appelé également madogramme ; les variogrammes robustes obtenus par quantiles des différences quadratiques (Armstrong and Delfiner, 1980), par moyenne des carrés d'incréments tronqués (Armstrong and Delfiner, 1980 ; Roustant, Dupuy, and Helbert, 2007), par quantiles des différences absolues (Genton 1998) ; un variogramme incluant une pondération des lois bivariables (Omre, 1984) ; des variogrammes relatifs à différents types de moyennes (Isaaks and Srivastava, 1989), les outils de covariance et corrélogramme dits non-ergodiques, dans lesquels les statistiques sont centrées et normées par des moyennes et écarts-types dépendant de la distance (Isaaks et Srivastava, 1988, Rossi et al. 1992) ; les indicatrices (Rossi et al. 1992) et autres transformations non-linéaires...

Le problème de valeurs extrêmes et de robustesse du variogramme est toujours d'actualité. Certaines transformations de la variable peuvent améliorer la robustesse de la structure estimée et la recherche d'anisotropie (tout en sachant que le variogramme de la transformée d'une variable ne correspond pas à celui de la variable originelle, sauf dans le cas extrêmement particulier du modèle mosaïque à valuations indépendantes). A moins qu'une valeur extrême originelle ne soit plus extrême après transformation (par exemple, après un écrêtage comme couramment pratiqué dans certains gisements d'or, ou bien après passage en indicatrice par un seuillage non extrême), un effectif conséquent reste nécessaire pour faire émerger une structure moyenne. Pour apprécier le problème soulevé par les valeurs extrêmes à l'état pur, on peut, comme Chilès et Delfiner (1999) le font, le schématiser par une variable indicatrice prenant peu fréquemment la valeur 1: cette variable restant binaire par toute transformation non triviale, le variogramme expérimental restera invariant (à l'échelle verticale près). Je reviendrai plus loin sur cette question des valeurs fortes, à propos de l'écrêtage, mais en me plaçant alors dans le contexte de la géostatistique non-linéaire (section 4.2).

Dans le cas de distributions de concentrations très dissymétriques, il est naturellement tentant de réduire l'influence des valeurs élevées en applicant une transformation de la variable. Le retour à la structure de la variable originelle (nécessaire pour calculer des variances de dispersion ou d'estimation) peut alors se faire si l'on dispose d'un modèle de la géostatistique non-linéaire. Ainsi on améliore souvent la structure en appliquant une transformation logarithmique, et on sait revenir à la structure de la variable originelle sous des hypothèses lognormales.

Confrontés à ce problème de distribution dissymétrique pour des données de concentrations en poissons, nous avons utilisé diverses approches pour améliorer la mise en évidence des structures (Guiblin et al., 1995). L'approche logarithmique apparaissait intéressante, mais les données comportaient un grand nombre de valeurs nulles. D'une part il était préférable de translater le variogramme avant d'en prendre le logarithme. D'autre part nous étions loin du cas lognormal. Nous avons alors établi et utilisé à titre heuristique une formule de retour similaire à celle du

lognormale, mais obtenue sans recours à la normalité (modèles de jetons valués multiplicatifs). Cette formule a permis de reconstituer le bon niveau de variabilité spatiale, tout en profitant de la stabilisation procurée par le passage au logarithme. Au prix d'une hypothèse de stationnarité sur la variable et pas seulement ses accroissements, la covariance non-centrée E(Z(x)Z(x+h)) nous est apparue comme un outil alternatif utile et non biaisé, en particulier exempt des biais qui s'introduisent dans la covariance centrée usuelle et dans le corrélogramme lors du centrage et de la normation (Journel et Huijbregts, 1978). Je reviendrai sur la robustesse de cet outil à propos de la géostatistique transitive.

Le manque de robustesse du variogramme expérimental peut être vu comme provenant de la différence entre la loi marginale empirique représentée par l'histogramme des valeurs de la variable, et la loi marginale correspondant à la loi bivariable empirique des valeurs à la distance h, utilisée pour le calcul du variogramme à cette distance. Omre (1984) propose un estimateur de variogramme robuste en affectant des pondérateurs de façon à réduire l'écart entre ces lois marginales. Cependant d'autres types de pondération peuvent être souhaitables, par exemple pour tenir compte d'échantillonnages irréguliers comme on le pratique couramment dans la modélisation d'histogrammes.

Dans Rivoirard (2000), j'ai proposé la notion générale de variogramme pondéré :

$$\gamma^w(h) = 0.5 \frac{\sum_{x_i - x_j = h} W_{ij} [Z(x_i) - Z(x_j)]^2}{\sum_{x_i - x_j = h} W_{ij}}$$

(où Z(x) représente la variable régionalisée, les x_i les points de données, et W_{ij} le poids affecté au couple $(Z(x_i), Z(x_j))$). Cet outil présente l'avantage de fournir des estimateurs non biaisés de la structure variographique, sans plus d'hypothèse que l'hypothèse intrinsèque classique présidant au variogramme. Par exemple, en calculant le variogramme à partir de chaque échantillon et en en faisant la moyenne, on obtient un « variogramme moyen par échantillon » s'appuyant par construction sur la loi marginale empirique. On peut bien entendu donner un poids différent w_i à chaque échantillon $Z(x_i)$ en cas d'échantillonnage irrégulier. Ceci peut être fait dans le variogramme moyen par échantillon, ou même directement dans l'estimateur classique du variogramme, qui devient alors un variogramme des échantillons pondérés :

$$0.5 \frac{\sum_{x_i - x_j \sim h} w_i w_j [Z(x_i) - Z(x_j)]^2}{\sum_{x_i - x_j \sim h} w_i w_j}$$

Celui-ci permet d'étendre les procédures courantes de déclusterisation de l'histogramme des données au variogramme, en se calant en particulier à la variance pondérée des échantillons. Ceci est utile en géostatistique non-linéaire et en simulations géostatistiques, lorsque l'on cherche à déterminer la structure de l'anamorphosée gaussienne supposée de moyenne nulle et de variance unité. A noter que la pondération des couples d'un variogramme pondéré fournit d'emblée un poids pour les points du variogramme, par exemple :

$$W(h) = \sum_{x_i - x_j \sim h} W_{ij} = \sum_{x_i - x_j \sim h} w_i w_j$$

ce qui est naturellement utile pour les ajustements automatiques (en particulier, un tel poids réduit l'influence des points du variogramme à très petites distances dans le cas de clusters, même si le niveau de variabilité correspondant n'est pas corrigé).

De façon générale, l'idée de pondérer les variogrammes (outre le lissage éventuel selon les distances) semble une bonne piste pour mieux estimer la structure en présence d'un échantillonnage irrégulier (Richmond, 2002 ; Emery et Ortiz, 2007, Reilly et Gelman 2007) ou de données dont la précision est hétérogène (Monestiez et al. 2006).

Références

Armstrong, M. and Delfiner, P., 1980, Towards a more robust variogram: A case study on coal, Technical report, Centre de Géostatisque, Fontainebleau.

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Cressie, N., and Hawkins, D.H., 1980, Robust estimation of the variogram: I, Math Geol, 12 (2), 115-125.

Emery X., and Ortiz J. M., 2007, Weighted sample variograms as a tool to better assess the spatial variability of soil properties, Geoderma 140, 81–89.

Genton, M., 1998, Highly Robust Variogram Estimation, Math Geol, 30 (2), 213-221.

Isaaks, E.H. and Srivastava, R.M., 1988. Spatial continuity for probabilistic deterministic geostatistics. Mathematical Geology, 20 : 313-341.

Isaaks, E.H., and Srivastava, R.M., 1989, An introduction to applied geostatistics. Oxford University Press, New York.

Journel, A. G., and Huijbregts, Ch. J. 1978. Mining geostatistics. Academic Press, London. x + 600 p.

Matheron, G. 1965. Les variables régionalisées et leur estimation. Masson, Paris. 306 p.

Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J-P., Guinet, C., 2006. Geostatistical modelling of spatial distribution of Balaenoptera physalus in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. Ecological Modelling 193:615-628.

Omre, H., 1984. The variogram and its estimation. In: Verly, G., David, M., Journel, A.G., Maréchal, A. (Eds.), Geostatistics for Natural Resources Characterization. Reidel, Dordrecht, 107–125.

Reilly, C., and Gelman, A., 2007, Weighted classical variogram estimation for data with clustering, Technometrics, Vol. 49, No. 2, 184-194.

Richmond, A., 2002. Two-point declustering for weighting data pairs in experimental variogram calculation. Computers & Geosciences 28 (2), 231–241.

Rossi, R. E., Mulla, D. J., Journel, A. G., & Franz, E. H., 1992, Geostatistical tools for modeling and interpreting ecological spatial dependence, Ecological Monographs 62(2), 277-314.

Roustant, O., Dupuy, D., and Helbert, C., 2007, Robust estimation of the variogram in computer experiments, Rapport interne, Département 3MI , Ecole des Mines de Saint-Etienne.

2.3 L'approche transitive pour le suivi des populations marines

Je résume, dans cette section, les développements décrits dans les papiers suivants :

Bez, N., Rivoirard, J., Guiblin, Ph., and Walsh, M., 1997. Covariogram and related tools for structural analysis of fish survey data. In Baafi and Schofield Eds, Geostatistics Wollongong'96, Kluwer Academic Publisher, Vol.2, 1316-1327.

Bez, N., and Rivoirard, J., 2000. On the role of sea surface temperature on the spatial distribution of early stages of mackerel using inertiograms, in ICES Journal of Marine Science, 57:383-392.

Bez, N., and Rivoirard, J., 2001. Transitive geostatistics to characterise spatial aggregations with diffuse limits: an application on mackerel ichtyoplankton. Fisheries research, 50, 41-58.

Woillez, M., Poulard, J-C., Rivoirard, J., Petitgas, P., and Bez, N., 2007. Indices for capturing spatial patterns and their evolution in time with an application on European hake (*Merluccius merluccius*) in the Bay of Biscay. ICES Journal of Marine Science, 64, 537-550. (Voir ANNEXE.)

Woillez, M., Rivoirard, J., and Petitgas, P., 2009, Notes on survey-based spatial indicators for monitoring fish populations. Aquat. Living Resour. 22, 155-164.

Moins puissante que la géostatistique intrinsèque classique, mais aussi moins gourmande en hypothèses, la géostatistique transitive, développée à la même époque par Matheron (1965), est peu utilisée. Rappelons qu'en géostatistique courante (dite intrinsèque), on modélise le comportement de la variable régionalisée à l'intérieur d'un champ, la variable étant implicitement supposée s'étendre avec le même comportement en dehors du champ ; dans l'approche transitive, au contraire, la variable s'éteint en s'éloignant du champ, et on ne cherche pas à faire la distinction entre champ et comportement sur le champ. L'outil structural de la géostatistique transitive, le covariogramme transitif, s'écrit simplement :

$$g(h) = \int z(x)\, z(x+h)\, dx$$

où z(x) est la variabile régionalisée et où l'intégrale est étendue théoriquement sur l'espace entier. L'approche transitive permet en pratique le calcul de la variance d'estimation d'une abondance

globale, à supposer que l'on dispose d'un échantillonnage non préférentiel et assez régulier, s'étendant au-delà des limites du champ.

L'approche transitive a été utilisée à deux dimensions pour l'estimation des œufs et larves de maquereaux à partir des campagnes Ifremer dans le Golfe de Gascogne (la variable régionalisée étant la densité surfacique en œufs et larves). Les premières applications remontent au début des années 80 par Francis Gohin (non publié ; j'en avais tiré un exercice pour les élèves de l'option « Géostatistique » de l'Ecole des Mines en 1983). Des applications plus récentes sur ces campagnes, ainsi que sur le poulpe marocain, figurent dans Bez (2002).

La géostatistique transitive a également été utilisée à une dimension, en prenant comme variable l'abondance cumulée par transect, lors de campagnes acoustiques basées sur des transects parallèles et équidistants : campagnes norvégiennes sur le hareng, en Mer de Norvège (thèse de Pierre Petitgas, 1991, encadrée par Matheron ; Petitgas, 1993), campagnes acoustiques au Sénégal (Samb and Petitgas, 1997). C'est une méthode extrêmement pratique pour estimer l'abondance avec sa variance d'estimation, basée sur le covariogramme transitif 1D. Nous l'avons également appliquée, dans le cadre du projet Européen GFSA (Geostatistics for Fish Stock Assessment, 1994-1998) aux campagnes norvégiennes sur le hareng séjournant dans les fjords (Rivoirard et Bez 1997, Rivoirard et al. 2000).

C'est à l'occasion du Projet Européen SEFOS (Shelf Edge Fisheries and Oceanographic Studies, 1993-1996), et avec Nicolas Bez dont j'encadrai la thèse à l'époque, que nous nous sommes intéressés de près à l'approche transitive, non pas tant pour obtenir des variances d'estimation globale d'abondance, que pour caractériser la structure spatiale des populations. En effet les distributions spatiales de concentrations de poisson sont en général dissymétriques, avec des valeurs fortes en petit nombre, et beaucoup de valeurs faibles ou nulles. Or, au contraire du variogramme de la géostatistique intrinsèque, le covariogram transitif possède de très bonnes propriétés de robustesse aux valeurs extrêmes, que ce soit les valeurs fortes des densités de poissons, ou bien les nombreuses valeurs faibles ou nulles qui peuvent poser problème pour la délimitation d'un champ. La formule du covariogramme (ou la formule empirique discrétisée correspondant à son estimation) permet d'ailleurs facilement d'apprécier la robustesse de l'outil. D'une part la contribution de z(x)z(x+h) devient nulle dès que z(x) ou z(x+h) est nul (ou supposé nul), d'où un raccord à l'assymptote g(∞) = 0. En outre, comme $|g(h)| \leq g(0)$ (théoriquement et empiriquement), le covariogramme présente un « bon » comportement à l'origine. Au total, le covariogramme, partant de g(0), avec un décrochement éventuel dû à l'effet de pépite, s'en va rejoindre l'assymptote g(h) = 0 aux très grandes distances (en outre, si z(x) est non-négatif, cas d'une densité de population, g(h) ne prend pas de valeur négative). D'une certaine façon, la covariance non-centrée dont il était question en section 2.2, qui s'écrit comme la moyenne (plutôt que la somme) des z(x) z(x+h), hérite en partie de ces bonnes propriétés de robustesse.

Toujours dans le cadre du projet Sefos, et avec Nicolas Bez, nous nous avons été amenés, dans la lignée de la géostatistique transitive, à utiliser et à développer des outils statistiques assez basiques permettant la description et le suivi des populations de poissons à partir de données de campagnes scientifiques, et présentant également une bonne robustesse aux zéros et aux valeurs fortes (thèse de Bez, 1997 ; Bez et al., 1997 ; Bez et Rivoirard, 2001). Ce travail a été poursuivi par Mathieu Woillez dont j'ai encadré la thèse, dans le cadre du projet Européen FISBOAT (Fisheries Independent Survey-Based Operational Assessment Tools, 2004-2007) (Thèse de Woillez, 2007 ; Woillez et al., 2007).

Un exemple simple de tels outils (pas original mais peu exploité à l'époque) est le centre de gravité d'une population et les statistiques associées (inertie, anisotropie mesurée sur les axes principaux) : la contribution d'un échantillon au calcul du centre de gravité est proportionnelle à la densité de poissons en ce point, et est en particulier nulle si celle-ci vaut zéro. Il est alors possible de suivre le déplacement d'une population en fonction des années ou de l'âge. Par exemple, avec l'âge, le merlu du Golfe de Gascogne se déplace vers le large (centre de gravité plus à l'Ouest) en se dispersant (inertie plus grande) (Woillez et al., 2007). Une question importante qui se pose est celle de savoir si de telles observations sont statistiquement significatives. Une répétition dans le temps, ou une tendance régulière dans le temps, sont des éléments en faveur d'une réalité représentée par ces observations. Il est possible également de comparer un mouvement de centre de gravité d'une population au mouvement éventuel du centre de gravité de l'échantillonnage (donc, non pondéré par la densité de poissons) pour voir si l'échantillonnage spatial n'est pas en cause (des aléas, tempêtes par exemple, peuvent en effet perturber le déroulement des campagnes…). La question d'un évènement significatif ou non reste posée cependant, si par exemple l'estimation du centre de gravité de la population vient soudainement de changer, sans que l'échantillonnage paraisse en cause.

Dans le même ordre d'idées, on peut s'intéresser aux conditions environnementales dans lesquelles une population se déploie à un moment donné: par exemple la distribution de la température par individu, autrement dit, l'abondance relative par classe de température. Une classe de température plus « habitée » ne représente d'ailleurs pas nécessairement un préférendum, en particulier si la population présente un comportement grégaire, car il faut bien que cette population soit quelque part. L'inertiogramme se propose de comparer la variance des valeurs de température de la population, aux variances que la population présenterait si on la translatait, par exemple (thèse de Bez, 1997 ; Bez et Rivoirard, 2000). L'observation d'une variance réelle minimale plaide alors en faveur d'un préférendum.

La recherche d'indices permettant de décrire l'état d'une population à partir de campagnes scientifiques est d'autant plus importante que la communauté halieute a quelques doutes sur la possibilité de déterminer le niveau absolu d'une abondance sur lequel prendre des décisions de management. Il peut s'agir d'indice vitaux (relatifs à la distribution en longueur, age, sexe, maturité, recruitement, mortalité) comme d'indices spatiaux (projet Européen Fisboat 2004-08; thèse de Woillez, 2007 ; Woillez et al., 2007 ; Woillez et al., 2009). Parmi ces derniers, on peut citer centre de gravité, inertie, anisotropie, nombre de patches, effet de pépite relatif du covariogramme, lesquels indices changent si l'on permute deux valeurs d'échantillons dans l'espace. D'autres indices spatiaux résument la distribution statistique des valeurs d'échantillons, et sont invariants par permutation (en maille régulière) : il s'agit par exemple de l'aire positive (aire couverte par les densités positives), de l'aire équivalente EA (Equivalent Area, aire qui serait occupée par la population si tous les individus avaient la même densité) ou de l'aire d'étalement SA (Spreading Area). Ces deux dernières sont liées respectivement au coefficient de variation $CV = \sigma / m$ et à l'indice de Gini G qui seraient calculés sur un champ A par les formules suivantes, mais ont l'avantage de recevoir une contribution nulle de la part des données nulles :

$$EA = \frac{Q^2}{\int z(x)^2 dx} = \frac{A^2 m^2}{A(m^2 + \sigma^2)} = \frac{A}{(1 + CV^2)}$$

$$\frac{SA}{A} + G = 1$$

où $Q = \int z(x)dx$ représente l'abondance de la population (Woillez et al. 2007). Pour revenir à l'exemple déjà cité du merlu dans le Golfe de Gascogne, bien que les stades les plus âgés voient leur dispersion augmenter (au sens de l'inertie), l'aire positive qu'ils occupent, l'aire équivalente ou l'aire d'étalement ont tendance à se réduire : ils occupent finalement une aire décroissante dans un territoire croissant (Woillez et al. 2007).

Références

Bez N. 1997, Statistiques individuelles et géostatistique transitive en écologie halieutique, Thèse de Docteur en Géostatistique, Ecole des Mines de Paris, 280 p.

Bez, N., 2002. Global fish abundance estimation from regular sampling: the geostatistical transitive method. Canadian Journal of Fisheries and Aquatic Sciences, 59 : 1921-1931.

Matheron, G., 1965. Les variables régionalisées et leur estimation. Masson, Paris. 306 p.

Petitgas, P., 1991. Contributions géostatistiques à la biologie des pêches maritimes. Thèse de Doctorat en Géostatistique, Ecole des Mines de Paris, 211 p.

Petitgas, P., 1993. Geostatistics for fish stock assessments: a review and an acoustic application. ICES Journal of Marine Science, 50: 285-298.

Rivoirard, J., and Bez, N., 1997, A 1D geostatistical analysis on Norwegian spring-spawning herring acoustic data in Ofotfjord (Dec. 1994). ICES CM 1997/Y:12, 9 p.

Rivoirard, J., Simmonds, J., Foote, K.G., Fernandes, P., and Bez, N., 2000. Geostatistics for estimating fish abundance. Blackwell Science, Oxford. 206 p.

Samb, B. and Petitgas, P. 1997. Estimation de la précision des campagnes acoustiques au Sénégal par la méthode géostatistique transitive à une dimension. Aquatic Living Resources 10: 75-82.

Woillez, M. 2007. Contributions géostatistiques à la biologie halieutique, Thèse de Doctorat en Géostatistique, Ecole des Mines de Paris, 173 p.

2.4 L'analyse structurale multivariable

Revenons à présent au cadre classique de la géostatistique, dans lequel on cherche à décrire le comportement d'une ou plusieurs variables régionalisées dans un champ. Les situations multivariables peuvent être extrêmement variées, mais dans une partie des cas, l'approche classique de la géostatistique linéaire, basée sur variogrammes croisés, cokrigeage, et par exemple modèle linéaire de corégionalisation, répondra au problème posé. Je voudrais plutôt ici citer certaines analyses multivariables auxquelles j'ai contribué, et d'où un modèle ou une propriété originale ont émergé.

2.4.1 Un modèle global pour le hareng écossais

Cette section résume:

Guiblin, P., Rivoirard, J., and Simmonds, E. J., 1996, Spatial distribution of length and age for Orkney-Shetland herring. ICES CM 1996/D:14, 20 p.

ainsi que la section 4.4 de :

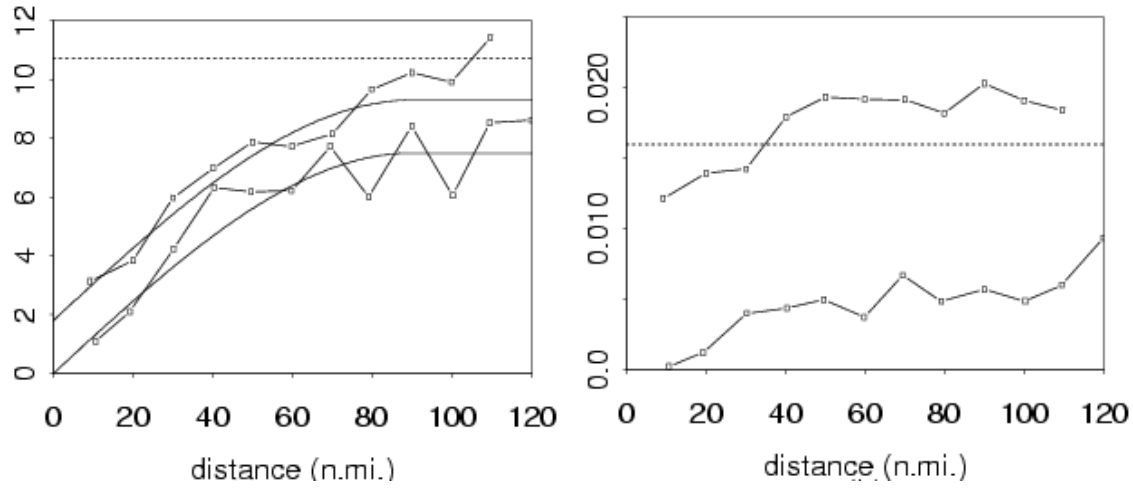
Rivoirard, J., Simmonds, J., Foote, K.G., Fernandes, P., and Bez, N., 2000. Geostatistics for estimating fish abundance. Blackwell Science, Oxford. 206 p.

Ce travail a été réalisé dans le cadre du projet GFSA (Geostatistics for Fish Stock Assessment, 1994-1998), avec Philippe Guiblin dont j'encadrais la thèse, et en collaboration avec le Marine Lab d'Aberdeen. Il s'agit des campagnes acoustiques dont il a déjà été question à propos du support d'une variable régionalisée (Simmonds et al., 1997). Le navire effectue des transects d'est en ouest et d'ouest en est de façon à couvrir la zone à échantillonner. L'écho acoustique indique la présence du hareng. De temps en temps, le navire fait un coup de chalut, de façon à confirmer l'espèce, à déterminer la distribution en classes de longueur en cette station et à déterminer l'age de certains individus par examen de leur otolithe. Ces individus serviront à établir la clé taille-age qui sera utilisée pour convertir les tailles en ages. Le nombre de poissons chalutés n'est, dans ces campagnes, pas un indice d'abondance. En un point 2D de l'espace, la densité de poissons est donnée par l'écho acoustique, corrigé par la longueur moyenne du poisson au sens des carrés. Traditionnellement, la distribution des tailles et l'application de la clé taille-age permet de désagréger les nombres d'individus en classes d'age.

L'analyse des données de plusieurs années de campagnes a permis de construire le modèle multivariable suivant (Guiblin et al., 1996; thèse de Guiblin, 1997 ; Rivoirard et al., 2000). Commençons par la longueur. La distribution des tailles en une station de chalutage est très concentrée sur sa moyenne. Autrement dit la taille moyenne (moyenne habituelle ou moyenne au sens des carrés, du coup très voisines) est très caractéristique du poisson en ce point de l'espace. Du point de vue spatial, le poisson est plus grand en allant vers le large, et la taille moyenne augmente avec la profondeur du fond. De plus le variogramme interannuel de cette longueur (ou de son résidu par rapport à la profondeur) est, contrairement à celui des âges, à peu près identique au variogramme moyen annuel, avec un petit décalage vertical apparaissant comme une composante pépitique (Fig. 5). Ce décalage peut être interprété comme une variation interannuelle des longueurs, la distribution dans l'espace de celles-ci étant sinon peu variable d'une année sur l'autre. Ainsi l'analyse variographique met en évidence, du point de vue écologique, le rôle stable que joue la taille du poisson dans l'occupation de l'espace. La cartographie de la longueur peut se faire par krigeage 2D (avec les données de l'année) ou 3D (avec les données de plusieurs années, en rajoutant la variabilité interannuelle), la profondeur jouant le rôle de dérive externe (Fig. 6).

Pour une année donnée, la relation entre longueur moyenne en un point et proportion au-dessus de chaque age est très serrée, et permet de passer, moyennant un petit résidu, d'une carte des longueurs à des cartes de proportions d'ages. Cependant les relations sont variables d'une année sur l'autre, de sorte que la distribution spatiale des ages présente une portée écologique moins

forte que celle de la taille. Pour une année donnée, les cartes de proportions d'age permettent finalement de désagréger par classes d'age la cartographie de la densité d'individus issue de l'acoustique. Il suffit ensuite de sommer les densités totale et par age pour en déduire les abondances totale et par age. Au total, on a donc un modèle multivariable du hareng incluant à la fois le modèle de la donnée acoustique pour une année, celui de la longueur (conjoint avec les autres années et dépendant de la sonde) et enfin celui des ages de l'année (dépendant de celui de la longueur de l'année). Une transcription de ce modèle global en terme de simulations géostatistiques a été réalisé de façon à combiner les sources d'incertitude spatiale sur les différentes variables, aboutissant à un CV de l'abondance compris en général entre 10% et 20% selon les âges (Woillez et al. 2009, voir section 5.1 du présent mémoire).



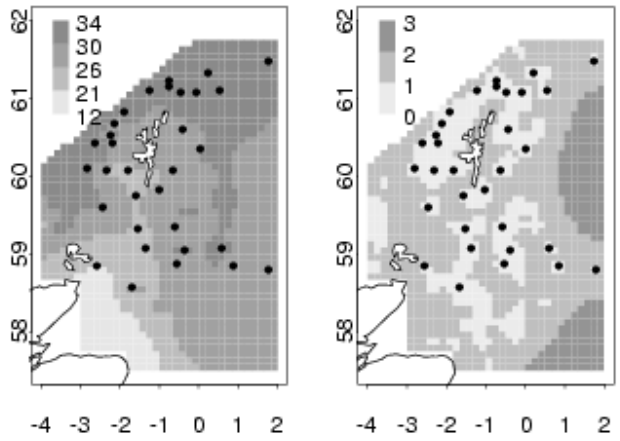
Figures 5.
Gauche : variogramme de la longueur moyenne en stations du hareng Ecossais (déduction faite de la part liée à la profondeur) ; le variogramme calculé à partir des paires interannuelles se déduit du variogramme moyen par année par l'ajout d'une petite composante pépitique, témoignant d'une faible variation interannuelle.
Droite : variogramme de la proportion d'age 4 et plus en stations : la variation interannuelle est prépondérante. (Extrait de Rivoirard et al., 2000.)

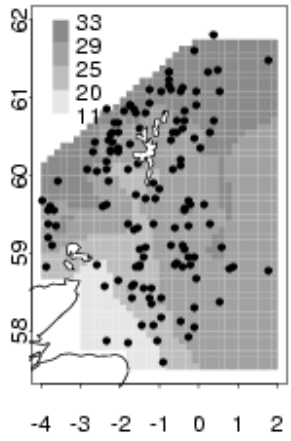
Références

Guiblin, Ph., 1997, Analyse statistique de campagnes (acoustique et chalutage) sur le hareng écossais, Thèse de Doctorat en Géostatistique, Ecole des Mines de Paris, 161 p.

Simmonds, E. J., Bailey, M., Toresen, R., Couperus, B., Pedersen, J., Reid, D. G., Fernandes, P. G. & Hammer, C. 1997. 1996 ICES coordinated acoustic survey of ICES Divisions IIIa, IVa, IVb, and VIa. ICES C.M. 1997 /H:11, 13 p.

Woillez, M., Rivoirard, J., and Fernandes, P.G. 2009. Evaluating the uncertainty of abundance estimates from acoustic surveys using geostatistical simulations – ICES Journal of Marine Science, 66:1377–1383.





Figures 6: (a) krigeage de la longueur du hareng pour l'année 1993 ; (b) l'écart-type de krigeage correspondant ; (c) carte générique pour une année quelconque. Les points représentent les stations de chalutages, pour 1993 (en haut), et pour les 6 années 1989-1994 (en bas). Les krigeages utilisent les données des six années, tenant compte de la petite variabilité interannuelle. (Extrait de Rivoirard et al., 2000.)

2.4.2 Des modèles adaptés au filtrage de données sismiques

Il m'a paru intéressant de rendre compte ici de certains développements réalisés avec Gaz-de-France dans le cadre d'une action FSH (Fonds de Soutien aux Hydrocarbures) :

Rivoirard, J., Renard, D., et Léger, M., 1993, Quantification de l'incertitude sur les profondeurs estimées par sismique et par puits. Rapport FSH, Centre de Géostatistique, ENSMP, 16 p. + planches. Confidentiel.

Je remercie ici Gaz-de-France d'avoir accepté que les résultats suivants puissent être utilisés à des fins pédagogiques et figurer dans le présent mémoire.

La cartographie d'un horizon géologique est problématique, lorsqu'on dispose de données variant régulièrement le long de profils, mais que les profils présentent des écarts de valeurs importants là où ils se croisent. Il s'agit en l'occurrence d'un problème rencontré sur des données de temps sismiques. Le variogramme moyen apparaît continu à l'origine, à cause de l'écrasante majorité de couples de valeurs provenant d'un même profil dans le calcul du variogramme à courtes distances. Faute d'une composante pépitique, il s'ensuit une cartographie par krigeage tourmentée au voisinage des croisements entre profils. Or il est apparu, d'après les analyses que nous avons faites, que le variogramme inter-profil est en grande partie identique au variogramme moyen par profil, à un décalage vertical près apparaissant comme une composante pépitique bien qu'il diminue légèrement avec la distance (on notera la similitude variographique entre ces données sismiques et les longueurs du hareng écossais…). On peut alors schématiquement interpréter la donnée sur un profil comme une variable propre, avec donc autant de variables que de profils, chaque profil étant la somme d'une variable cachée qui serait la vraie valeur, et d'une erreur propre au profil (la variable cachée et les différentes erreurs étant indépendantes). Un cokrigeage dans ce modèle permet alors de prendre en compte les écarts entre données aux croisements, en considérant qu'ils proviennent des erreurs propres à chacun des profils, et fournit une carte bien plus cohérente que le krigeage précédent à ces endroits (Fig. 7 - 8).

Naturellement ceci n'explique en rien l'origine des écarts (peut-être due à des problèmes de migration chers aux géophysiciens, et qu'il y aurait intérêt à réduire en amont) mais montre la puissance et la souplesse de l'outil géostatistique pour rendre cohérentes des mesures qui ne le sont pas tout à fait. Il est à ce sujet tout à fait remarquable que de tels traitements géostatistiques ne nécessitent aucunement de maillage régulier. Dans le même ordre d'idées, l'extraction par analyse krigeante de composantes spatiales issue d'une modèle linéaire de corégionalisation peut fournir un filtrage efficace de données sismiques 4D. Ainsi, en croisant la sismique de deux campagnes, on peut extraire la structure spatiale de la partie commune, et estimer par cokrigeage factoriel partie commune et différences par rapport à celle-ci (Beucher et al. 2001 ; Coléou et al., 2002).

Références

Beucher H., Galli A., Rivoirard J., 2001, Pravap 10 : Geostatistical analysis of the 4D seismic. Rapport technique, Centre de Géostatistique, Ecole des Mines de Paris, 56 p. Confidentiel.

Coléou, T., Hoeber, H., Lecerf, D., 2002, Multivariate geostatistical filtering of time-lapse seismic data for an improved 4D signature. 73rd Ann. Intern. Mtg., S.E.G., Expanded Abstracts.

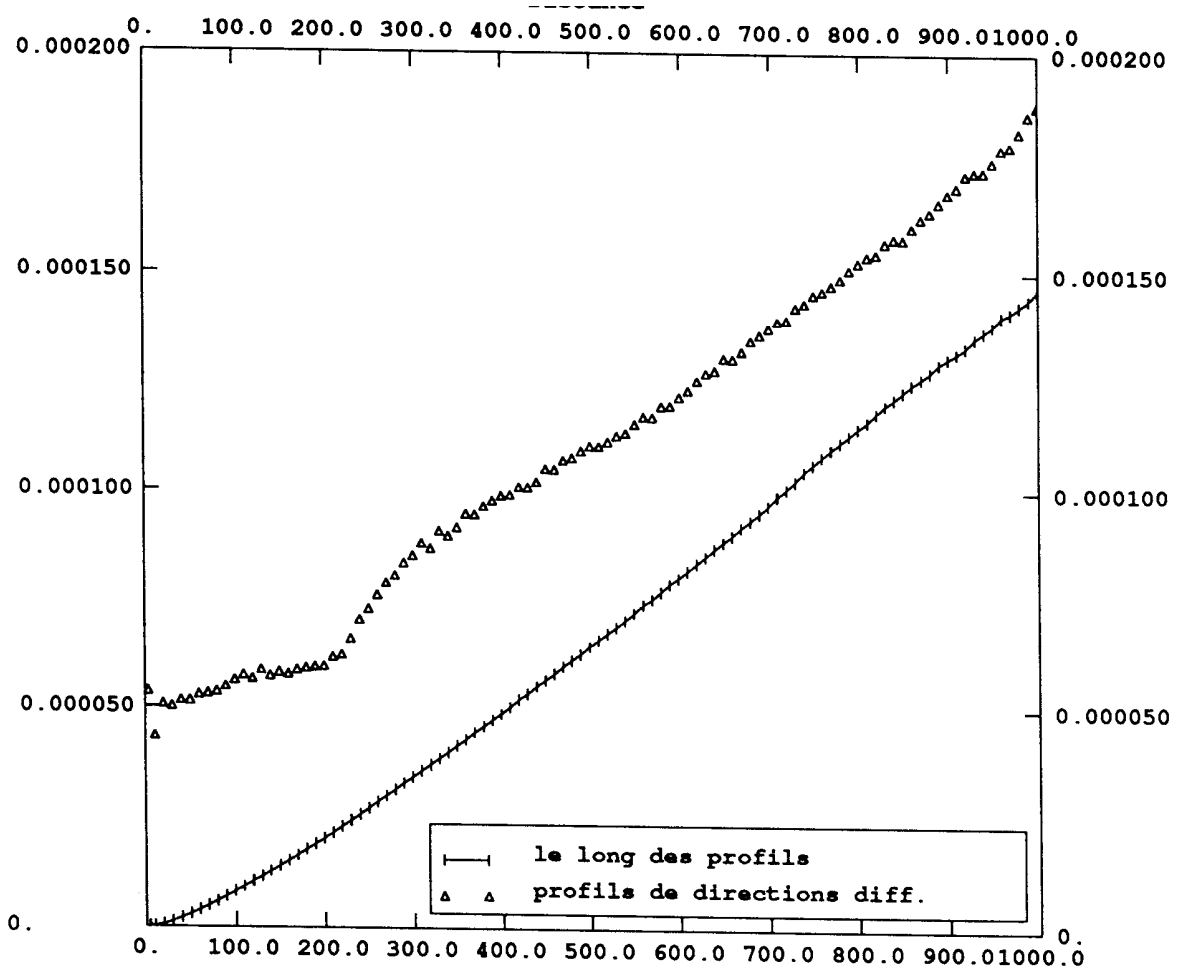
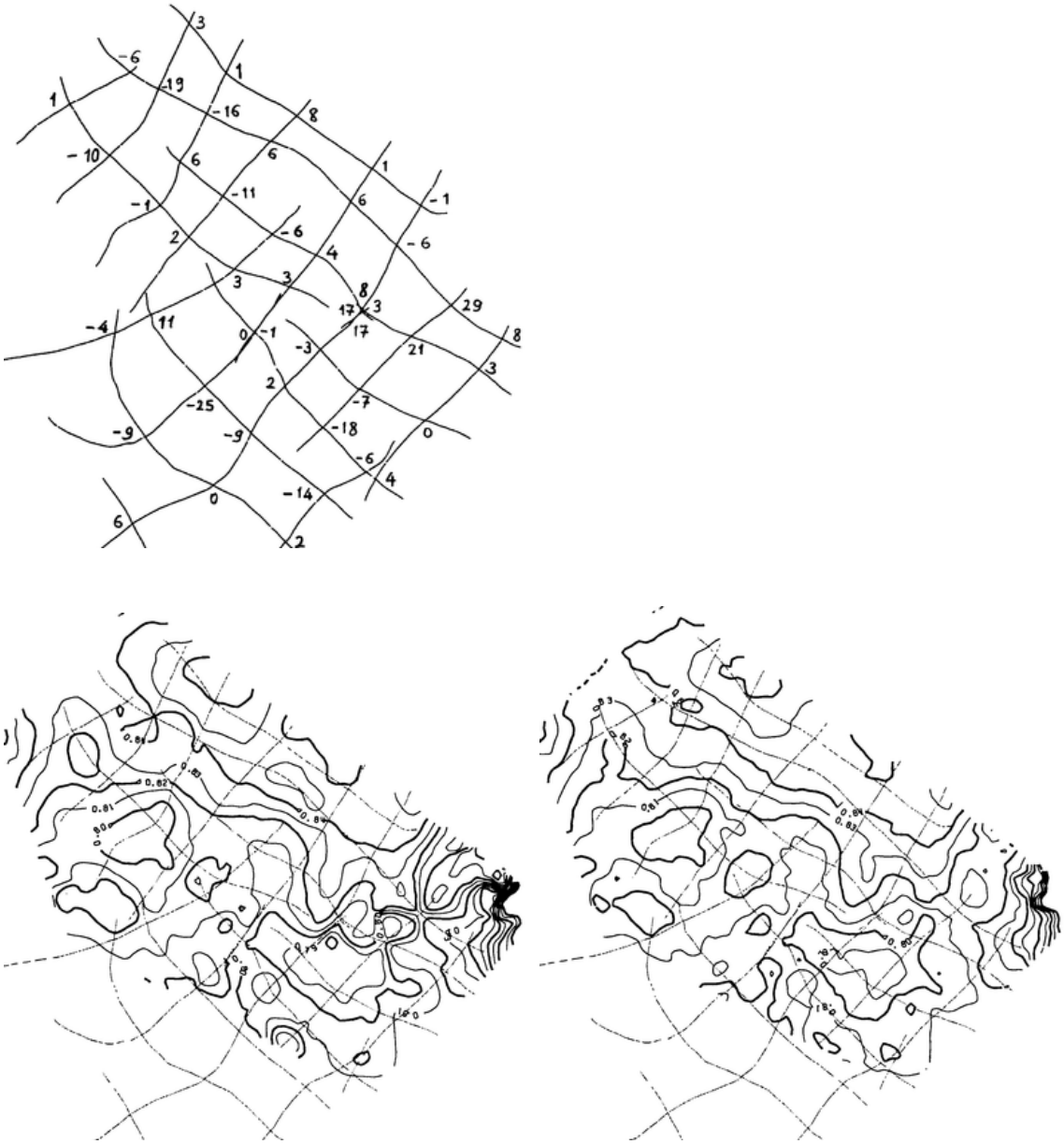


Figure 7 : Temps sismiques le long de profils : variogramme moyen le long des profils (pratiquement égal au variogramme moyen), et variogramme entre profils de directions différentes : en gros, une composante pépitique permet de passer de l'un à l'autre. (Extrait de Rivoirard et al., 1993.)



Figures 8. Dans l’ordre : Implantation des profils, et écarts aux croisements. Cartographie par krigeage du temps sismique avec la structure continue du variogramme moyen. Cokrigeage avec composante pépitique entre profils : les écarts aux croisements sont correctement pris en compte. (Extrait de Rivoirard et al., 1993.)

2.4.3 Le modèle à résidu, exemple d'une hiérarchie entre minerai et métal

Dans cette section, je résume certains travaux, effectués dans le cadre d'un important projet de recherche financé par Cogéma, intégrant la thèse de Jean-Luc Bordessoule que j'ai encadrée, et ayant conduit à la construction du modèle à résidu :

Bordessoule, J.-L., Demange, C., et Rivoirard, J., 1988, Estimation de réserves géologiques en uranium par variables utiles à résidu autokrigeable. *Sciences de la Terre*, série Informatique géologique, *28*, pp 27-51.

Bordessoule, J.-L., Demange, C., and Rivoirard, J., 1989, Using an orthogonal residual between ore and metal to estimate in-situ uranium resources. In M. ARMSTRONG (ed.), *Geostatistics*, Vol. 2. Dordrecht : Kluwer, 1989, 923-934. (Voir ANNEXE.)

Rivoirard, J., et Bordessoule, J.-L., 1987, Etude géostatistique sur le gisement de Tréviels. 83 p. Confidentiel.

Le modèle linéaire de corégionalisation est un outil puissant. En outre un ajustement automatique des paliers des variogrammes simples et croisés facilite grandement la modélisation, qu'on peut alors utiliser en cokrigeage ou simulation. Mais il est parfois intéressant d'observer les choses d'un peu plus près.

Le gisement d'uranium de Tréviels (Hérault), qui fut exploité à ciel ouvert par la Cogéma, est constitué de couches sédimentaires à léger pendage vers le sud, décalées par un jeu de failles subverticales est-ouest plus ou moins importantes. Certaines de ces couches, d'épaisseur métrique, sont porteuses d'uranium, la ressource minéralisée constituant 4% du volume. L'exploration est menée par sondages subverticaux à maille 25 m. L'exploitation est faite par gradins de 5 mètres, les trous de tir répartis à maille 3 mètres, et le contrôle des teneurs opéré sur la radiométrie (convertie en uranium) mesurée tous les 20 cm le long de ces trous de tir de 5 mètres.

Une étude géostatistique approfondie a été réalisée sur ces données de trous de tir à maille fine en vue de construire un modèle apte à représenter de tels gisements (Rivoirard et Bordessoule, 1987 ; Bordessoule et al., 1988, 1989 ; thèse de Bordessoule, 1990). On a travaillé sur les « variables utiles » additives minerai (noté T) et métal (Q), plus exactement proportion de minerai et métal correspondant, dont le rapport est égale à la teneur du minerai (variable non additive puisque son moyennage nécessite justement sa pondération par le minerai). Il s'agit en somme de l'équivalent 3D des variables utiles puissance et accumulation-métal d'une couche minéralisée intégrée selon une direction donnée pour être traitée à 2D.

Au niveau des trous de tir, le métal augmente fortement avec le minerai, avec un coefficient de corrélation de 0.76. La teneur augmente plus doucement avec le minerai, présentant une corrélation de 0.44. La variographie des variables utiles révèle une forte anisotropie des variables, avec continuité maximale dans la direction de plongement des couches. Le minerai, qui représente la géométrie de la minéralisation, manifeste une continuité beaucoup plus marquée que le métal. Enfin, le plus remarquable est l'extraordinaire similitude entre le variogramme croisé et le variogramme du minerai. Ainsi, lorsque l'on croise le métal Q avec le minerai T, la partie

propre au métal disparaît, et ce qui reste est très voisin du minerai (à un facteur d'échelle près, naturellement) : Fig. 9. On peut donc écrire $Q(x) = E(Q(x)|T(x)) + R(x)$, avec R(x) résidu de la régression de Q(x) sur T(x), et donc non corrélé à T(x), le variogramme de la régression étant proche de celui de T(x). Remarquons qu'on peut toujours écrire une telle relation entre deux variables T et Q, mais cela ne présente un intérêt en géostatistique que si R(x) est *spatialement* non corrélé avec la régression (donc entre deux points différents et pas seulement en un même point x), ce qui est précisément le cas ici. Le modèle multivariable ci-dessus est un peu compliqué à utiliser pour un estimation des ressources, à cause de la présence simultanée de T(x) et de $E(Q(x)|T(x))$ qui est une fonction de T(x). Pour un cokrigeage, il y a donc tout intérêt à le simplifier en utilisant la régression linéaire de Q(x) sur T(x), soit

$Q(x) = aT(x) + b + R(x)$, ou même

$Q(x) = mT(x) + R(x)$,

en forçant le modèle à passer par l'origine ($T(x) = 0 \Rightarrow Q(x) = 0$; m représente ici la teneur moyenne du minerai, et la teneur Q/T n'est plus corrélée au minerai).

Le *modèle à résidu,* que nous proposons ici, est bien entendu très particulier, mais me semble extrêmement intéressant. Dans ce modèle, il existe une *hiérarchie* entre les variables, avec une *variable maîtresse* (T, la géométrie), et une *variable subordonnée* à cette variable maîtresse (Q), qui s'obtient par addition d'un résidu sans corrélation spatiale avec la variable maîtresse ($R = Q - mT$). Le variogramme croisé entre T et Q est égal à m fois le variogramme de T, et le variogramme de Q est la somme de m^2 fois le variogramme de T et du variogramme du résidu. Ici le résidu apparaît peu structuré, car c'est lui qui permet de passer du minerai au métal, en prenant en charge les teneurs Q/T dans toute leur variabilité.

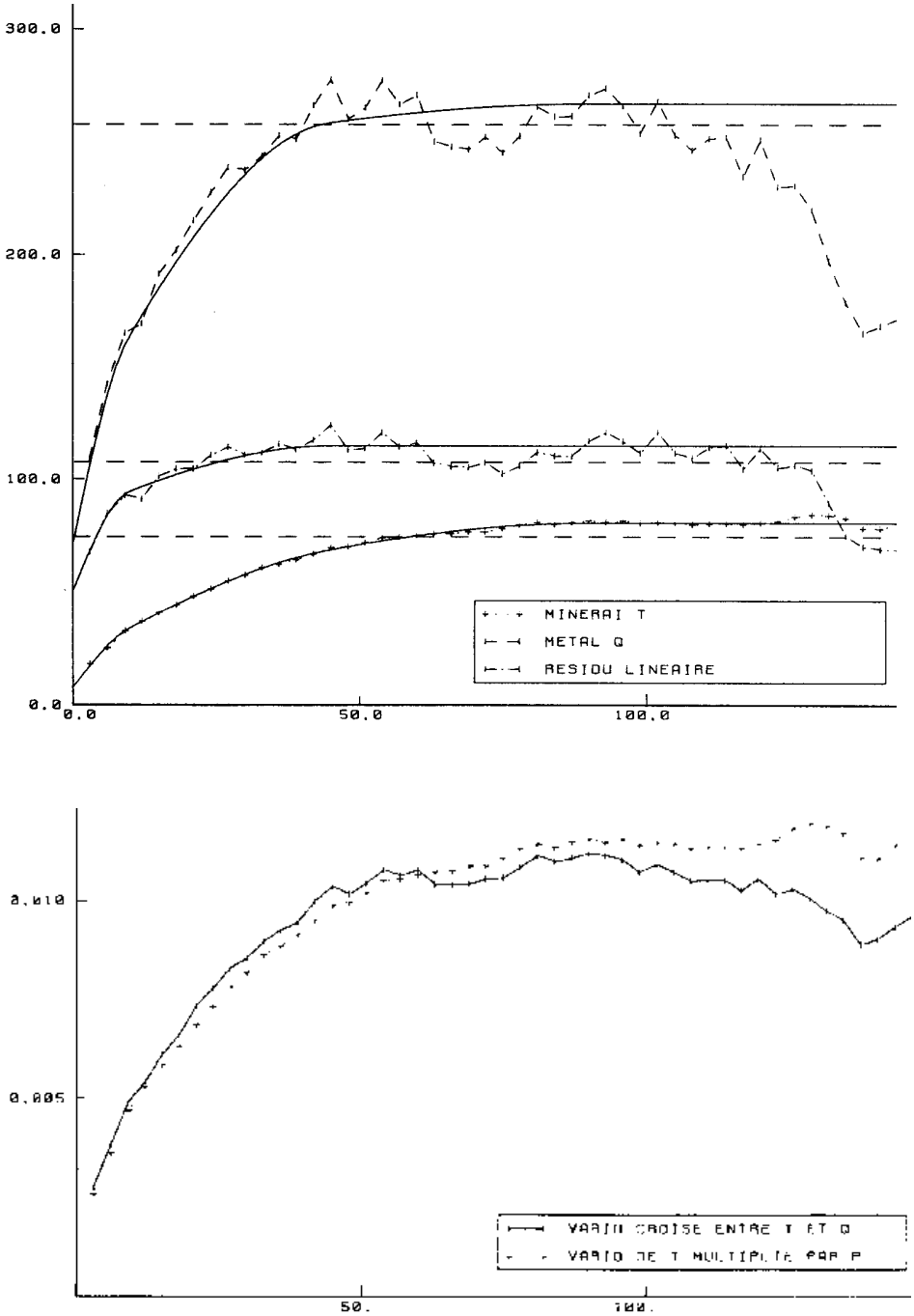
Il semblerait qu'un tel modèle à résidu entre deux variables, détectable par un variogramme croisé identique au variogramme de l'une des variables, s'observe en fait assez souvent, avec un résidu pépitique ou non : exemple du poids de pierres subordonné à leur nombre, observé dans des gisements de diamants (comm. pers. C. Lantuéjoul); du nombre d'arbres subordonné à une information photographique dans le domaine forestier (de Fouquet et Mandallaz, 1992) ; cas des données de chalut subordonnées à l'acoustique en halieutique (thèse de Mireille Bouleau, 2005, encadrée par Nicolas Bez) ; pollution estivale en NO2 subordonnée à la pollution hivernale (de Fouquet, 2006) ; etc. Ceci est d'autant plus appréciable que ce modèle jouit de remarquables propriétés en estimation, comme on va le voir au chapitre suivant. On remarquera sinon avec intérêt que la hiérarchie provient de la similitude entre structures (la croisée et une simple), mais non de la continuité plus ou moins forte des différentes structures : rien n'empêcherait que la variable maîtresse soit la variable la moins structurée.

Références

Bouleau, M., 2005, Combinaison géostatistique de l'acoustique et des captures dans les campagnes scientifiques de pêche par chalutage. Thèse de doctorat en géostatistique, Ecole des Mines de Paris, 189 p.

de Fouquet, C., 2006, La modélisation géostatistique des milieux anthropisés. Habilitation à Diriger des Recherches, Sciences de la Terre, Paris, Université Pierre et Marie Curie.

de Fouquet, C., and Mandallaz, D., 1992, Using geostatistics for forest inventory with air cover: an example. In: A. Soares, Editor, Geostatistics Troia'92, Vol. 2, Kluwer, Dordrecht, The Netherlands (1992), 875–886.



Figures 9. Variographie de variables utiles minerai et métal. En haut, variogrammes simples du minerai, du métal, et du résidu : on notera le contraste entre la structure du minerai (assez continue), et celle du métal (beaucoup plus variable), due au résidu. En bas : le variogramme croisé entre minerai et métal est pratiquement proportionnel au variogramme simple du minerai. (Extrait de Bordessoule et al., 1988.)

3 Krigeage et cokrigeage

3.1 Repousser les frontières du voisinage de krigeage ?

Je résume dans cette section des travaux sur le choix du voisinage de krigeage:

Rivoirard, J., 1984 a, Looking for a kriging plan in a stockwerk deposit. In : G. Verly et al. (eds.), *Geostatistics for natural resources characterization*, Dordrecht : Reidel, Part 2, p 951-963.

Rivoirard, J., 1984 b. *Le comportement des poids de krigeage*. Thèse de Docteur-Ingénieur en Sciences et techniques Minières, Option Géostatistique, ENSMP, 72 p.

Rivoirard, J., 1987 b, Two key parameters when choosing the kriging neighbourhood. *Mathematical geology, 19* (8), 851-856.

Le krigeage est l'interpolateur linéaire classique de la géostatistique (Matheron, 1965 ; Journel et Huijbregts, 1978 ; Chilès et Delfiner, 1999). La valeur estimée en un point (ou sur un bloc, un domaine…) est une moyenne pondérée des valeurs du voisinage, et éventuellement d'un paramètre de moyenne. Ainsi le « krigeage simple » (KS, ou krigeage à moyenne connue) fait-il usage d'un paramètre de moyenne, supposé alors stationnaire. Le krigeage ordinaire (KO ou encore krigeage à moyenne inconnue), au contraire, n'en fait pas usage. Le voisinage est dit « unique » lorsque l'ensemble des données est utilisé, ce qui est possible lorsque leur nombre n'est pas trop grand. Il est « glissant », incluant seulement les données les plus voisines, lorsqu'il se déplace avec la cible, point ou bloc. Les données les plus proches, et donc les mieux corrélées avec la cible, font souvent écran aux plus lointaines, facilitant le choix du voisinage.

Mes premiers contacts avec le krigeage ont été aussi problématiques et stimulants que ma rencontre avec le variogramme (Rivoirard, 1984 a). Il s'agissait du même gisement d'uranium en stockwerk dont j'ai parlé à propos de l'influence des valeurs extrêmes sur le variogramme. En possession d'un modèle de variogramme présentant un beau palier, je m'apprêtais à effectuer un krigeage ordinaire de blocs quand j'observais la chose suivante : certes les poids de krigeage les plus forts correspondaient aux données intérieures au bloc à estimer, mais les échantillons périphériques recevaient eux aussi des poids notables. Et plus j'agrandissais le voisinage, plus la périphérie s'éloignait du bloc à estimer, emportant avec elle une bonne partie des poids. Bref, j'étais loin du cas d'école dans lequel le voisinage s'arrête là où les poids deviennent négligeables, les données les plus proches faisant écran aux plus lointaines. Ainsi que j'ai pu le montrer, le phénomène que j'observais était finalement dû au contraire à un effet d'écran inverse, qui s'explique par la décomposition du krigeage ordinaire (KO ou encore krigeage à moyenne inconnue) à l'aide du krigeage simple (KS, ou krigeage à moyenne connue) et du krigeage de la moyenne. En effet le KO n'est autre qu'un KS, dans lequel la moyenne est remplacée par son estimation par krigeage (Matheron, 1970) :

$$Z^{KS} = \sum \lambda_\alpha Z(x_\alpha) + \left(1 - \sum \lambda_\alpha\right) m$$

$$Z^{KO} = \sum \lambda_\alpha Z(x_\alpha) + \left(1 - \sum \lambda_\alpha\right) m^K$$

Or l'estimation par krigeage de la moyenne, autrement dit de la moyenne sur l'espace entier, privilégie les données périphériques, qui font écran aux données plus intérieures : d'où l'écran inverse. Naturellement ceci ne se fait sentir en KO que si le poids attribué à la moyenne en KS n'est pas petit. D'où l'importance, dans le choix du voisinage du KO, de ce paramètre qu'est le poids de la moyenne en KS $\lambda_m = 1 - \sum \lambda_\alpha$ (Rivoirard, 1984 b, 1987 b).

Le krigeage peut être vu de différentes façons, notamment comme un interpolateur linéaire optimal, sans biais et de variance d'estimation minimale. En estimation minière cependant, il représente plutôt une approximation de la valeur probable d'un bloc connaissant les échantillons du voisinage (Espérance Conditionnelle, laquelle nécessite des hypothèses beaucoup plus fortes). Son but est de réduire la variabilité trop forte des échantillons (effet de support), et de s'approcher du non-biais conditionnel E(Z|Z*) = Z*, qui garantit qu'en sélectionnant les blocs (estimés) les plus riches, on récupère bien en moyenne la teneur attendue (il s'agit là de l'origine du krigeage dans les années 50, avec les travaux de Krige (1951) et Matheron (1955), alors dans un contexte lognormal). Une sélection basée sur un estimateur trop variant (comme la valeur des échantillons) conduit en effet à une surestimation du riche, le résultat étant systématiquement inférieur à la prévision. A noter que les écarts provenant ainsi du seul effet de support peuvent être attribués à tort à un problème de biais d'échantillonnage. On peut s'apercevoir de la cause réelle des écarts lorsqu'on constate que le minerai rejeté au stérile est, lui, moins pauvre que prévu. A la limite (exemple d'un gisement d'or purement pépitique) toute sélection est illusoire, et conduirait à un stérile en réalité aussi riche que le minerai sélectionné.

D'où l'intérêt du krigeage, pour éviter la surestimation des valeurs riches. Encore faut-il qu'il soit effectivement proche du non-biais conditionnel, dont la formulation nécessite des hypothèses mathématiques autres que celles du krigeage : la courbe de régression E(Z|Z*) doit être linéaire et égale à la première bissectrice, or il n'y a pas de raison pour que cette courbe, qu'on ne connaît pas, soit effectivement linéaire. Pour s'approcher du non-biais conditionnel, on essaiera alors de s'approcher du cas d'une droite de régression pZ* + b égale à la première bissectrice, autrement dit d'une pente de régression linéaire égale à 1, ou tout au moins proche de l'unité (Rivoirard, 1987 b). Cette pente *p*, égale à cov(Z,Z*)/var(Z*), est effectivement égale à 1 pour le KS ou krigeage à moyenne connue.

Or le KS conduit, dans la logique de la stationnarité, à des estimations se rapprochant de la moyenne lorsque l'information se raréfie : le manque d'information est compensé par la moyenne connue. Dans un gisement minier par exemple, ceci conduirait tout naturellement à donner à des zones moins reconnues (parce que plus difficiles d'accès, plus profondes, ou simplement jugées a priori moins intéressantes) des estimations fortement influencées par la valeur moyenne. La stationnarité à l'échelle du champ est une hypothèse potentiellement forte, pas forcément contrôlable (les valeurs des quelques échantillons profonds sont-ils significativement différents des autres ?). Aussi un souci fréquent en géostatistique, est-il de réduire les hypothèses de stationnarité, par exemple en se bornant à une stationnarité locale permettant un krigeage ordinaire, ou krigeage à moyenne inconnue. Cependant la pente de régression linéaire p = cov(Z,Z*)/var(Z*) n'a plus de raison d'être voisine de l'unité (à moins évidemment que le poids

de la moyenne en KS soit nul auquel cas KO = KS). Or on peut montrer qu'un estimateur quelconque, qui serait plus variant que la quantité à estimer, présente une pente de régression inférieure à 1: c'est ce qui se produit en KO à partir d'une seule donnée, l'estimateur étant alors égal à la valeur de la donnée. Mais de façon plus pratique, il existe beaucoup de cas où la pente de régression est trop faible, par la faute du voisinage trop restreint, et où cette pente s'améliore en agrandissant ce voisinage. Ainsi, au final, il peut être utile d'utiliser les deux paramètres théoriques que sont le poids de la moyenne en KS et la pente de régression linéaire du KO pour choisir le voisinage de KO et se prémunir en particulier d'un biais conditionnel trop important (Rivoirard 1984 b (thèse), 1987 b). En particulier un voisinage fournissant un poids de la moyenne en KS λ_m proche de 0, donnera une pente de régression théorique p proche de 1 : en effet p est nécessairement compris entre 1- λ_m et 1, que λ_m soit positif ou non (Chilès et Delfiner, 1999). L'erreur quadratique moyenne observée expérimentalement en validation croisée (on estime chaque donnée par ses voisines) peut aussi servir à délimiter un voisinage. Partant d'un voisinage trop petit, elle aura tendance à diminuer (et la précision à augmenter) si l'on aggrandit ce voisinage. Mais elle aura tendance à augmenter à nouveau si le voisinage commence à être trop grand vis-à-vis d'une stationnarité seulement locale (non publié).

A l'heure actuelle, la question plus générale du choix du voisinage de krigeage est toujours un thème de recherche. Depuis longtemps, le choix des données voisines est guidé, non seulement par la proximité à la cible, mais aussi par leur direction, vues de la cible (Isaaks and Srivastava, 1989 ; Chilès et Delfiner, 1999). Ceci est en particulier fort utile dans le cas de données très denses le long de lignes, pour éviter des voisinages déséquilibrés par la sélection d'un grand nombre de données situées d'un même coté de la cible et provenant d'une même ligne. Cependant, même dans des configurations plus classiques, les données les plus proches ne sont pas nécessairement les plus utiles (Emery, 2009 a).

Références

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Emery, X., 2009 a, The kriging update equations and their application to the selection of neighboring data, Computational. Geosciences, 13:269–280.

Isaaks, E.H., Srivastava, R.M., 1989, An introduction to applied geostatistics. Oxford University Press, New York.

Journel, A. G., and Huijbregts, Ch. J., 1978. Mining geostatistics. Academic Press, London. x + 600 p.

Krige, D.G., 1951, A Statistical. Approach to some Mine valuation and allied problems on the Witwatersrand, Master Thesis, University of Witwatersrand.

Matheron, G., 1955, Applications des méthodes statistiques à l'évaluation des gisements, Annales des Mines, Paris, No 12, 50-75.

Matheron, G., 1965, Les variables régionalisées et leur estimation. Masson, Paris. 306 p.

Matheron, G. 1970, La théorie des variables régionalisées et ses applications, Cahiers du Centre de Morphologie Mathématique 5. Ecole des Mines de Paris, Fontainebleau.

3.2 Eliminer les discontinuités de cartes krigées

Dans cette courte section, je résume le papier suivant:

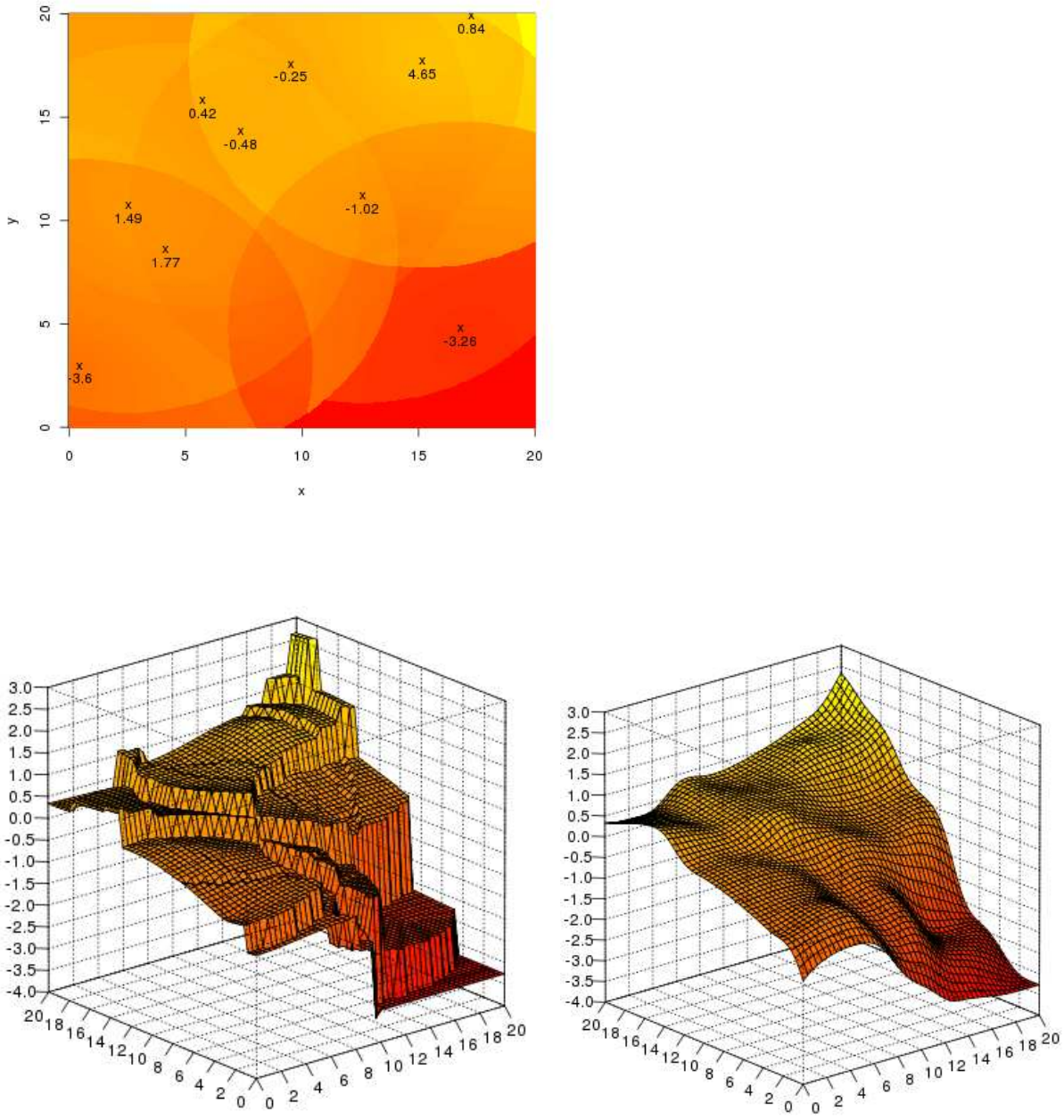
Rivoirard J., and Romary T., 2011, Continuity for kriging with moving neighborhood, Mathematical geosciences 43: 469–481.

L'utilisation d'un voisinage glissant, quand celui-ci a une influence sur les poids, présente cependant un inconvénient dans le cas d'une estimation ponctuelle : celui de donner des discontinuités sur la carte krigée, provoquées par des données sortant ou rentrant dans le voisinage glissant. De telles discontinuités ne sont pas réalistes, ni souhaitables dans certaines applications comme la modélisation numérique de terrains utilisée en hydrologie. Une méthode assez compliquée, basée sur des transformations *ad hoc*, avait été proposée par Gribov et Krivoruchko (2004).

Avec Thomas Romary, alors post-doc au Centre de Géosciences, nous avons développé une méthode simple et générale pour éliminer de telles discontinuités, même dans des cas non stationnaires (Rivoirard et Romary, 2011) : Fig. 10. Il s'agit de pénaliser les valeurs les plus éloignées dans le voisinage, de façon à ce que leur poids soit nul lorsqu'elles sont abandonnées par le voisinage glissant. Cette méthode se trouve offrir un formalisme très intéressant à la méthode proposée par Gribov et Krivoruchko (2004). La pénalisation des données apparaît comme un moyen extraordinairement souple et puissant, et offre des perspectives tout à fait intéressantes en terme de mise à jour de carte comme dans le cas de grands jeux de données.

Références

Gribov, A., and Krivoruchko, K., 2004, Geostatistical mapping with continuous moving neighbourhood. Math Geol 36(2):267-281.



Figures 10 : Exemple 2D synthétique avec 10 points de données. Dans l'ordre : Krigeage ordinaire, le voisinage glissant (rayon = 10) étant responsable de discontinuités ; idem en vue perspective ; krigeage continu. (Extrait de Rivoirard et Romary, 2011.)

3.3 Cokrigeage et autokrigeabilité

Dans cette section, je résume certains résultats sur le cokrigeage, propres au modèle à résidu, développés dans les papiers suivants :

Bordessoule, J.-L., Demange, C., Rivoirard, J., 1988, Estimation de réserves géologiques en uranium par variables utiles à résidu autokrigeable. *Sciences de la Terre*, série Informatique géologique, *28*, 27-51.

Bordessoule, J.-L., Demange, C., Rivoirard, J., 1989, Using an orthogonal residual between ore and metal to estimate in-situ uranium resources. In M. ARMSTRONG (ed.), *Geostatistics*, Vol. 2. Dordrecht : Kluwer, 923-934. (Voir ANNEXE.)

Le cokrigeage, estimation linéaire optimale de plusieurs variables, s'utilise dans des circonstances diverses (Chilès et Delfiner, 1999 ; Wackernagel, 1998). L'analyse krigeante, qui permet la cartographie des facteurs de décomposition d'une ou plusieurs variables, utilise par exemple une forme de cokrigeage. Le cokrigeage est par ailleurs très utile pour le conditionnement des simulations multivariables de transformées gaussiennes. Mais naturellement le cokrigeage peut être aussi utilisé pour estimer simultanément plusieurs variables dans leur distribution originelle non gaussienne, de même que des transformées non gaussiennes (des indicatrices par exemple). Il peut alors être d'un maniement délicat. Ainsi un cokrigeage d'indicatrices présentera couramment des valeurs estimées hors de (0,1), et de façon souvent plus marquée qu'un krigeage d'indicatrice. Egalement, les variables secondaires intervenant par leurs accroissements (positifs ou négatifs) en cokrigeage ordinaire, le cokrigeage de variables positives sera plus facilement négatif que leur krigeage (Isaaks and Srivastava, 1989).

Néanmoins le cokrigeage permet parfois des estimations, non seulement plus précises par construction que le krigeage (à voisinage identique), mais également plus cohérentes. Le cas du gisement d'uranium de Tréviels en donne un exemple original et intéressant (Bordessoule et al. 1988, 1989). On a estimé le minerai T et le métal Q, et déduit la teneur Q/T, pour des blocs de 25x25x5m à partir d'une maille de 25x25m identique à celle des sondages d'exploration, et comparé les résultats aux valeurs « vraies » calculées à l'aide de l'ensemble des trous de tir de chaque bloc (Fig. 11). En ce qui concerne les valeurs vraies de blocs, il ressort que la teneur Q/T est légèrement corrélée au minerai (la teneur de quelques blocs contenant très peu de minerai étant plus basse). Le cokrigeage rend bien le peu de lien entre teneur et minerai, mais avec des teneurs peu dispersées: manifestement les teneurs sont trop variables pour pouvoir être estimées avec précision. Par contre le krigeage séparé de T et de Q fournit une teneur qui décroît nettement lorsque le minerai augmente, ce qui est irréaliste. Ce phénomène s'explique par le contraste entre la structure de T, plus continue, et celle de Q, plus erratique. Le lissage, l'attraction vers la moyenne est plus important pour Q que pour T, d'où le résultat. Dans cet exemple, le cokrigeage tire profit du contraste entre les structures pour renforcer la cohérence des résultats.

L'exemple de Tréviels a été l'occasion de redécouvrir une propriété passée inaperçue jusque là et figurant dans une note de Matheron (1979) : l'autokrigeabilité (Wackernagel, 1998). Est autokrigeable une variable dont le cokrigeage est identique au krigeage, pour toute configuration isotopique (c'est-à-dire dans laquelle toutes les variables sont informées aux mêmes points de

données, contrairement au cas hétérotopique). Cette dernière précision (variables toutes connues aux mêmes points) est importante : il est clair que l'identité entre krigeage et cokrigeage n'est plus garantie si des données viennent à manquer sur la variable autokrigeable. Par contre les données sur les autres variables peuvent manquer sans affecter l'identité puisqu'elles reçoivent sinon un poids nul en cokrigeage. La condition nécessaire et suffisante d'autokrigeabilité (krigeage et cokrigeage simple ou ordinaire) est la proportionnalité du variogramme croisé avec le variogramme simple de la variable autokrigeable. L'autokrigeabilité correspond donc au modèle à résidu décrit plus haut, la variable maîtresse étant la variable autokrigeable.

Dans le cas de Tréviels, le minerai T, variable maîtresse, est autokrigeable : son cokrigeage est égal à son krigeage. En outre, le métal Q peut se décomposer en minerai et résidu :

$Q = mT + R,$

ces deux variables n'étant pas spatialement corrélées. Il s'ensuit que le cokrigeage de Q peut s'obtenir par un krigeage séparé du minerai T et du résidu $R = Q - mT$ calculé aux points de données.

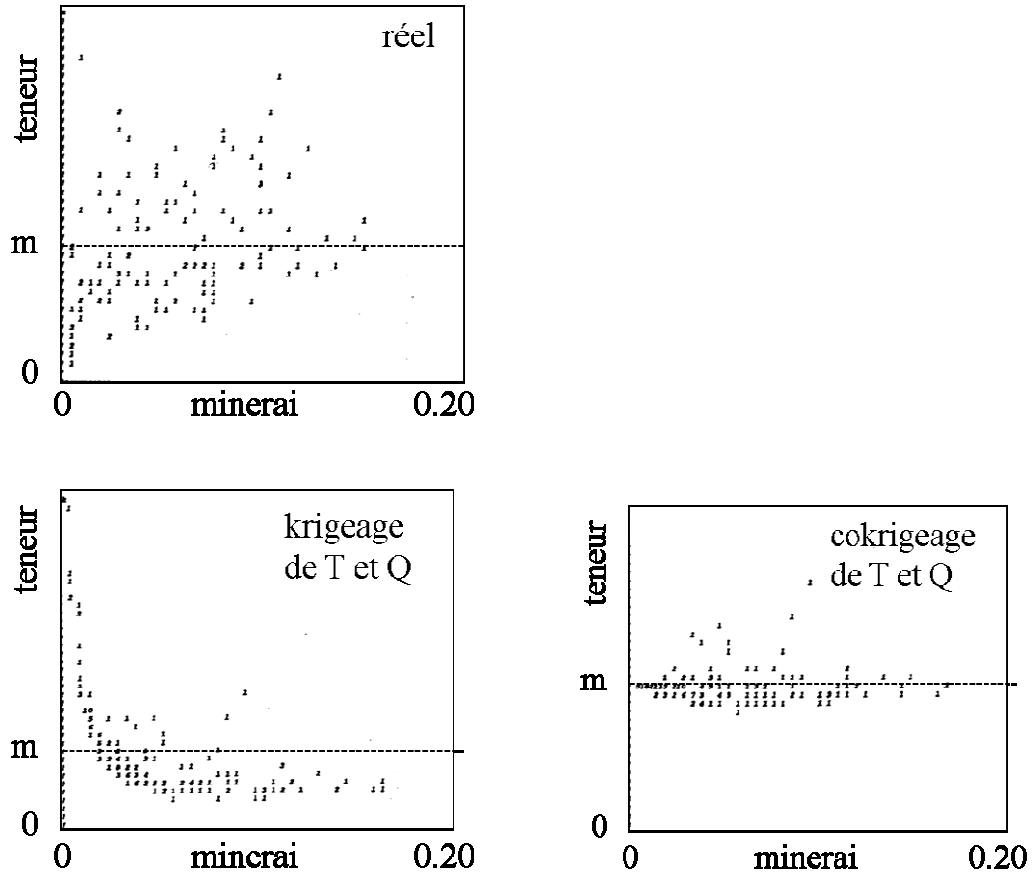
Références

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Isaaks, E.H., Srivastava, R.M., 1989, An introduction to applied geostatistics. Oxford University Press, New York.

Matheron, G., 1979, Recherche de simplification dans un problème de cokrigeage, Technical Report N-628, Centre de Géostatistique, Fontainebleau, France, 19 p.

Wackernagel, H., 1998. Multivariate geostatistics: an introduction with applications, 2nd ed., Springer, Berlin, 291 p.



Figures 11. Nuages de corrélation entre la proportion de minerai (selon x) et la teneur de ce minerai (selon y) pour des panneaux de 25 x 25 x 5 m. Dans l'ordre : valeurs « vraies » obtenues par maille fine 3 x 3 m de trous de tir ; valeurs déduites des krigeages séparés du minerai et du métal à partir d'une maille de sondages de 25 x 25 m ; valeurs déduites du cokrigeage de ces variables. (D'après Bordessoule et al., 1989.)

3.4 Le voisinage de cokrigeage, dans la prise en compte de variables auxiliaires

Je résume dans cette section les resultants essentiels décrits dans les papiers suivants :

Rivoirard, J., 2001, Which models for collocated cokriging ? *Mathematical geology*, vol. 33, no. 2, 117-131.

Rivoirard, J., 2002 a, On the structural link between variables in kriging with external drift. *Mathematical geology*, vol. 34, no. 7,797-808.

Rivoirard, J., 2004, On some simplifications of cokriging neighbourhood. *Mathematical geology*, vol. 36, no. 8, 899-915. (Voir ANNEXE.)

La qualité d'un krigeage ou cokrigeage est souvent liée au choix de son voisinage. Pour le krigeage, le choix d'un voisinage est essentiellement fonction de l'implantation et en particulier de la densité des points de données, ainsi que de la structure variographique, et nous avons vu des paramètres pouvant aider à ce choix. Choisir un voisinage de cokrigeage peut être beaucoup plus difficile : les différentes variables ne sont pas forcément connues toutes aux mêmes points (isotopie ou hétérotopie), et on est amené à mesurer et à traiter des variables de plus en plus nombreuses et des données de plus en plus denses. Il s'agit donc de sélectionner les données à bon escient, et de retenir celles qui apportent le maximum d'information. Ce n'est pas une mince affaire. On conçoit que le voisinage peut dépendre de la variable : on ira chercher loin les valeurs d'une variable peu informée, mais on se contentera de la richesse d'information d'une variable dense à proximité du point cible à estimer. Toutefois on essaiera également de caler les variables entre elles en retenant également les valeurs de la variable dense là où la variable peu informée est connue. D'où l'importance de la colocation dans le choix du voisinage. A noter que le voisinage n'est plus nécessairement une région de l'espace dont toutes les données sont utilisées pour l'estimation, mais l'ensemble des données géométriquement réparties retenu pour l'estimation.

Un exemple classique de configuration hétérotopique, emprunté au pétrole, est celui de la profondeur d'un horizon géologique connue de façon précise mais éparse par des puits peu nombreux, et par une sismique connue de façon dense (pratiquement, partout où l'on veut). Cet exemple est à l'origine du krigeage en dérive externe, la sismique étant utilisée pour donner la forme de la dérive de l'horizon (les détails de la méthode sont bien documentés dans Chilès et Delfiner (1999)). Dans cette méthode extrêmement pratique et populaire, la donnée sismique (variable secondaire) est utilisée seulement à l'endroit des puits (où l'on connaît la variable primaire) ainsi qu'au point cible : il s'agit d'un voisinage colocalisé, duquel est exclue la variable secondaire aux autres points de l'espace.

Le cokrigeage colocalisé (collocated cokriging) a connu une énorme vogue dans le milieu géostatistique pendant les années 90. Il s'agissait, au tout départ, de ne retenir de la variable secondaire, supposée dense, que sa valeur au point cible (Xu et al., 1992). Dans un modèle où la structure croisée était proportionnelle à la structure de la variable primaire (baptisé alors modèle de Markov, puis Markov 1), seule était alors nécessaire, de la corégionalisation, le coefficient de corrélation entre les variables qui fixe l'amplitude de la structure croisée, la structure de la variable secondaire n'étant pas utilisée. Bien vite il apparut que l'usage de la seule colocation au point cible était pauvre, et le cokrigeage colocalisé fut étendu au même voisinage colocalisé que le krigeage en dérive externe (ceci ne semble pas avoir fait l'objet d'une publication spécifique). Le modèle dit de Markov, associé au cokrigeage colocalisé, ne garantit cependant pas que le cokrigeage complet pût, sous ces hypothèses, se réduire au cokrigeage colocalisé, comme l'ont remarqué certains auteurs (Wackernagel, 1998 ; Chilès et Delfiner, 1999). En effet, dans ce modèle, les données de la variable secondaire là où la variable primaire est inconnue n'ont pas nécessairement des poids nuls, et les négliger peut représenter une perte substantielle d'information, notamment en simulation lorsque l'on cherche à reproduire les structures (Haas et al., 1998).

J'ai montré qu'en réalité, le cokrigeage colocalisé n'est rigoureusement un cokrigeage complet que dans un modèle à résidu dont la variable maîtresse est la variable secondaire, non la variable primaire (modèle Markov 2) (Rivoirard, 2001). Il s'agit d'ailleurs au fond à peu près des mêmes hypothèses que celles du krigeage en dérive externe, basé sur le même voisinage colocalisé, comme je l'ai expliqué dans Rivoirard (2002 a). Par ailleurs dans ce modèle à résidu, le cokrigeage, alors cokrigeage colocalisé, n'est autre que la traditionnelle technique du krigeage du résidu (Rivoirard, 2001). Hors de ce modèle, et toujours dans le cas d'une variable secondaire connue partout où l'on veut, le cokrigeage colocalisé est une version théoriquement plus précise qu'un simple krigeage du résidu, car il correspond à un cokrigeage du résidu sur le même voisinage colocalisé, mais il ne correspond pas à un cokrigeage complet.

Fort de ces résultats, je me suis intéressé à des conditions plus générales dans lesquelles le voisinage de cokrigeage pouvait se simplifier (Rivoirard, 2004). De façon malicieuse, il apparaît, comme je l'ai montré, que le modèle de Markov 1 (la variable primaire étant la variable maîtresse), qui ne justifie donc pas le cokrigeage colocalisé, justifie au contraire le cokrigeage délocalisé à moyennes connues (« dislocated cokriging ») dans le cas d'une structure résiduelle pépitique : en effet, là où la variable primaire est connue, la connaissance de la variable secondaire équivaut à celle du résidu, ce qui n'apporte pas d'information, et c'est seulement là où la variable primaire est inconnue, que la variable secondaire peut apporter de l'information… Ce modèle n'est d'ailleurs nullement une vue de l'esprit, il correspond en effet à une situation où les mesures exactes seraient rares (variable primaire) et où on disposerait par ailleurs de mesures entâchées d'une erreur pépitique et indépendante (variable secondaire). Une simplification moindre est également présente en cokrigeage ordinaire, c'est-à-dire à moyennes inconnues, les quantités indépendantes participant alors à l'estimation implicite des moyennes. J'ai finalement recensé un certain nombre de modèles, bi- ou multi-variables, et de configurations géométriques dans lesquels des simplifications s'opéraient (Rivoirard, 2004). Les configurations géométriques sont basées sur la distinction entre les ensembles de points où variables primaire et secondaires sont connues ensemble ou séparément, ainsi que sur le fait que chacun de ces ensembles est vide ou non. Les modèles sont essentiellement des modèles à résidus, dans lesquels variable maîtresse ou résidu sont éventuellement pépitiques.

Il est assez remarquable que de tels résultats, somme toute assez basiques, du moins après coup, n'aient pas vu leur apparition plus tôt. J'y vois au moins une raison. Il était connu depuis longtemps que le cokrigeage se simplifiait en krigeage dans le cas isotopique dans un modèle où toutes les structures sont proportionnelles (Matheron, 1965). Cela valait pour toutes les variables à la fois, mais a sans doute occulté la raison profonde qui est la décomposition possible en un modèle à résidu, chaque variable pouvant être considérée tour à tour comme variable maîtresse. Matheron (1965) avait qualifié ce modèle à structures proportionnelles de corrélation intrinsèque : il entendait par là que le coefficient de corrélation entre deux variables était le même, quel que soit le support des variables et le champ parcouru par le support, et avait donc une signification intrinsèque. Hors de ce modèle en effet, le coefficient de corrélation entre deux variables ne présente, en fait, pas grande signification. Dans le modèle à résidu, c'est la pente de la régression linéaire de la variable secondaire sur la variable primaire qui présente ce caractère intrinsèque : hors du modèle à résidu, cette pente n'a qu'une signification relative.

Références

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Haas, A., Biver, P., et Moulière, D., 1998. Simulations stochastiques en cascade, *in* Cahiers de Géostatistique, Fasc. 6, ENSMP, 31-43.

Matheron, G. 1965. Les variables régionalisées et leur estimation. Masson, Paris. 306 p.

Wackernagel, H., 1998. Multivariate geostatistics: an introduction with applications, 2nd ed., Springer, Berlin, 291 p.

Xu, W., Tran, T. T., Srivastava, R. M., and Journel, A. G. 1992, Integrating seismic data in reservoir modeling: The collocated cokriging alternative. SPE paper 24742, 67th Annual Technical Conference and exhibition, 833-842.

4 Les modèles non-linéaires

4.1 L'estimation au support ponctuel et les dépassements de seuils

Je résume dans cette section les résultats décrits dans le papier suivant (voir ANNEXE):

Rivoirard, J., 1989, Models with orthogonal indicator residuals. In M. Armstrong (ed.), *Geostatistics*, Dordrecht : Kluwer, 1989, vol. 1, 91-107.

La géostatistique est utile pour estimer des fonctions non-linéaires de la variable (dépassement de seuil notamment), ainsi que pour accéder à des estimateurs non-linéaires plus puissants qu'un simple krigeage, ou à la variance conditionnelle. On sait que, dans un modèle (multi-)gaussien, les lois conditionnelles sont également gaussiennes, avec une espérance conditionnelle linéaire, qui correspond donc au krigeage, et une variance conditionnelle, identique à la variance non conditionnelle, indépendante des valeurs des données. On a donc directement accès, dans le cadre de ce modèle, aux probabilités conditionnelles de dépasser tel ou tel seuil, comme à l'estimateur le meilleur (au sens des moindres carrés), et à la variance conditionnelle. Moyennant une transformation de la variable, et toujours sous des hypothèses multi-gaussiennes de la transformée, ces heureuses propriétés se conservent (modèle gaussien anamorphosé). Ce modèle multi-gaussien a été popularisé par Verly (1983).

Dans les autres situations, on fait appel de façon très générale, à des techniques basées sur les indicatrices (par exemple $\text{ind}(Z(x) \geq z)$ pour différents seuils z). Il peut ainsi s'agir du Krigeage Disjonctif, théoriquement un cokrigeage des indicatrices (Matheron, 1976 a). Une panoplie impressionnante de modèles ont été proposés par Matheron dans les années 70 : ce sont des modèles *isofactoriels*, dans lesquels le cokrigeage se simplifie en krigeage séparé de facteurs. Ainsi les modèles bi-gaussiens, hermitiens, gamma, binomial négatif, Poisson… (je renvoie à Chilès et Delfiner (1999) pour la bibliographie des œuvres de Matheron sur ce sujet). Dans ces modèles, les indicatrices ne sont plus au devant de la scène, et les hypothèses restent marquées par le type de distribution auquel on fait appel pour la variable régionalisée ou une transformée. Un modèle de diffusion discret a toutefois été également proposé par Matheron (1984) et mis en œuvre par Lajaunie et Lantuéjoul (1989), qui permet une approche plus empirique.

La mise en œuvre de tous ces modèles est cependant, et était surtout à l'époque, une opération assez lourde et délicate. D'où l'approche basée sur un simple krigeage séparé des indicatrices, proposée par l'Ecole de Stanford (Journel, 1982), approche résolument plus simple, même si elle n'est en général pas optimale et ignore les relations structurales pouvant exister entre différentes indicatrices (lesquelles sont toujours corrélées). Matheron (1982) a montré que le krigeage des indicatrices était optimal, identique alors au cokrigeage d'indicatrices, dans le modèle mosaïque à valuations indépendantes. Dans ce modèle, toutes les structures sont identiques, et il n'existe pas ce phénomène courant qu'est la destructuration des hautes (ou basses) valeurs, autrement dit la dégradation du variogramme dans certaines gammes de coupures.

J'ai apporté ma pierre à cet édifice dans Rivoirard (1989), en m'intéressant à ce qu'on peut appeler les effets de bord, pour décrire les relations entre indicatrices. Si $z' > z$, alors le rapport entre le variogramme croisé des indicatrices au-dessus des coupures z et z' et le variogramme

simple de l'indicatrice à coupure z, représente la probabilité que Z(x) dépasse z', sachant que Z(x) dépasse déjà z mais que Z(x+h) est inférieur à z :

$$\frac{\gamma_{1_{Z\geq z}1_{Z\geq z'}}(h)}{\gamma_{1_{Z\geq z}}(h)} = P(Z(x)\geq z' \mid Z(x)\geq z, Z(x+h)<z)$$

Au fond, on s'intéresse à la probabilité de dépasser un seuil plus élevé, à l'intérieur de l'ensemble constitué par les valeurs supérieures à z, et ceci en fonction de la plus ou moins grande proximité aux frontières de cet ensemble. Cependant on notera que la distance à la frontière n'intervient pas explicitement dans l'outil proposé, et on n'essaie pas, par exemple, de mesurer la probabilité de dépassement de z' en fonction d'une telle distance. On se borne à imposer que le point voisin x+h reste à l'extérieur de cet ensemble. L'outil bi-point proposé est naturellement beaucoup plus commode pour l'analyse structurale à partir d'un ensemble discret de points de données, qui ne permet en général pas de mesurer directement un effet de bord en fonction de la distance à une frontière.

Un modèle particulièrement intéressant est alors celui dans lequel il n'existe aucun effet de bord, lorsqu'on monte en valeur (Rivoirard 1989) (ce modèle se transpose naturellement à une absence d'effet de bord en descendant). Une probabilité telle que celle ci-dessus est alors constante, elle ne dépend plus de h. Dans ce modèle, les lois bivariables vérifient la propriété remarquable :

$$P(Z(x)\geq z', Z(x+h)\geq z) = P(Z(x)\geq z' \mid Z(x)\geq z)\, P(Z(x)\geq z, Z(x+h)\geq z)$$
$$= \frac{T(z')}{T(z)} P(\min(Z(x), Z(x+h))\geq z)$$

où T(z) représente la loi marginale $P(Z(x)\geq z)$. J'ai montré qu'un tel modèle est nécessairement de type mosaïque, en ce sens que la probabilité d'observer la même valeur de la variable régionalisée en des points voisins n'est pas nulle. Les fonctions booléennes à fonction primaire cylindrique sont un exemple de modèle vérifiant cette absence d'effet de bord (Jeulin, 1991). Vue l'absence d'effet de bord, la structure croisée des deux indicatrices étant identique à la structure simple de la première, ces indicatrices obéissent à un modèle à résidu, donc hiérarchique, dans lequel la variable maîtresse est la première indicatrice, celle à coupure la plus basse. L'indicatrice à coupure plus haute apparaît alors comme la somme de la première indicatrice, à laquelle elle est subordonnée, et du résidu entre indicatrices :

$$\frac{1_{Z(x)\geq z'}}{T(z')} = \frac{1_{Z(x)\geq z}}{T(z)} + R(x)$$

avec $R(x) = \frac{1_{Z(x)\geq z'}}{T(z')} - E\left[\frac{1_{Z(x)\geq z'}}{T(z')} \mid 1_{Z(x)\geq z}\right] = \frac{1_{Z(x)\geq z'}}{T(z')} - \frac{1_{Z(x)\geq z}}{T(z)}$

Le modèle s'étend à un nombre quelconque d'indicatrices, correspondant par exemple aux seuils de discrétisation de la variable régionalisée. A l'intérieur de l'ensemble constitué par les valeurs au-dessus d'un seuil quelconque, il y a absence d'effet de bord.

Ce modèle se factorise à l'aide des résidus d'indicatrices successifs. La coupure la plus basse fournit la pièce maîtresse, puis on passe à la coupure immédiatement au-dessus par son résidu d'indicatrices, et ainsi de suite :

$$\frac{1_{Z(x)\geq z_{i+1}}}{T_{i+1}} = \frac{1_{Z(x)\geq z_1}}{T_1} + \left(\frac{1_{Z(x)\geq z_2}}{T_2} - \frac{1_{Z(x)\geq z_1}}{T_1} \right) + \ldots + \left(\frac{1_{Z(x)\geq z_{i+1}}}{T_{i+1}} - \frac{1_{Z(x)\geq z_i}}{T_i} \right)$$

Il s'agit donc d'un modèle isofactoriel, dans lequel les facteurs sont les résidus d'indicatrices. Il suffit de kriger séparément l'indicatrice la plus basse et les résidus d'indicatrices successifs pour obtenir le cokrigeage des indicatrices et donc l'estimation par Krigeage Disjonctif.

Ce modèle isofactoriel, mais basé sur une construction géométrique et non sur une loi de distribution statistique, rend précisément bien compte de certaines régionalisations marquées par la géométrie. Il peut s'appliquer à des gisements en veines comme le gisement d'uranium de Tréviels décrit plus haut (et d'ailleurs le modèle à résidus d'indicatrices est une extension du modèle minerai-métal, l'indicatrice à coupure z représentant le minerai ponctuel à coupure z). Dans le domaine halieutique, la distribution du hareng en Mer du Nord, par exemple, répond bien aux hypothèses de ce modèle (absence d'effet de bord en montant, détectable par comparaison des variogrammes simples et croisés d'indicatrices) (thèse de Petitgas, 1991 ; Petitgas, 1993).

Finalement on peut distinguer deux grandes familles parmi les modèles de la géostatistique non-linéaire : les modèles de type diffusif (gaussien, gamma, etc), et les modèles de type mosaïque. Dans cette dernière famille se rangent le modèle mosaïque à valuations indépendantes ainsi que le modèle à résidu d'indicatrices. On remarquera que le modèle mosaïque à valuations indépendantes ne présente aucun effet de bord, ni en montant, ni en descendant. Mais également que le modèle à résidus d'indicatrices, s'il ne présente pas d'effet de bord en montant, en présente cependant en descendant : Fig. 12.

Références

Jeulin, D., 1991. Modèles morphologiques de structures aléatoires et de changement d'échelle, Thèse de Docteur ès Sciences Physiques, Université de Caen.

Journel, A.G., 1982. The indicator approach to estimation of spatial data. Proceedings of the 17th APCOM, Port. City Press (New York), 793-806.

Lajaunie, C., et Lantuéjoul, C., 1989. Setting up the general methodology for discrete isofactorial models. *In* Geostatistics, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 323-334.

Matheron, G., 1976 a. A simple substitute for conditional expectation: the disjunctive kriging, in Advanced Geostatistics in the Mining Industry, M. Guarascio, M. David, and C. Huijbregts eds. Reidel, Dordrecht, Holland, 221-236.

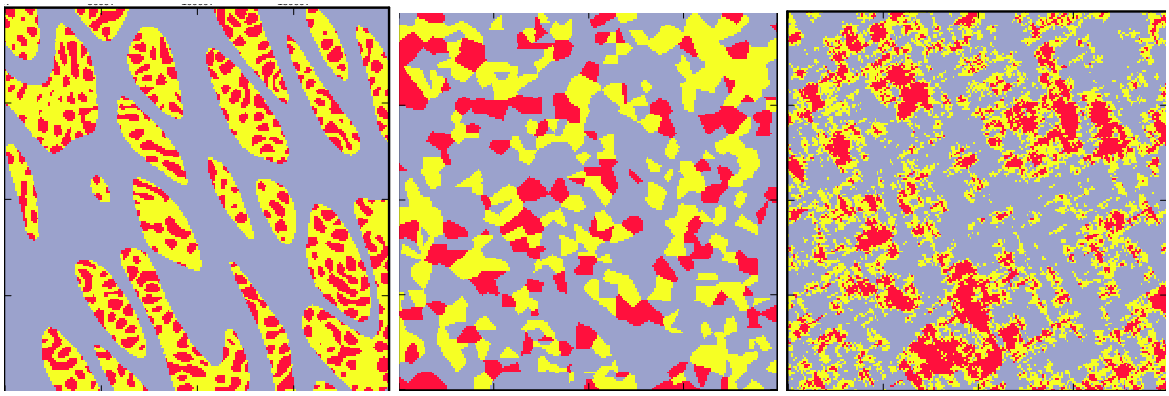
Matheron, G., 1982. La destructuration des hautes teneurs et le krigeage des indicatrices, Ecole des Mines de Paris.

Matheron, G., 1984. Une méthodologie générale pour les modèles isofactoriels discrets. Science de la Terre, Série Informatique Géologique, 21, 1-21.

Petitgas P., 1991. Contributions géostatistiques à la biologie des pêches maritimes. Thèse de Doctorat en Géostatistique, Ecole des Mines de Paris, 211 p.

Petitgas, P. 1993. Use of disjunctive kriging to model areas of high pelagic fish density in acoustic fisheries surveys. Aquatic Living Resources, 6: 201-209.

Verly, G., 1983. The multigaussian approach and its applications to the estimation of local reserves, Math. Geol. Vol. 15, No 2, 259-286.



Figures 12. Trois modèles avec mêmes proportions de valeurs faibles (gris), moyennes (jaune), et fortes (rouge). De gauche à droite : modèle à résidus d'indicatrices (absence d'effets de bords en montant) ; modèle mosaïque à valuations indépendantes (aucun effet de bord) ; modèle de diffusion (gaussienne seuillée, le jaune sert de transition entre gris et rouge). Figures réalisées par Hélène Beucher.

4.2 L'écrêtage dans les mines d'or

Je résume dans cette section les papiers suivants:

Rivoirard, J. 2002 b. Analyse de la sensibilité à l'écrêtage à une coupure haute dans l'évaluation de l'or récupérable. Rapport interne, Centre de Géostatistique, Ecole des Mines de Paris, 25 p. Confidentiel.

Rivoirard, J., Demange, C., Freulon, X., Lécureuil, A., Bellot, N., A top-cut model for deposits with very skewed grade distribution, submitted to Mathematical Geosciences.

L'influence des valeurs extrêmes pose couramment problème au niveau de la variographie (voir section 2.2) et bien évidemment aussi au niveau de l'estimation. Dans les gisements d'or, caractérisés par une distribution de teneurs très dissymétrique avec valeurs extrêmes, une pratique courante consiste à écrêter les valeurs à un seuil. Ceci permet d'éviter d'estimer des blocs par des valeurs irréalistement fortes. Cependant l'écrêtage soulève deux questions. Premièrement, à moins que les valeurs fortes ne soient dues à des erreurs, l'intégralité des valeurs devrait être prise en compte pour une estimation globale : dès lors où est passée la partie de métal ôtée par l'écrêtage ? Comme toute sous-estimation, cette pratique peut réserver de bonnes surprises (au final, trouver plus qu'attendu), comme elle peut avoir pour conséquence le rejet de projets en réalité viables. Deuxièmement, si écrêtage il convient de faire, à quel seuil ?

Le choix du seuil d'écrêtage est en général assez arbitraire, même si la technique relève d'un « reasonable compromise » (selon Chilès et Delfiner (1999)), entre conservation et suppression des valeurs extrêmes. Certains auteurs (Snowden et al., 1994) adoptent un seuil minimal égal au quantile 97.5% (lequel serait censé éliminer un biais liée à la moyenne arithmétique en lognormal). Parker (1991) propose une approche plus élaborée qu'un simple écrêtage: la population des teneurs y est vue comme la combinaison d'un histogramme d'allure modérée et d'une queue de distribution très dissymétrique (et estimée par une approche lognormale). Les deux populations sont conjuguées via une indicatrice, pour l'estimation des blocs, de sorte qu'il n'y a pas de perte globale de métal. Le choix du seuil correspond à un décrochage de statistiques (coefficient de variation). Il n'y a cependant pas de validation structurale du modèle sous-jacent.

Ce problème d'écrêtage m'a été posé par Aréva (alors Cogéma), qui a mis à ma disposition plusieurs jeux de données minières réels. Une analyse fine de la variographie de ces jeux m'a conduit à proposer un modèle d'écrêtage (Rivoirard, 2002 b), dans lequel la teneur se décompose en trois termes :

$Z = Z^z + k\, 1\{Z>z\} + R$

- l'écrêtée, minimum entre teneur et seuil ;

- un second terme proportionnel à l'indicatrice $1\{Z>z\}$;

- un résidu R, seul terme dans lequel apparaissent individuellement les valeurs fortes de Z.

En moyenne, le résidu est nul, et le deuxième terme est égal au bénéfice conventionnel de la géostatistique à coupure z, qui représente ici paradoxalement la perte moyenne provoquée par l'écrêtage. Les deux premiers termes sont corrélés, mais un tel modèle est essentiellement intéressant s'il n'existe pas de corrélation spatiale entre le résidu et les deux premiers termes. Ceci a lieu lorsqu'il n'existe pas d'effet de bord dans l'ensemble des valeurs dépassant le seuil. D'où un moyen pratique de choisir un tel seuil lors de l'analyse variographique. Si de plus le résidu est pépitique, on obtient une estimation de Z par cokrigeage des deux premiers termes :

$Z^* = [Z^z]^{CK} + k\, [1\{Z>z\}]^{CK}$

Dans un tel modèle, l'estimation de l'écrêtée est complétée par celle de la crête, qui est distribuée là où les valeurs de Z ont le plus de chances de dépasser le seuil (estimation de l'indicatrice). Ce modèle a été appliqué effectivement sur plusieurs jeux de données à Aréva, notamment à l'occasion de stages, et est en cours de publication (Rivoirard et al.). En pratique il apparaît qu'on

peut choisir un seuil bien inférieur à celui couramment utilisé (par exemple 5 g/t au lieu de 50 g/t), ce qui serait inenvisageable si l'estimation se bornait à celle de l'écrêtée, vue la perte de métal occasionnée. L'estimation par cokrigeage permet une meilleure délimitation des zones les plus pauvres, et cantonne le minerai le plus riche dans les zones où leur présence est la plus probable (Fig. 13).

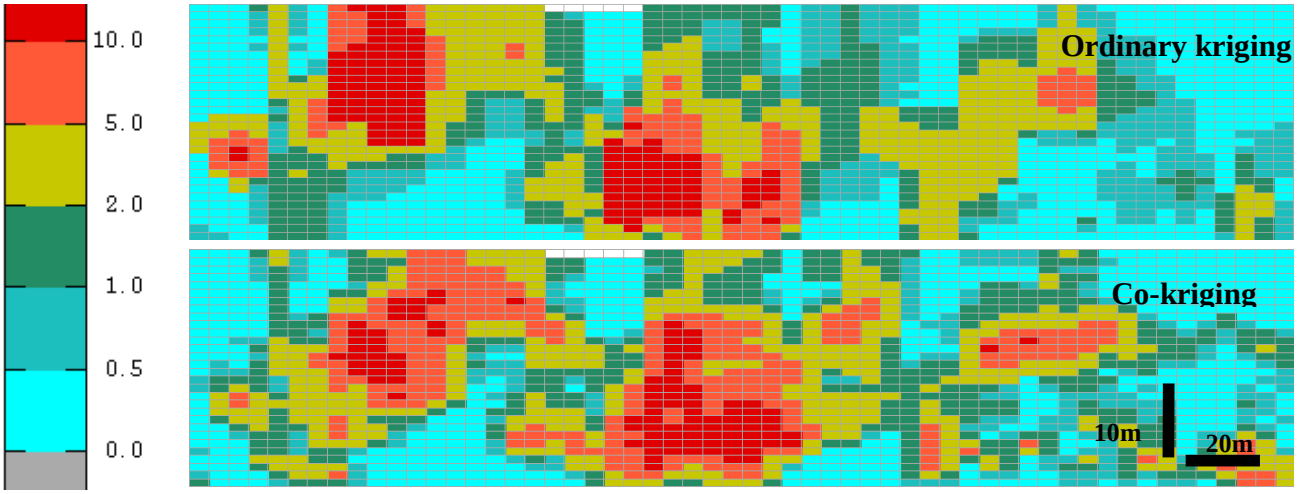
Ce modèle peut être vu comme une variante du modèle à résidus d'indicatrices, réduit à une coupure unique, mais dans lequel les valeurs sont conservées dans leur intégralité (elles ne sont pas simplement discrétisées en une valeur moyenne pour chacune des deux classes).

Références

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Parker, H. M., 1991, Statistical treatment of outlier data in epithermal gold deposit reserve estimation, Math. Geol., Vol. 23, No 2, 175-199.

Snowden, V., Moore, C., and Kelly, D., 1994, Using Geostatistics to Assist in Optimising Grade Control Estimation at KCGM's Fimiston Open Pit, Proceedings of the 4th Large Open Pit Mining Conference, Perth, 5-9 September 1994, 117-131.



Figures 13. Section verticale dans une mine d'or (blocs de 5 m de coté par 1 m de haut). Comparaison entre un krigeage ordinaire (haut) et un cokrigeage dans le modèle d'écrêtage (bas). Dans ce dernier cas, l'extension des valeurs élevées se fait là où leur présence est la plus probable.

4.3 Modélisation de l'histogramme : combler les queues de distribution

Dans cette section, je décris un moyen de mieux modéliser les histogrammes en géostatistique, présenté dans :

Deraisme, J., and Rivoirard, J., 2009, Histogram modelling and simulations in the case of skewed distributions with a 0-effect: issues and new developments, IAMG 2009 Conference, 23-29 August, Stanford (Powerpoint presentation).

La modélisation de l'histogramme est une opération indispensable, que ce soit en vue du changement de support (voir plus loin), ou de simulation, notamment à partir de transformées gaussiennes (Chilès et Delfiner, 1999). Très souvent, les données (éventuellement désagrégées) fournissent un histogramme assez proche de ce que serait l'histogramme exhaustif sur le champ considéré. Des différences existent néanmoins, responsables en particulier d'un écart sur les moyennes, dont la variance peut être calculée par la géostatistique linéaire dans certaines conditions (variance d'estimation globale : Matheron (1970), Journel et Huijbregts (1978)). Mais c'est surtout au niveau des queues de distribution (par exemple, les valeurs élevées et rares) que la modélisation est la plus critique. Cette partie de la distribution, peu utilisée dans le calcul classique des réserves récupérables aux coupures minières habituelles, est sollicitée dans le calcul d'intervalles de confiance, ainsi qu'en simulation. L'usage répandu des modèles transformés gaussiens $Z = \varphi(Y)$ a fait que la distribution de la variable d'intérêt Z est souvent modélisée via sa fonction d'anamorphose φ , dont la connaissance est équivalente à celle de la distribution. Pour des raisons en partie historiques, cette modélisation se fait très souvent par un ajustement par polynomes d'Hermite, indispensable au Krigeage Disjonctif gaussien, notamment pour la prévision des réserves récupérables (Matheron, 1976 a et b). Cependant l'ajustement des queues de distributions est problématique, d'autant plus qu'un développement polynomial tronqué est difficile à controler aux grandes valeurs de la gaussienne, comme je le décris dans Rivoirard (1985).

J'ai regardé à nouveau récemment ce problème avec la société Géovariances (en particulier Jacques Deraisme), qui souhaitait améliorer la modélisation des queues d'anamorphose gaussienne proposée par le logiciel géostatistique Isatis® dont elle assure le développement. Il est finalement apparu plus simple de revenir à la source, autrement dit, à l'histogramme empirique des données. J'ai proposé une façon de modéliser un tel histogramme, possédant des valeurs fortes isolées par exemple, qui consiste à augmenter sa continuité grâce à une dispersion de chacune des données autour de sa valeur : par exemple (pour un histogramme de valeurs positives) une dispersion lognormale dont la variance augmente avec le carré de la valeur. Ainsi modélisé, l'histogramme peut être discrétisé à souhait (et être ajusté par un développement polynomial au besoin). Dans cette opération de dispersion, la moyenne de l'histogramme est conservée, et la variance légèrement augmentée. Ceci est d'ailleurs souhaitable, car on doit s'attendre à ce que la variance de dispersion des données soit un peu inférieure à la variance ponctuelle dans le champ, d'une quantité en général faible, mais dont l'ordre de grandeur est précisément celui de la variance d'estimation de la moyenne : celle-ci peut donc servir à contrôler la dispersion des valeurs de l'histogramme (Deraisme and Rivoirard, 2009).

Références

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Journel, A. G., and Huijbregts, Ch. J. 1978. Mining geostatistics. Academic Press, London. x + 600 p.

Matheron, G. 1970, La théorie des variables régionalisées et ses applications, Cahiers du Centre de Morphologie Mathématique 5. Ecole des Mines de Paris, Fontainebleau.

Matheron, G., 1976 a. A simple substitute for conditional expectation: the disjunctive kriging, in Advanced Geostatistics in the Mining Industry, M. Guarascio, M. David, and C. Huijbregts eds. Reidel, Dordrecht, Holland, 221-236.

Matheron, G., 1976 b, Forecasting block grade distributions: the transfer functions. In Advanced geostatistics in the mining industry, M. Guarascio, M. David, and C. Huijbregts eds. Reidel, Holland, 237-251.

Rivoirard, J., 1985, Convergence des développements en polynômes d'Hermite. *Sciences de la terre, série Informatique géologique*, no. 24, 129-159.

4.4 Le changement de support : du bloc de sélection minier au panneau

Je résume dans cette section certains travaux sur la prévision des réserves récupérables :

Deraisme, J., Rivoirard, J., and Carrasco Castelli, P., 2008, Multivariate Uniform Conditioning and Block simulations with Discrete Gaussian Model: Application to Chuquicamata Deposit, VIII International Geostatistics Congress, GEOSTATS 2008, 69-78.

Rivoirard, J., 1984 c, Une méthode d'estimation du récupérable local multivariable, Note interne non confidentielle, Centre de Géostatistique, Ecole des Mines de Paris, 8 p.

Les techniques non-linéaires vues dans les sections précédentes permettent de prédire la distribution locale d'une variable, prise au support quasi-ponctuel de son échantillonnage. Or à supposer que la distribution globale de la variable régionalisée sur son champ soit suffisamment bien connue (histogramme), la distribution globale correspondant à un support plus important (le support d'intérêt, bloc de sélection minier par exemple) ne l'est pas. Et les distributions locales non plus. D'où l'utilité des corrections de support, ou des modèles de changement de support, développés à cet effet, qui permettent de prévoir d'une part la distribution globale, et d'autre part, à l'aide de techniques non-linéaires identiques à celles du cas ponctuel, les distributions locales.

Le modèle gaussien discret de changement de support est peut-être le plus utilisé (Matheron, 1976 b ; Chilès et Delfiner, 1999). Le champ est partitionné (« discrétisé ») en blocs identiques, et les échantillons sont supposés implantés aléatoirement et indépendamment dans leur bloc. Variables ponctuelles Z(x) comme variables sur support bloc Z(v) sont considérées comme

transformées de variables gaussiennes réduites Y(x) et Y_v , sur lesquelles on fait les hypothèses bi- ou multi-gaussiennes (le modèle se prête également bien aux simulations dans un cadre gaussien). Naturellement Z(v) est une régularisée de la variables ponctuelle Z(x). Cependant, malgré les transformations non-linéaires, Emery et Ortiz (2005) ont montré que, dans les hypothèses du modèle gaussien discret, la gaussienne réduite bloc n'était, elle aussi, autre que la régularisée (normée) de la gaussienne ponctuelle. Sans rentrer dans le détail de ce résultat (auquel je suis étranger), cette avancée récente, dans un domaine où l'essentiel date d'une bonne vingtaine d'années, ouvre de nouvelles possibilités pour le calcul des paramètres en jeu. De plus ces possibilités s'étendent aussi au cas multivariable, en particulier pour des simulations (Emery, 2009 b; Emery et Ortiz, 2010).

Une difficulté rencontrée dans l'application des techniques non-linéaires réside dans les hypothèses de stationnarité nécessaires (Rivoirard, 1994 a). Nous avons déjà parlé de ce problème lors du krigeage linéaire. La solution consistait alors à remplacer le krigeage simple par un krigeage ordinaire (avec cependant la difficulté de choisir un voisinage de façon à ne pas trop s'écarter du non-biais conditionnel). Pour pallier à ce problème en non-linéaire, une des solutions consiste à utiliser la technique du conditionnement uniforme (CU) par panneau (thèse de Armando Remacre (1984) dirigée par Matheron ; Guibal and Remacre, 1984 ; Rivoirard, 1994 a). En bref on cherche à prévoir la distribution des blocs dans un panneau, en conditionnant celle-ci par la valeur du panneau, dans le cadre d'un modèle gaussien discret élargi jusqu'à la taille de ce panneau (à noter que dans cette technique, l'inversion des valeurs d'échantillons en gaussiennes n'est pas nécessaire, ce qui favorise son utilisation lorsque par exemple les échantillons présentent un grand nombre de valeurs nulles). La distribution des valeurs de blocs dans le panneau est alors calée, par construction, à la valeur injectée du panneau. Celle-ci peut être alors choisie de façon à s'adapter à une stationnarité locale, par exemple un krigeage ordinaire du panneau. Cette technique me semblant prometteuse, j'en avais établi une version multivariable, devant permettre la prévision des réserves récupérables d'un gisement polymétallique par coupure sur l'une des variables (Rivoirard, 1984 c).

Depuis, l'usage du CU monovariable s'est répandu dans l'industrie minière, probablement du fait d'un bon compromis entre simplicité et performance, et ce, notamment grâce au travail de consultants tels que Daniel Guibal en Australie (Vann and Guibal, 1998). Récemment la méthode de CU multivariable que j'avais proposée en 1984 a suscité l'intérêt d'industriels miniers, et conduit la société Géovariances à créer en partenariat avec l'Ecole des Mines un consortium regroupant une demi-douzaine de compagnies minières (Multivariate Recoverable Resources Consortium, 2006-2009) dans le but initial d'implémenter la méthode dans le logiciel Isatis. La mise en œuvre a été décrite dans le papier que nous avons publié (Deraisme et al., 2008).

A cette occasion d'ailleurs, nous avons fini par élucider un point épineux déjà présent dans le CU monovariable. Alors que la technique requiert la variance de la teneur du panneau, cette teneur n'est qu'estimée : dès lors, faut-il utiliser la variance de la teneur vraie, ou celle de la teneur estimée (laquelle peut changer d'un panneau à l'autre, selon la configuration géométrique d'estimation du panneau à partir des données) ? Il apparaît en fait que c'est bien la variance de la teneur estimée qui se justifie, le formalisme pouvant être étendu à ce cas, à supposer que cet estimateur soit sans biais conditionnel : hypothèse qui parait implicite au fait de vouloir conditionner l'estimation de la distribution des blocs justement à cette valeur de panneau (si on récupérait tout les blocs du panneau, c'est bien cette valeur qu'en moyenne on s'attend à obtenir).

Références

Emery, X., and Ortiz, J. M., 2005. Internal consistency and inference of change of support isofactorial models, *in* Geostatistics Banff 2004, Proceedings of the 7th International Geostatistics Congress, O. Leuangthong and C.V. Deutsch (eds.), Springer, Dordrecht, The Netherlands, Vol. 2, p. 1057-1066.

Emery, X., 2009 b. Change-of-support models and computer programs for direct block-support simulation, Computers & Geosciences 35, 2047–2056.

Emery, X., and Ortiz J. M., 2010. Two approaches to direct block-support conditional co-simulation, Computers and Geosciences, in press.

Guibal, D., and Remacre, A. Z., 1984, Local estimation of the recoverable reserves: comparing various methods with the reality on a porphyry copper deposit, *in* Verly, G., David, M., Journel, A. G., and Maréchal, A., eds., Geostatistics for Natural Resources Characterization: Reidel, Dordrecht, 435–448.

Matheron, G., 1976 b, Forecasting block grade distributions: the transfer functions. In Advanced geostatistics in the mining industry, M. Guarascio, M. David, and C. Huijbregts eds. Reidel, Holland, 237-251.

Remacre, A., 1984. L'estimation du récupérable local - Le conditionnement uniforme, Thèse de Doct.-Ing. en Sciences et Techniques Minières, Ecole des Mines de Paris.

Rivoirard, J., 1994 a, Introduction to disjunctive kriging and non-linear geostatistics. Oxford : Clarendon Press, 181 p.

Vann, J., and Guibal, D., 1998. Beyond Ordinary Kriging – An overview of non-linear estimation. Proceedings of Beyond Ordinary Kriging Seminar, Geostatistical Association of Australasia, Perth, Western Australia.

5 Simulations

5.1 Simuler des abondances par les modèles gaussiens

Dans cette section, je parle de travaux de simulation effectués lors des thèses de Jean-Marc Chautru et de Mathieu Woillez, que j'ai dirigées :

Chautru, J.-M., 1989. Modélisation probabiliste de gisements de nodules polymétalliques, Thèse de doctorat en géostatistique, Ecole des Mines de Paris.

Woillez, M., 2007. Contributions géostatistiques à la biologie halieutique, Thèse de Doctorat en Géostatistique, Ecole des Mines de Paris, 173 p.

Woillez, M., Rivoirard, J., and Fernandes, P.G. 2009. Evaluating the uncertainty of abundance estimates from acoustic surveys using geostatistical simulations – ICES Journal of Marine Science, 66:1377–1383.

Les simulations géostatistiques sont un moyen de reproduire un phénomène dans sa variabilité spatiale. Les simulations conditionnelles, de plus, passent par les points expérimentaux. Naturellement une simulation géostatistique suppose un modèle complet de fonction aléatoire, et pas simplement un variogramme. Le modèle classique est le modèle gaussien (après anamorphose), grâce d'une part au théorème central limite pour ce qui concerne les simulations non conditionnelles, et d'autre part à ses facilités de conditionnement (Journel et Huijbregts, 1978 ; Chilès et Delfiner, 1999 ; Lantuéjoul, 2001). L'anamorphose permet, au départ d'inverser les données brutes en valeurs gaussiennes, et à l'arrivée de transformer la simulation de la variable gaussienne en simulation de la variable originale. Alors que le krigeage vise à estimer la valeur d'une variable régionalisée en un point ou sur un domaine, ou que la géostatistique non-linéaire permet le calcul d'intervalles de confiance sur un support déterminé, il est possible d'estimer des quantités plus complexes à partir de simulations conditionnelles.

Dans le domaine minier, notamment, les simulations permettent la prévision de réserves récupérables compte tenu de contraintes d'exploitation : ainsi pour les contraintes de pente accessible aux engins d'exploitation, par exemple, dans la récupération des nodules polymétaliques gisant par 5000m de fond dans le pacifique nord. Dans un premier temps, il s'agissait donc de simuler la distribution des nodules sur le fond marin, couplée à celle de la topographie (thèse de Jean-Marc Chautru (1989), suite à des travaux de Chauvet (1976), Rivoirard (1981), et Lantuéjoul et Rivoirard (1982)) : Fig. 14. Malgré l'existence d'une relation non-linéaire entre concentration en nodules et pente, la simulation bi-variable a été grandement facilitée par l'existence d'un résidu spatialement indépendant de la topographie. L'exploitation de ces simulations montrait ensuite que seul 43% de l'abondance des nodules était récupérable, ce qui correspondait à 34% de la surface du fond marin.

Toujours en mer, mais cette fois-ci pour l'abondance en poissons. Les simulations géostatistiques permettent de chiffrer l'incertitude de cette abondance par classes d'age ou de taille, en combinant par exemple simulations des densités acoustiques de poisson et de la répartition par classes

(exemple du hareng Ecossais déjà cité en section 2.4.1, Thèse de M. Woillez (2007) ; Woillez et al., 2009) : Fig. 15.

Dans le modèle gaussien transformé, les données brutes sont traditionnellement inversées en valeurs dont au moins l'histogramme doit être gaussien. Ceci pose problème lorsque l'anamorphose n'est pas inversible. Ceci se produit notamment lorsqu'il existe un fort effet zéro, ce qui était le cas dans l'exemple précédent (Woillez et al. 2009), où la majorité des valeurs de la densité étaient nulles. Pour résoudre cette difficulté, nous nous sommes inspirés de la méthode utilisée en simulation gaussienne seuillée, où le passage entre gaussienne et gaussienne seuillée n'est pas inversible non plus. Nous avons donc eu recours à un échantillonneur de Gibbs pour commencer par simuler des valeurs gaussiennes à l'endroit de ces valeurs nulles de densité. Naturellement ces premières valeurs gaussiennes simulées doivent obéir à la structure spatiale des gaussiennes du modèle, laquelle ne peut être connue directement, et il était donc nécessaire de procéder au préalable à l'inférence indirecte de celle-ci.

Références

Chauvet, P., 1976, Mission NIXO 20. Note technique, Centre de Morphologie Mathématique, Fontainebleau. Confidentiel.

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Journel, A. G., and Huijbregts, Ch. J. 1978. Mining geostatistics. Academic Press, London. x + 600 p.

Lantuéjoul, C., 2001, Geostatistical Simulation: Models and Algorithms, Springer.

Lantuéjoul, C., et Rivoirard, J., 1982, Etude de la distribution des pentes fortes. Etude CNEXO. 43 p. Confidentiel.

Rivoirard, J., 1981, Précision de l'estimation du récupérable d'un champ de nodules. 52 p. Confidentiel.

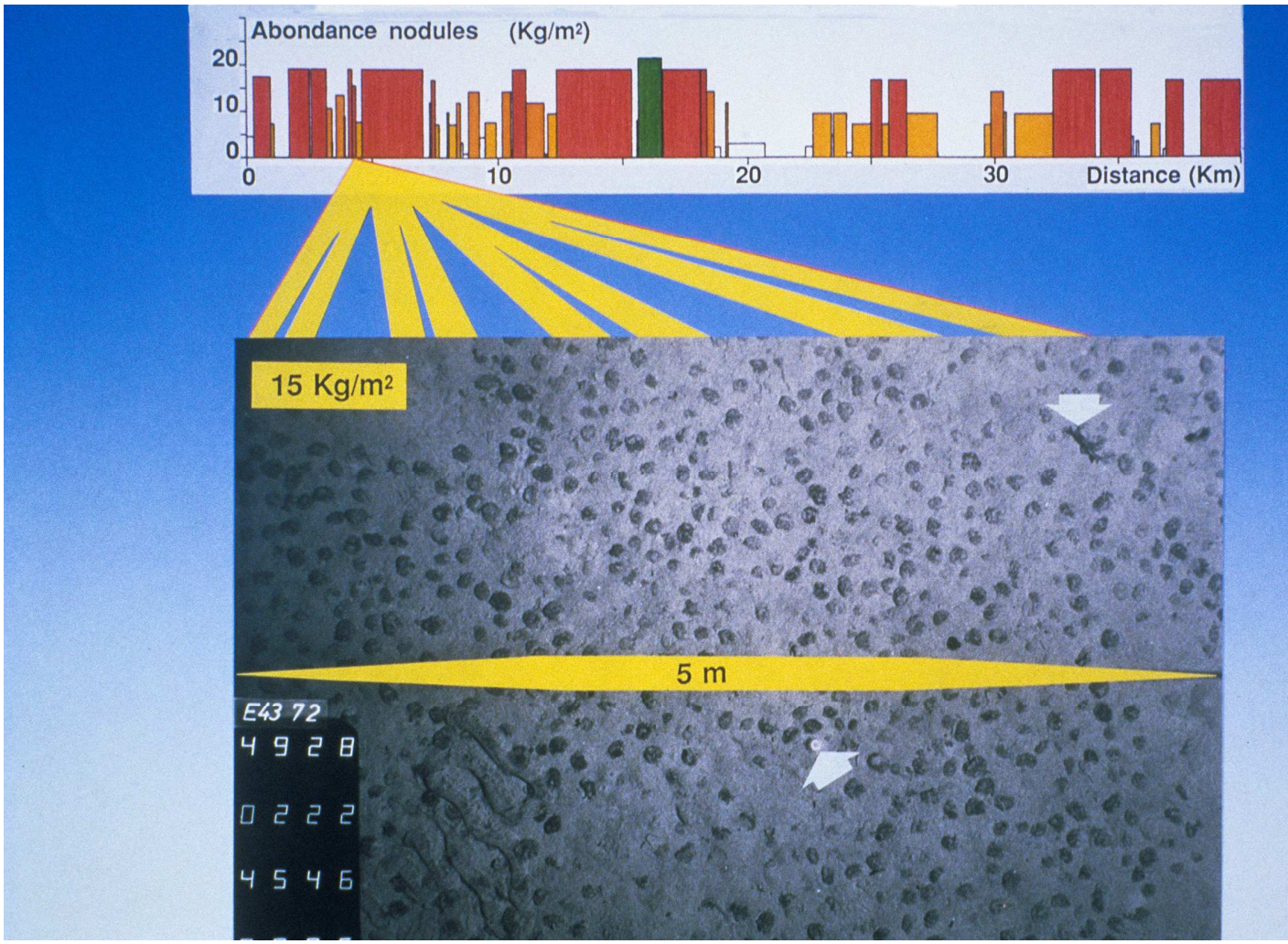


Figure 14-a : Nodules polymétalliques sur le fond du Pacifique Nord. L'échantillonnage se fait à partir de photos et de prélèvement, le long de transects. (Extrait de « Stratégie de dragage/Nodule mining strategy », document audio-visuel, Ifremer-Gemonod, 1988)

Figure 14-b. Nodules polymétalliques sur le fond du Pacifique Nord. Ssimulation géostatistique sur un champ de 22.5 x 50 km : l'abondance augmente du blanc au rouge puis au vert ; les pentes supérieures à 10% sont superposées en noir. (Extrait de « Stratégie de dragage/Nodule mining strategy », document audio-visuel, Ifremer-Gemonod, 1988)

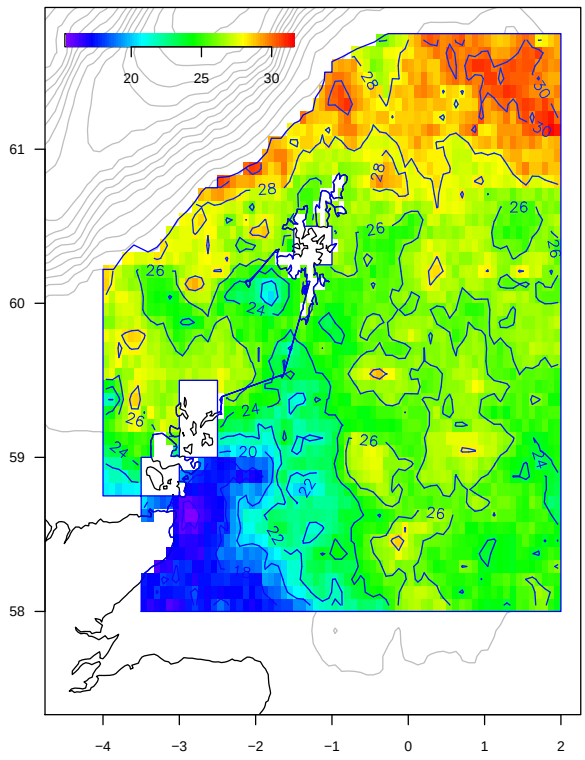


Figure 15. Simulation de la longueur du hareng Ecossais (en cm), année 2003 (Woillez et al., 2009).

5.2 Simuler des corps sédimentaires par gaussiennes seuillées

Dans cette section, je décris quelques développements liés à l'application de la méthode des gaussiennes seuillées sur le champ de gaz de Tunu et décrits dans des rapports d'étude suivants :

Le Loc'h, G., Rivoirard, J. et Groupe Heresim, 1989, Geostatistical study of Tunu, Phase 1. Etude CG-IFP. 36 p. Confidentiel.

Le Loc'h, G., Rivoirard, J. and Groupe Heresim, 1990, Geostatistical study of Tunu, Phases 1 + 2. Etude CG-IFP. 139 p. Confidentiel.

Le Loc'h, G., Rivoirard, J., Renard, D., Ravenne, C., Groupe Heresim, 1992, Geostatistical study of Tunu gas field. Phase III : Non stationary simulations. 17 p. + annexes + fig. Confidentiel.

Rivoirard, J., Renard, D., Ravenne, C. and Heresim Group, 1993, Geostatistical study of Tunu gas field. Phase IV. Confidentiel.

Le problème posé par la simulation des valeurs nulles dans la section précédente est voisin de celui posé par la simulation de faciès par gaussiennes seuillées. Cette méthode a été développée initialement par l'Ecole des Mines et l'IFP (groupe Heresim) pour représenter des réservoirs sédimentaires hétérogènes (Matheron et al., 1988 ; Armstrong et al., 2003). Les indicatrices de faciès y sont considérées comme résultant du seuillage de gaussiennes sous-jacentes. Pour bénéficier des heureuses propriétés des fonctions aléatoires gaussiennes en matière de conditionnement, il est avantageux de procéder à la simulation préalable des gaussiennes à l'endroit des données. Ceci se fait par échantillonneur de Gibbs, tandis que l'inférence de la structure des gaussiennes se fait par inversion de celle des indicatrices de faciès. De tels modèles 3D de réservoirs hétérogènes sont utiles aux géologues pour représenter et visualiser des versions 3D plausibles de leurs réservoirs, comme aux ingénieurs de réservoir pour effectuer des simulations d'écoulement.

Avec d'autres collègues au sein du groupe Heresim, j'ai participé à la réalisation de simulations du champ de Tunu en Indonésie, constitué de multiples dépôts sableux dans un environnement deltaïque, et exploité par Total (Seigneurin et al. 1993). Avec Gaëlle Le Loch en particulier, nous avons procédé à l'analyse structurale et aux premières simulations par gaussiennes seuillées (rapports d'étude des phases I, II et III). Il en résultait des simulations de faciès conditionnées aux données de faciès des puits. Cependant ces simulations n'étaient pas cohérentes au niveau des fluides connus aux puits : ainsi du gaz présent dans un puits pouvait se trouver connecté par du sable à des niveaux d'eau situés plus haut dans d'autres puits. Comme il s'agissait de formations ramifiées, nous avons solutionné le problème, avec Didier Renard (rapports d'étude des phases III et IV), en développant une sorte de simulateur d'écoulement.

Nous partions des données de fluide (eau ou gaz) connues à l'emplacement des sables le long des puits. Puis nous injections, de proche en proche à partir de ces données de sable, soit du gaz qui avait tendance à monter, à partir des données de gaz, soit de l'eau qui avait tendance à descendre, à partir des données d'eau. Ainsi les fluides se propageaient, informant en fluide les cellules de sable de la simulation de faciès, d'abord au voisinage des puits, puis au-delà. A chaque pas de temps, nous tirions au sort une cellule parmi les cellules de sable périphériques aux cellules déjà informées, en privilégiant celles possédant le plus de cellules informées voisines en position horizontale, ou en position haute ou basse selon le fluide. Cette cellule de sable se voyait alors informée par le fluide de ses voisines. Ainsi chaque fluide envahissait rapidement les corps sableux les plus gros qu'il rencontrait (l'eau commençant par les parties inférieures des corps, le gaz par les parties supérieures), et ralentissant énormément dans les parties de sable les plus ramifiées où finissaient par se concentrer les incohérences (eau au-dessus du gaz). Il suffisait alors de remplacer le sable par un bouchon d'argile à l'endroit de ces contacts pour effacer les incohérences. Vus les volumes infimes qui étaient concernés, ceci ne changeait pratiquement pas les proportions de faciès simulés. Bien entendu les composantes connexes sableuses non connectées aux puits étaient informées directement en fluide, selon la probabilité locale de trouver de l'eau ou du gaz. In fine nous obtenions des simulations conditionnelles à la fois en faciès et en fluides.

Références

Armstrong, M., Galli, A.G., Loch, G., Eschard, R., Geffroy, F., 2003, Plurigaussian Simulations in Geosciences, Springer. 149 p.

Matheron, G., Beucher, H., de Fouquet, C., Galli, A. et Ravenne, C., 1987, Simulation conditionnelle à trois faciès dans une falaise de la formation du Brent. Sciences de la Terre, Série Informatique Géologique, 28, 213-249.

Seigneurin, A., Muller, D., Galli, A., and Ravenne, C., 1993, Optimization of the well-spacing with a geostatistical model Tunu field - Mahakam area. In : Proc. Indonesian Petroleum Association, Twenty Second Annual Convention, Jakarta, Indonesia.

5.3 Les corps sédimentaires comme objets de simulations booléennes

Je me suis intéressé au modèle booléen, à l'occasion du Consortium Booléen (1993-1994) regroupant Ecole des Mines, IFP, GDF, Elf, et BP. Les résultats que nous avons obtenus alors avec ma collègue Hélène Beucher ont été consignés dans deux rapports d'étude :

Rivoirard, J., et Beucher, H., 1993, Validité du modèle booléen et tests sur simulations. 44 p. Confidentiel.

Rivoirard, J., 1994 b, Validité du modèle booléen et tests sur simulations (Second rapport). 50p. Confidentiel.

Bien que modélisant de manière indirecte la géométrie et l'arrangement des faciès, le modèle gaussien (ou pluri-gaussien) seuillé mentionné précédemment est capable de restituer des formes d'un réalisme étonnant. De plus il se prête bien aux simulations conditionnelles. Naturellement, lorsqu'on veut simuler un ensemble de corps ou objets, on peut chercher à le faire directement, pourvu que l'on sache simuler au préalable ces objets individuellement. Dans le modèle booléen classique (Matheron, 1967), les objets individuels, eux-mêmes aléatoires et de même loi, sont implantés au hasard (selon un processus de Poisson de densité constante) et indépendamment. Le modèle booléen lui-même est l'ensemble aléatoire constitué par la réunion des objets. Ce modèle a retenu l'attention des pétroliers, depuis les travaux précurseurs de Haldorsen (Haldorsen, 1983 ; Haldorsen and Lake, 1984), qui proposait de conditionner la distribution de lentilles gréseuses (schématisées par des objets parallélépidédiques de taille aléatoire) aux traversées de sable observées le long de puits verticaux espacés. En particulier Chessa (1995) et Berkhout et al. (1996) ont amélioré la méthode en remarquant qu'une lentille avait d'autant plus de chances d'être recontrée par un puits que sa taille était grande, et donc que la distribution de probabilité de la taille des lentilles rencontrées par un puits ne devait pas coïncider avec leur distribution a priori.

Les développements théoriques fondamentaux du schéma booléen sont dus à Matheron (1967), et ont été replacés dans la perspective plus générale des ensemble aléatoires par Matheron (1975). Un des gros intérêts du modèle booléen réside dans ses possibilités de calcul (lesquelles proviennent du fait qu'il s'agit d'un ensemble indéfiniment divisible pour la réunion : comme les objets sont indépendants, tout schéma booléen peut être considéré comme la réunion de schémas booléens de densité moindre). En voici quelques éléments utiles dans la suite.

La capacité de Choquet du modèle booléen (analogue de la fonction de répartition d'une variable aléatoire) prend une forme très simple (non détaillée ici). Par exemple un point de l'espace n'appartiendra pas au modèle booléen, s'il n'est recouvert par aucun objet. La probabilité correspondante est celle avec laquelle le nombre de points poissonniens tombant dans le volume correspondant à un objet est égal à 0. Cette probabilité vaut donc $\exp(-\theta E(K))$, où θ est la densité d'implantation des objets et $E(K)$ l'espérance du volume d'un objet générique K. Elle représente en moyenne la proportion de volume de l'espace n'appartenant pas aux objets.

De même le variogramme de l'ensemble booléen s'exprime (par une formule similaire) à l'aide de la densité et du covariogramme géométrique des objets. Du fait que le covariogramme ne peut prendre de valeurs négatives, on peut montrer que le variogramme du modèle booléen ne peut pas dépasser la variance et ne peut donc pas présenter d'effets de trou au niveau du palier.

Lorsqu'on chemine le long d'une droite (pouvant représenter un puits pétrolier), se trouver hors des objets (mettons des lentilles de sable) n'augmente en rien la probabilité de rencontrer du sable plus loin, puisque ces objets sont distribués aléatoirement et indépendamment. Aussi la loi de probabilité des traversées linéaires entre objets (disons L) a un taux de « survie » au-delà d'une longueur h, $P(L-h > dh \mid L>h)$, qui ne peut diminuer quand h augmente. Ce taux est soit constant en h (les traversées sont alors de loi exponentielle, et c'est le cas d'objets convexes : lorsqu'on en quitte un, c'est pour toujours), ou augmente avec h (en particulier à petite distance, à cause de la possibilité de rencontrer à nouveau le même objet, s'il n'est pas convexe).

Dans le Consortium Booléen, avec ma collègue Hélène Beucher, nous nous sommes intéressés à la manière de valider le modèle booléen en contexte pétrolier (Rivoirard et Beucher 1993, Rivoirard 1994 b). Nous avons fait ce travail en utilisant des simulations variées : simulations booléennes (avec objets connexes ou non), simulations d'objets non booléennes (modèles d'objets émincés, évitant la proximité entre objets), et simulations gaussiennes seuillées. Nous avons fait choix de retenir, comme outils de validation, le calcul de statistiques à une dimension, telles que l'on peut effectuer le long de puits pétroliers espacés : absence d'effet de trou au niveau du palier du variogramme ; survie des traversées entre objets (ou plus exactement des traversées de l'ensemble complémentaire, cette statistique peut d'ailleurs servir à déterminer si c'est l'ensemble ou son complémentaire qui peut être vu comme un schéma booléen d'objets) ; et d'autres statistiques non détaillées ici (comme l'inégalité triangulaire renforcée du variogramme).

Les résultats essentiels sont les suivants. Déjà, la présence d'effet de trou au niveau du palier du variogramme (témoignant par exemple de l'absence de proximité entre objets) est de fait un excellent moyen de réfuter le modèle booléen. De façon plus fine sinon, un « taux de survie » des traversées hors objets qui diminue invalide assez nettement le modèle booléen. En pratique, la longueur des traversées mesurables étant finie, c'est aux petites distances que ceci doit être observé : un taux qui commence par décroitre indique un déficit de traversées courtes par rapport au modèle booléen, cas par exemple d'objets qui auraient tendance à s'éviter au lieu d'être implantés indépendamment.

Ainsi traversées entre objets et variogramme sans effet de trou au niveau du palier apparaissent-ils comme de bons critères de validation, du moins tant qu'on reste dans un cadre stationnaire. Cependant, dans la réalité, la distribution des lentilles de sable n'est le plus souvent pas homogène, en particulier dans la direction verticale, et il était intéressant de voir aussi comment valider ou invalider un modèle booléen non stationnaire (Rivoirard, 1994 b). Ceci est apparu

faisable dans le cas d'objets homogènes et d'une densité de Poisson monotone dans la direction verticale (ainsi le « taux de survie » observé ne doit pas diminuer dans la direction de décroissance de la densité), mais beaucoup plus ardu autrement. Au final le modèle booléen sensu stricto apparaît exigeant dans le cas stationnaire (l'hypothèse d'objets implantés aléatoirement et indépendamment est forte), mais l'introduction de non-stationnarités en fait un modèle beaucoup plus général et difficile à réfuter.

L'inférence de modèles booléens non stationnaires a fait plus tard l'objet de la thèse de Maria Benito Garcia-Morales (2003), encadrée par Hélène Beucher. Toutes choses égales par ailleurs, les objets sont plus fréquents là où la densité (non stationnaire) est plus élevée, ce qui se traduit par une proportion accrue du faciès correspondant. Inversement, on peut faire l'inférence de la densité non stationnaire en déconvoluant la carte des proportions estimées.

Références

Benito Garcia-Morales, M., 2003. Non stationnarité dans les modèles de type booléen : application à la simulation d'unités sédimentaires, Thèse de Docteur en Géostatistique, Ecole des Mines de Paris.

Berkhout, R. J., Chessa, A. G., and Martinius, A. W., 1996, A statistical adjustment of Haldorsen's conditioned Boolean simulation algorithm, Math Geol, Vol 28, No 6, 791-810.

Chessa, A. G., 1995, Conditional simulation of spatial stochastic models for reservoir heterogeneity, PhD TU Delft.

Haldorsen, H. H., 1983, Reservoir Characterization Procedures for Numerical Simulation, Ph.D. Dissertation, The University of Texas at Austin.

Haldorsen, H. H., Lake, L.W., 1984, A new approach to shale management in field-scale models, SPE Jour., v. 24, no. 4, 447-457.

Matheron, G., 1967, Éléments pour une théorie des milieux poreux, Masson, Paris.

Matheron, G., 1975, Random sets and integral geometry, Wiley, New-York.

5.4 Modèles génétiques de réservoirs chenalisés méandriformes

Dans cette section, je parle des modèles génétiques aléatoires de réservoirs chenalisés méandriformes développés à l'Ecole des Mines, dans le cadre de la thèse de Simon Lopez (2003) que j'ai encadrée avec Isabelle Cojan, ainsi que des consortiums pétroliers MCRC (Meandering Channelized Reservoirs Consortium, 2002-2004), puis FLUMY (Fluvial meandering channelized reservoirs, a process-based stochastic modelling, depuis 2005) dirigés par Isabelle Cojan et moi-même :

Cojan, I., Fouché, O., Lopez, S., and Rivoirard, J., 2005, Process-based reservoir modelling in the example of meandering channel. In: O. Leuangthong and C.V. Deutch (eds.), *Geostatistics Banff 2004*, Dordrecht: Springer, 2005. p. 611-619.

Lopez, S., 2003, Modélisation de reservoirs chenalisés méandriformes, approche génétique et stochastique, Ph. D., Ecole des Mines de Paris.

Lopez S., Cojan I., Rivoirard J., and Galli A., 2008. Process-based stochastic modelling: meandering channelized reservoirs. Spec. Publ. Int. Assoc. Sedimentol. 40 - 139:144.

Rivoirard J., Cojan I., and Geffroy F., 2007, Controlling the location of deposits in meandering channelized reservoir models. In: Extended Abstracts Book, Petroleum Geostatistics 2007, Cascais, Portugal, European Association of Geoscientists & Engineers, Houten, The Netherlands, P19. 5 p.

Rivoirard J., Cojan I., Renard D. and Geffroy F., 2008, Advances in quantification of process-based models for meandering channelized reservoirs, VIII International Geostatistics Congress, GEOSTATS 2008, 607-616. (Voir ANNEXE.)

A la différence des modèles booléens, dans lequels il n'existe pas de processus sédimentaire de dépôt des objets, les modèles aléatoires *génétiques* (« genetic » ou « process-based » en anglais) visent à simuler les dépôts en reproduisant les processus majeurs de la sédimentation. Apparus très tôt (voir par exemple : Jacod et Joathon, 1971), ces modèles ont connu beaucoup de développements depuis une quinzaine d'années, profitant de la puissance de calcul des ordinateurs actuels et comblant l'insuffisance des modèles de simulations purement géostatistiques. Par exemple Teles et al. (1998) et Teles (1999) simulent les dépôts d'une plaine alluviale en respectant les lois empiriques de la migration des méandres de la rivière.

Il existe en effet des géométries, des agencements, qu'il est très difficile de simuler, sauf à essayer de mimer la nature, autrement dit de simuler les processus à l'oeuvre. C'est en particulier le cas des réservoirs constitués par les barres de méandres et autres sédiments d'un système fluviatile chenalisé. A la suite d'échanges entre sédimentologue (Isabelle Cojan) et géostatisticien (Alain Galli) de l'Ecole des Mines de Paris, des modèles génétiques aléatoires de réservoirs chenalisés méandriformes ont été développés dans le cadre des consortiums pétroliers MCRC et FLUMY que j'ai dirigés avec Isabelle Cojan (thèse de Simon Lopez en 2003, Cojan et al. 2005, Lopez et al. 2008). La figure 16 montre un exemple d'une telle modélisation. A noter que des modèles comparables ont été développés à la même époque à l'Université d'Alberta (Pyrcz, 2004 ; Pyrcz et al., 2009). Ces modèles sont basés sur des processus physiques régis par des équations apparues dans la littérature spécialisée dans les années 80 pour ce qui concerne l'évolution des méandres et donc la migration (notamment Ikeda et al., 1981). Naturellement, des approximations sont présentes, liées en particulier à l'échelle de travail (celle des réservoirs, on simule des champs de plusieurs km de coté, et non une barre de méandre dans tous ses détails) et au fait que ce qu'on cherche à simuler est essentiellement l'architecture des dépôts qui vont subsister. Trois processus sédimentaires majeurs sont à l'œuvre. D'abord effectivement la migration du chenal, mouvement essentiellement latéral, accompagnée des dépôts sableux de barres de méandre. Ensuite l'agradation du système lors des crues de débordement, correspondant aux dépôts sableux de fond de chenal, aux levées silteuses et aux dépôts alluviaux argileux d'épaisseur allant décroissante plus loin sur la plaine ; du fait de l'aggradation, le chenal a tendance à se hausser au-dessus de sa plaine, protégé par ses levées. Et enfin les avulsions, lorsqu'à l'occasion de ruptures de levées le chenal va chercher un chemin plus bas dans la plaine,

où l'agradation se fera alors, équilibrant la topographie du système. L'extension latérale des méandres est limitée par l'occurrence de cutoffs, définissant la ceinture de chenaux. Dans un modèle tel que FLUMY, un changement très léger du chenal initial provoque, par le jeu des cutoffs, des évolutions de même allure mais différentes, ce qui correspond à une situation pré-aléatoire. Par ailleurs les occurrences de crues de débordement et d'avulsions obéissent à une chronologie périodique ou aléatoire.

Alors qu'en réalité la succession des dépôts se fait selon l'état initial et les conditions courantes, la situation est différente si l'on cherche à modéliser un champ. En particulier une simulation conditionnelle va être conditionnée par les données dont on dispose (ex : données de puits), autrement dit des données qui représentent le futur par rapport à la simulation temporelle de la sédimentation. Il apparaît assez naturel de construire des simulations conditionnelles en essayant de reproduire les conditions locales de dépôt telles qu'indiquées par les données : ainsi de s'arranger pour faire passer de préférence le chenal là où on l'a observé, et où on veut donc déposer du sable, même si le procédé n'est pas efficace à 100%. Lopez (2003) a imaginé de réduire l'érodabilité dans de tels endroits, de façon à y attirer le chenal migrant, faisant jouer à ce paramètre un rôle travesti mais intéressant. A contrario, une érodabilité plus forte préservera de l'argile déjà déposée là où on veut. Lors d'une avulsion, il convient de même de générer un nouveau tracé de chenal qui passe au voisinage de puits où, à ce moment-là, on veut déposer du sable, et d'éviter les dépôts d'argile à préserver. De façon à éviter une combinatoire compliquée entre puits « attractifs » pour l'avulsion, et puits « répulsifs », nous avons imaginé de laisser le nouveau chenal choisir de lui-même où aller, en imprimant à la topographie des sortes de gouttières-entonnoirs fictifs conduisant aux puits à sable, et de barrières de protection pour les puits à argile (Rivoirard et al. 2007).

Au contraire des gaussiennes seuillées et de leur conditionnement aux faciès, ces modèles génétiques stochastiques et ce type de conditionnement ne relèvent pas à proprement parler de la géostatistique (le conditionnement des fluides précédemment évoqué en section 5.2 non plus, d'ailleurs). En particulier, les simulations dites conditionnelles ne coïncident vraisemblablement pas à ce qu'elles devraient être en théorie, autrement dit, des simulations non conditionnelles qui passeraient par les données.

De façon assez détournée, cependant, la géostatistique s'avère utile dans ces modèles de réservoirs chenalisés. Pour être opérationnel, un modèle comme FLUMY doit reposer sur un nombre limité de paramètres clés, disons une dizaine (les autres étant moins importants). Or ces paramètres, qui permettent la construction du modèle, ne sont pas directement accessibles à partir des données disponibles, exception faite de certains (profondeur du chenal, par exemple). Il n'est donc pas aisé de faire le choix de l'ensemble des paramètres clés, même si l'on sait que, par exemple, des crues de débordement plus intenses ou plus fréquentes vont favoriser l'agradation et donc au final diminuer le net/gross (proportion de sable). De façon à choisir des ordres de grandeur appropriés pour les paramètres, nous avons fait un usage heuristique des formules du modèle booléen. Nous avons plus précisément considéré qu'entre deux avulsions, les corps sableux constitués par les points de l'espace balayés par le chenal pouvaient être considérés comme les objets d'un schéma booléen (en section transverse du champ). La fraction de l'espace complémentaire au net/gross s'exprime alors comme :

$$1 - \text{proportion de sable} = \exp\left[-\theta K\right]$$

où K représente la section moyenne transverse d'un objet, et θ la densité de ces objets (en nombre d'objets par unité de surface transverse). K augmente essentiellement avec la section du chenal, la vitesse de migration et la période d'avulsions ; θ avec la fréquence d'avulsions et la période des crues de débordement. En reliant ainsi ces paramètres K et θ aux paramètres d'entrée du modèle et à des paramètres dérivés, nous avons constaté que la formule booléenne était effectivement capable de fournir un ordre de grandeur intéressant pour le N/G (Rivoirard et al. 2008). En exploitant ces différents types de relations, il a même été possible de proposer à l'utilisateur non-expert de FLUMY un pré-calculateur, informant l'ensemble du modèle à partir des trois paramètres que sont la profondeur du chenal, l'extension latérale des corps sableux, et le N/G.

Références

Ikeda S., Parker G., and Sawai K. 1981. Bend theory of river meanders. Part 1. Linear development, *Journal of Fluid Mechanics*, vol. 112, pp. 363-377.

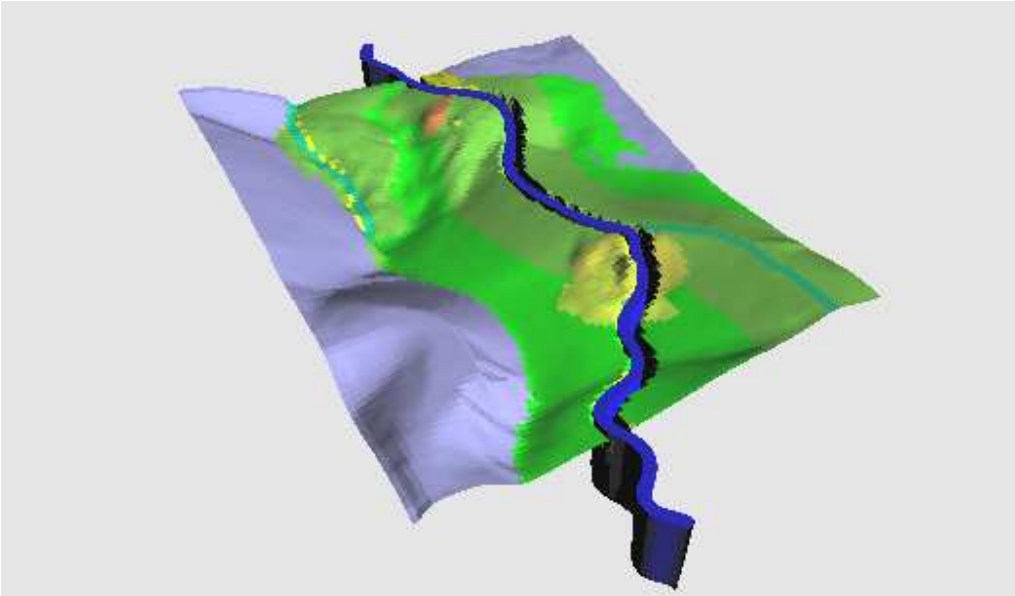
Jacod J. et Joathon P., 1971, Use of random-genetic models in the study of sedimentary processes, Math. Geol., vol. 3, no. 3, 265-279.

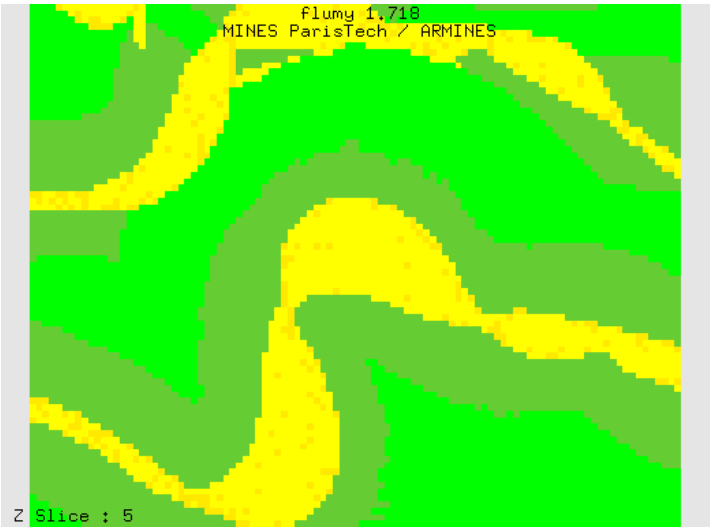
Pyrcz, M.J., 2004. Integration of geologic information into geostatistical models. Ph.D. Dissertation, University of Alberta, Edmonton, Canada, 296 p.

Pyrcz, M.J., Boisvert, J. B., and Deutsch, C.V., 2009, Alluvsim: a Conditional Event-based Fluvial Model, Computers & Geosciences, vol. 35, no8, 1671-1685.

Teles, V., 1999, Construction de Réservoirs Aquifères Alluviaux par Modèle Génétique de mise en place des Alluvions. PhD Thesis, University Pierre et Marie Curie Paris VI, Mémoire des Sciences de la Terre, 99-03.

Teles, V., Marsily, G. de, Perrier, E., 1998, Sur une nouvelle approche de modélisation de la mise en place des sédiments dans une plaine alluviale pour en représenter l'hétérogénéité. CR Acad. Sci. Paris, 327 , 597–606.





Figures 16. Modélisation génétique de réservoirs chenalisés méandriformes (modèle FLUMY développé à l'Ecole des Mines). Haut : le chenal migre en déposant des barres de méandres ; agrade en construisant les levées (vert foncé) et en déposant des fines sur la plaine (vert clair) ; et à l'occasion de ruptures de levée, fait des dépôts de crevasse (jaune, marron) et peut choisir un trajet différent (avulsion). Echelle verticale exagérée. Bas : exemple de section horizontale montrant (en jaune) le développement des barres de méandre.

6 Perspectives

Dans la continuité de mes travaux de recherche, je vois plusieurs secteurs fondamentaux dans lesquels la méthodologie géostatistique pourrait encore s'améliorer.

6.1 Outils structuraux plus souples, plus généraux

Dans la pratique géostatistique, un des aspects les plus excitants, importants, et critiques est l'analyse structurale. Excitant, parce que c'est à cette étape que l'on peut dévoiler la structure ou les liens structuraux des variables. Important, parce que c'est de leur modélisation que dépend la suite des opérations. Mais critique, parce qu'il est souvent difficile de mettre en évidence les structures. A la question : « comment dégager au mieux les structures ? », la réponse est encore loin d'être simple.

La présence de clusters (paquets de données regroupées géographiquement) dans l'échantillonnage, par exemple dans des zones plus riches et plus variables, semble nécessiter une transformation de la variable pour rendre la variabilité à petite distance compatible avec la variabilité à plus grande distance. Des variations locales de la structure (ex : directions principales d'anisotropie) rendent insuffisant le calcul de variogrammes moyens et suggèrent le développement de modèles globaux, non stationnaires et cohérents (Paciorek and Schervish, 2006 ; Adams and Stegle, 2008 ; Machuca-Mory, 2010 ; Magneron et al., 2010). Une meilleure exploitation des continuités spatiales peut aussi être obtenue, en multivariable comme en non-linéaire, par combinaison des variables continues ou catégorielles.

La dépendance à d'autres variables, en particulier la dépendance causale à des variables mieux connues, suggère l'utilisation de régressions statistiques multivariables (exemple des Generalized Additive Models populaires en biologie marine, pour des variables explicatives environnementales : Hastie and Tibshirani (1990), Denis, Lejeune, and Robin (2002), Trenkel (2006)…), mais dans lesquels la notion de support chère à la géostatistique peut jouer un rôle important. Par ailleurs des modèles plus économiques en hypothèses pourraient être intéressants : ainsi, quitte à se contenter d'estimateurs faiblement biaisés, travailler sur des écarts quadratiques plutôt que des variances quand se posent des problèmes de séparation entre dérive et résidu (Rivoirard et Guiblin, 1997).

L'inférence de modèles spécifiques ayant fait leur preuve mérite plus de recherche. Je pense d'abord au modèle gaussien seuillé, dans sa version non-stationnaire où le seuil (équivalent à la proportion de faciès) varie dans l'espace. Comment estimer ces proportions à partir des données d'indicatrices de faciès? Comment distinguer, au niveau de ces données, les variations qui relèvent de la gaussienne du modèle, de celles liées à la proportion ? Quelles conséquences ont ces deux sources de variabilité concurrentes sur l'objectivité du modèle ? Je pense ensuite au modèle spécifiquement pluri-gaussien : comment ajuster au mieux le domaine des différents faciès dans le « drapeau » (à 2D) dont les axes sont les deux gaussiennes ?

Enfin la spécification de la validité du modèle en terme de distance m'apparaîtrait comme un usage souhaitable. La stationnarité (c'est-à-dire l'invariance par translation) est ce qui permet l'inférence du modèle à partir d'une réalisation unique, mais jusqu'où est valable la stationnarité (de la variable, de ses accroissements, etc) dans le modèle ajusté: est-ce à l'échelle du champ, ou

est-elle seulement locale ? Et jusqu'à quelle distance environ, dans ce cas ? Point qui nous amène au point suivant : le choix des voisinages en krigeage et cokrigeage.

6.2 Voisinage de krigeage et cokrigeage

Le choix d'un voisinage glissant est justifié, soit par une hypothèse de stationnarité locale, le modèle de variogramme par exemple ayant une validité limitée en distance, soit par un nombre important de données. Comment alors réduire le nombre de données ? On peut utiliser des critères de proximité, intégrant éventuellement des notions de directions (octants) ; regrouper certaines données, par exemple les plus lointaines ; on peut aussi chercher à sélectionner les données les plus importantes vis-à-vis de la précision (Emery, 2009 a).

Le krigeage pénalisant offre la possibilité d'obtenir un krigeage continu avec un voisinage limité, et devrait bien se prêter à des problèmes de raccordements, ou de mise à jour de carte. Il serait naturellement intéressant de l'adapter à des voisinages plus élaborés, à base d'octants par exemple. Mais « faire sortir » les données des octants avec un poids nul est peut-être maladroit, et sans doute faut-il développer des voisinages privilégiant des recherches directionnelles et mieux adaptés à une pénalisation. Le krigeage pénalisant est aussi un moyen souple de traiter de grands jeux de données. Dans le cas de cokrigeage, le recours à des modèles à effets d'écran me semble un moyen de réduire les données en limitant la perte d'information. Une situation assez difficile est celle de l'analyse krigeante, où l'on doit traiter différentes échelles en même temps, dans le but précisément de pouvoir les extraire individuellement. C'est sur ces questions que je souhaite travailler.

6.3 Krigeage Disjonctif empirique

La géostatistique non-linéaire a été essentiellement développée dans les années 70 et 80, alors en vue de la prévision des réserves récupérables minières (qu'il conviendrait maintenant, à la suite du code minier australien, le JORC, de rebaptiser en ressources récupérables). L'extraordinaire moisson de résultats n'a été jusque là que peu exploitée, hors le modèle gaussien, rafraîchi ces dernières années par Xavier Emery (notamment Emery et Ortiz (2005)). On dispose, dans des cartons, de nombre de modèles basés sur des lois gamma, binomiale négative, Poisson, etc, avec leur changement de support. Ces modèles permettent l'estimation par Krigeage Disjonctif. D'un point de vue théorique, le Krigeage Disjonctif est identique au cokrigeage d'indicatrices. L'approche directe par indicatrices, extrêmement populaire à une époque, semble moins en faveur aujourd'hui (Journel, 1983, Vann and Guibal, 1998). Cela me semble dommage, vu l'intérêt que présentent les indicatrices lorsque des objets géométriques sont en jeu. La raison en est due en partie, je pense, au manque de considération des liens entre indicatrices dans l'approche par indicatrices. De ce point de vue, les modèles à résidu d'indicatrice et d'écrêtage font exception. Le cokrigeage d'indicatrices me semble mériter davantage de recherche. Ce pourrait être sur des modèles directs de structures d'indicatrices (Pardo-Igúzquiza et Dowd, 2005). Ou sur des méthodes de factorisation, notamment les MAF : les Min-max Autocorrelation Factors sont à la géostatistique ce que l'analyse en composantes principales est à la statistique. Ils permettent de représenter un ensemble de variables par des facteurs, combinaisons linéaires non corrélées entre elles pour la distance h=0 ainsi que pour une autre distance privilégiée, par exemple celle de la maille d'échantillonnage, et ces combinaisons sont classées selon la valeur de leur autocorrélation à cette distance. Sont naturellement les plus importants pour l'estimation locale les facteurs qui ont l'autocorrélation la meilleure, et les moins importants, ceux qui sont

purement pépitiques (et ceci, quelle que soit la part de variabilité expliquée par ces facteurs, à la différence des Composantes Principales). Appliqués aux indicatrices, ces facteurs pourraient orienter vers le choix de modèles connus (par exemple dans un modèle gaussien, le premier facteur est la gaussienne, dans un modèle à résidus d'indicatrices, c'est l'indicatrice de base). Mais ils pourraient aussi conduire à des modèles originaux, généralisant les modèles gaussiens discrets développés par Matheron (1984) et Lajaunie et Lantuéjoul (1989), basés sur des processus de diffusion. En somme ces MAF d'indicatrices permettraient une sorte de Krigeage Disjonctif empirique. La recherche devra considérer celle de modèles cohérents du point de vue des lois de probabilité impliquées, les questions de continuité et de discrétisation, le changement de support, et l'adaptation à une stationnarité locale. Naturellement la recherche gagnera à être étendue aux cas multivariables.

Références

Adams, R., and Stegle, O., 2008. Gaussian process product models for nonparametric nonstationarity, Proceedings of the 25th International Conference on Machine Learning, ACM Press, 1-8.

Denis, V., Lejeune, J., and Robin, J. P. 2002. Spatio-temporal analysis of commercial trawler data using General Additive models: patterns of Loliginid squid abundance in the north-east Atlantic. ICES Journal of Marine Science, 59: 633–648.

Emery, X., and Ortiz J. M., 2005. Internal consistency and inference of change of support isofactorial models, *in* Geostatistics Banff 2004, Proceedings of the 7th International Geostatistics Congress, O. Leuangthong and C.V. Deutsch (eds.), Springer, Dordrecht, The Netherlands, Vol. 2, p. 1057-1066.

Emery, X., 2009 a. The kriging update equations and their application to the selection of neighboring data, Computational Geosciences, vol. 13, no3, 269-28.

Hastie, T. J., and Tibshirani, R. J., 1990. Generalized Additive Models. Chapman & Hall/CRC.

Journel, A.G., 1983: Nonparametric estimation of spatial distributions, Mathematical Geology, v.15, no.3, p.445-468.

Lajaunie, C., et Lantuéjoul, C., 1989. Setting up the general methodology for discrete isofactorial models. *In* Geostatistics, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 323-334.

Machuca-Mory, D. F., 2010. Geostatistics with Location-Dependent Statistics, PhD in geostatistics, University of Alberta.

Magneron, C., Jeannée, N., Le Moine, O., and Bourillet, J.-F., 2010. Integrating Prior Knowledge and Locally Varying Parameters with Moving-GeoStatistics: Methodology and Application to Bathymetric Mapping, GeoENV VII, Quantitative Geology and Geostatistics, Volume 16, 405-415.

Matheron, G., 1984. Une méthodologie générale pour les modèles isofactoriels discrets. Science de la Terre, Série Informatique Géologique, 21, 1-21.

Paciorek, C. J., and Schervish, M. J., 2006, Spatial modelling using a new class of nonstationary covariance functions, Environmetrics. 2006; 17(5): 483–506.

Pardo-Igúzquiza, E. and Dowd, P.A., 2005. On the automatic inference and modelling of a set of indicator covariances and cross-covariances. In: O. Leuangthong and C.V. Deutch (eds.), *Geostatistics Banff 2004*, Dordrecht: Springer, p. 185-194.

Rivoirard, J., and Guiblin, Ph., 1997. Global estimation variance in presence of conditioning parameters. In Baafi and Schofield Eds, Geostatistics Wollongong'96, Kluwer Academic Publisher, Vol.1, 246-257.

Trenkel, V., 2006. Méthodes d'évaluation des ressources marines vivantes pour une exploitation durable. HDR. Université Montpellier 2.

Vann, J., and Guibal, D., 1998. Beyond Ordinary Kriging – An overview of non-linear estimation. Proceedings of Beyond Ordinary Kriging Seminar, Geostatistical Association of Australasia, Perth, Western Australia.

7 Références

Adams, R., and Stegle, O., 2008. Gaussian process product models for nonparametric nonstationarity, Proceedings of the 25th International Conference on Machine Learning, ACM Press, 1-8.

Armstrong, M., and Delfiner, P., 1980, Towards a more robust variogram: A case study on coal, Technical report, Centre de Géostatisque, Fontainebleau.

Armstrong, M., Galli, A.G., Loch, G., Eschard, R., Geffroy, F., 2003, Plurigaussian Simulations in Geosciences, Springer. 149 p.

Benito Garcia-Morales, M., 2003. Non stationnarité dans les modèles de type booléen : application à la simulation d'unités sédimentaires, Thèse de Docteur en Géostatistique, Ecole des Mines de Paris.

Berkhout, R. J., Chessa, A. G., and Martinius, A. W., 1996, A statistical adjustment of Haldorsen's conditioned Boolean simulation algorithm, Math Geol, Vol 28, No 6, 791-810.

Beucher H., Galli A., Rivoirard J., 2001, Pravap 10 : Geostatistical analysis of the 4D seismic. Rapport technique, Centre de Géostatistique, Ecole des Mines de Paris, 56 p. Confidentiel.

Bez, N., 1997, Statistiques individuelles et géostatistique transitive en écologie halieutique, Thèse de Docteur en Géostatistique, Ecole des Mines de Paris, 280 p.

Bez, N., 2002. Global fish abundance estimation from regular sampling: the geostatistical transitive method. Canadian Journal of Fisheries and Aquatic Sciences, 59 : 1921-1931.

Bez, N., and Rivoirard, J., 2000. On the role of sea surface temperature on the spatial distribution of early stages of mackerel using inertiograms, in ICES Journal of Marine Science, 57:383-392.

Bez, N., and Rivoirard, J., 2001. Transitive geostatistics to characterise spatial aggregations with diffuse limits: an application on mackerel ichtyoplankton. Fisheries research, 50, 41-58.

Bez, N., Rivoirard, J., Guiblin, Ph., and Walsh, M., 1997. Covariogram and related tools for structural analysis of fish survey data. In Baafi and Schofield Eds, Geostatistics Wollongong'96, Kluwer Academic Publisher, Vol.2, 1316-1327.

Bordessoule, J.-L., Demange, C., et Rivoirard, J., 1988, Estimation de réserves géologiques en uranium par variables utiles à résidu autokrigeable. *Sciences de la Terre*, série Informatique géologique, *28*, pp 27-51.

Bordessoule, J.-L., Demange, C., and Rivoirard, J., 1989, Using an orthogonal residual between ore and metal to estimate in-situ uranium resources. In M. ARMSTRONG (ed.), *Geostatistics*, Vol. 2. Dordrecht : Kluwer, 1989, 923-934.

Bouleau, M., 2005, Combinaison géostatistique de l'acoustique et des captures dans les campagnes scientifiques de pêche par chalutage. Thèse de doctorat en géostatistique, Ecole des Mines de Paris, 189 p.

Chautru, J.-M., 1989. Modélisation probabiliste de gisements de nodules polymétalliques, Thèse de doctorat en géostatistique, Ecole des Mines de Paris.

Chauvet, P., 1976, Mission NIXO 20. Note technique, Centre de Morphologie Mathématique, Fontainebleau. Confidentiel.

Chessa, A. G., 1995, Conditional simulation of spatial stochastic models for reservoir heterogeneity, PhD TU Delft.

Chilès, J.-P., and Delfiner, P., 1999, Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.

Cojan, I., Fouché, O., Lopez, S., and Rivoirard, J., 2005, Process-based reservoir modelling in the example of meandering channel. In: O. Leuangthong and C.V. Deutch (eds.), *Geostatistics Banff 2004*, Dordrecht: Springer, 2005. p. 611-619.

Coléou, T., Hoeber, H., Lecerf, D., 2002, Multivariate geostatistical filtering of time-lapse seismic data for an improved 4D signature. 73rd Ann. Intern. Mtg., S.E.G., Expanded Abstracts.

Cressie, N., and Hawkins, D.H., 1980, Robust estimation of the variogram: I, Math Geol, 12 (2), 115-125.

Denis, V., Lejeune, J., and Robin, J. P. 2002. Spatio-temporal analysis of commercial trawler data using General Additive models: patterns of Loliginid squid abundance in the north-east Atlantic. ICES Journal of Marine Science, 59: 633–648.

Deraisme, J., and Rivoirard, J., 2009, Histogram modelling and simulations in the case of skewed distributions with a 0-effect: issues and new developments, IAMG 2009 Conference, 23-29 August, Stanford (Powerpoint presentation).

Deraisme, J., Rivoirard, J., and Carrasco Castelli, P., 2008, Multivariate Uniform Conditioning and Block simulations with Discrete Gaussian Model: Application to Chuquicamata Deposit, VIII International Geostatistics Congress, GEOSTATS 2008, 69-78.

Emery, X., 2009 a, The kriging update equations and their application to the selection of neighboring data, Computational. Geosciences, 13:269–280.

Emery, X., 2009 b. Change-of-support models and computer programs for direct block-support simulation, Computers & Geosciences 35:2047–2056.

Emery, X., and Ortiz J. M., 2005. Internal consistency and inference of change of support isofactorial models, *in* Geostatistics Banff 2004, Proceedings of the 7th International Geostatistics Congress, O. Leuangthong and C.V. Deutsch (eds.), Springer, Dordrecht, The Netherlands, Vol. 2, p. 1057-1066.

Emery X., and Ortiz J. M., 2007, Weighted sample variograms as a tool to better assess the spatial variability of soil properties, Geoderma 140, 81–89.

Emery, X., and Ortiz J. M., 2010. Two approaches to direct block-support conditional co-simulation, Computers and Geosciences, in press.

de Fouquet, C., 2006, La modélisation géostatistique des milieux anthropisés. Habilitation à Diriger des Recherches, Sciences de la Terre, Paris, Université Pierre et Marie Curie.

de Fouquet, C., and Mandallaz, D., 1992, Using geostatistics for forest inventory with air cover: an example. In: A. Soares, Editor, Geostatistics Troia'92, Vol. 2, Kluwer, Dordrecht, The Netherlands (1992), 875–886.

Genton, M., 1998, Highly Robust Variogram Estimation, Math Geol, 30 (2), 213-221.

Guibal, D., and Remacre, A. Z., 1984, Local estimation of the recoverable reserves: comparing various methods with the reality on a porphyry copper deposit, *in* Verly, G., David, M., Journel, A. G., and Maréchal, A., eds., Geostatistics for Natural Resources Characterization: Reidel, Dordrecht, 435–448.

Gribov, A., and Krivoruchko, K., 2004, Geostatistical mapping with continuous moving neighbourhood. Math Geol 36(2):267-281.

Guiblin, Ph., 1997, Analyse statistique de campagnes (acoustique et chalutage) sur le hareng écossais, Thèse de Doctorat en Géostatistique, Ecole des Mines de Paris, 161 p.

Guiblin, P., Rivoirard J., et Simmonds, E. J., 1995, Analyse structurale de données à répartition dissymétrique : exemple du hareng écossais. Cahiers de Géostatistique, 5, ENSMP, 137-159.

Guiblin, P., Rivoirard, J., and Simmonds, E. J., 1996, Spatial distribution of length and age for Orkney-Shetland herring. ICES CM 1996/D:14, 20 p.

Haas, A., Biver, P., et Moulière, D., 1998. Simulations stochastiques en cascade, *in* Cahiers de Géostatistique, Fasc. 6, ENSMP, 31-43.

Haldorsen, H. H., 1983, Reservoir Characterization Procedures for Numerical Simulation, Ph.D. Dissertation, The University of Texas at Austin.

Haldorsen, H. H., and Lake, L.W., 1984, A new approach to shale management in field-scale models, SPE Jour., v. 24, no. 4, 447-457.

Hastie, T. J., and Tibshirani, R. J., 1990. Generalized Additive Models. Chapman & Hall/CRC.

Ikeda S., Parker G., and Sawai K. 1981. Bend theory of river meanders. Part 1. Linear development, *Journal of Fluid Mechanics*, vol. 112, pp. 363-377.

Isaaks, E.H., and Srivastava, R.M., 1988. Spatial continuity for probabilistic deterministic geostatistics. Mathematical Geology, 20 : 313-341.

Isaaks, E.H., and Srivastava, R.M., 1989, An introduction to applied geostatistics. Oxford University Press, New York.

Jacod, J., et Joathon, P., 1971, Use of random-genetic models in the study of sedimentary processes, Math. Geol., vol. 3, no. 3, 265-279.

Jeulin, D., 1991. Modèles morphologiques de structures aléatoires et de changement d'échelle, Thèse de Docteur ès Sciences Physiques, Université de Caen.

Journel, A.G., 1982. The indicator approach to estimation of spatial data. Proceedings of the 17th APCOM, Port. City Press (New York), 793-806.

Journel, A.G., 1983: Nonparametric estimation of spatial distributions, Mathematical Geology, v.15, no.3, p.445-468.

Journel, A. G., and Huijbregts, Ch. J., 1978, Mining geostatistics, Academic Press, London. x + 600 p.

Krige, D.G., 1951, A Statistical. Approach to some Mine valuation and allied problems on the Witwatersrand, Master Thesis, University of Witwatersrand.

Lajaunie, C., et Lantuéjoul, C., 1989. Setting up the general methodology for discrete isofactorial models. *In* Geostatistics, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 323-334.

Lantuéjoul, C., 2001, Geostatistical Simulation: Models and Algorithms, Springer.

Lantuéjoul, C., et Rivoirard, J., 1982, Etude de la distribution des pentes fortes. Etude CNEXO. 43 p. Confidentiel.

Le Loc'h, G., Rivoirard, J. et Groupe Heresim, 1989, Geostatistical study of Tunu, Phase 1. Etude CG-IFP. 36 p. Confidentiel.

Le Loc'h, G., Rivoirard, J. and Groupe Heresim, 1990, Geostatistical study of Tunu, Phases 1 + 2. Etude CG-IFP. 139 p. Confidentiel.

Le Loc'h, G., Rivoirard, J., Renard, D., Ravenne, C., Groupe Heresim, 1992, Geostatistical study of Tunu gas field. Phase III : Non stationary simulations. 17 p. + annexes + fig. Confidentiel.

Lopez, S., 2003, Modélisation de reservoirs chenalisés méandriformes, approche génétique et stochastique, Ph. D., Ecole des Mines de Paris.

Lopez S., Cojan I., Rivoirard J., and Galli A., 2008. Process-based stochastic modelling: meandering channelized reservoirs. Spec. Publ. Int. Assoc. Sedimentol. 40 - 139:144.

Machuca-Mory, D. F., 2010. Geostatistics with Location-Dependent Statistics, PhD in geostatistics, University of Alberta.

Magneron, C., Jeannée, N., Le Moine, O., and Bourillet, J.-F., 2010. Integrating Prior Knowledge and Locally Varying Parameters with Moving-GeoStatistics: Methodology and Application to Bathymetric Mapping, GeoENV VII, Quantitative Geology and Geostatistics, Volume 16, 405-415.

Matheron, G., 1955, Applications des méthodes statistiques à l'évaluation des gisements, Annales des Mines, Paris, No 12, 50-75.

Matheron, G., 1965, Les variables régionalisées et leur estimation. Masson, Paris. 306 p.

Matheron, G., 1967, Éléments pour une théorie des milieux poreux, Masson, Paris.

Matheron, G. 1970, La théorie des variables régionalisées et ses applications, Cahiers du Centre de Morphologie Mathématique 5. Ecole des Mines de Paris, Fontainebleau.

Matheron, G., 1975, Random sets and integral geometry, Wiley, New-York.

Matheron, G., 1976 a. A simple substitute for conditional expectation: the disjunctive kriging, in Advanced Geostatistics in the Mining Industry, M. Guarascio, M. David, and C. Huijbregts eds. Reidel, Dordrecht, Holland, 221-236.

Matheron, G., 1976 b, Forecasting block grade distributions: the transfer functions. In Advanced geostatistics in the mining industry, M. Guarascio, M. David, and C. Huijbregts eds. Reidel, Holland, 237-251.

Matheron, G., 1979, Recherche de simplification dans un problème de cokrigeage, Technical Report N-628, Centre de Géostatistique, Fontainebleau, France, 19 p.

Matheron, G., 1982. La destructuration des hautes teneurs et le krigeage des indicatrices, Ecole des Mines de Paris.

Matheron, G., 1984. Une méthodologie générale pour les modèles isofactoriels discrets. Science de la Terre, Série Informatique Géologique, 21, 1-21.

Matheron, G., Beucher, H., de Fouquet, C., Galli, A. et Ravenne, C., 1987, Simulation conditionnelle à trois faciès dans une falaise de la formation du Brent. Sciences de la Terre, Série Informatique Géologique, 28, 213-249.

Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J-P., Guinet, C., 2006. Geostatistical modelling of spatial distribution of Balaenoptera physalus in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. Ecological Modelling 193:615-628.

Omre, H., 1984. The variogram and its estimation. In: Verly, G., David, M., Journel, A.G., Maréchal, A. (Eds.), Geostatistics for Natural Resources Characterization. Reidel, Dordrecht, 107–125.

Paciorek, C. J., and Schervish, M. J., 2006, Spatial modelling using a new class of nonstationary covariance functions, Environmetrics. 2006; 17(5): 483–506.

Pardo-Igúzquiza, E., and Dowd, P.A., 2005. On the automatic inference and modelling of a set of indicator covariances and cross-covariances. In: O. Leuangthong and C.V. Deutch (eds.), *Geostatistics Banff 2004*, Dordrecht: Springer, p. 185-194.

Parker, H. M., 1991, Statistical treatment of outlier data in epithermal gold deposit reserve estimation, Math. Geol., Vol. 23, No 2, 175-199.

Petitgas, P., 1991. Contributions géostatistiques à la biologie des pêches maritimes. Thèse de Doctorat en Géostatistique, Ecole des Mines de Paris, 211 p.

Petitgas, P. 1993. Geostatistics for fish stock assessments: a review and an acoustic application. ICES Journal of Marine Science, 50: 285-298.

Petitgas, P. 1993. Use of disjunctive kriging to model areas of high pelagic fish density in acoustic fisheries surveys. Aquatic Living Resources, 6: 201-209.

Petitgas, P., and Rivoirard, J., 1993. Global estimation : s^2/n and the geostatistical estimation variance, in "Report of the Workshop on the Applicability of Spatial Statistical Methods to Acoustic Survey Data", Copenhagen: ICES Cooperative Research Report 195, p.42-60.

Pyrcz, M.J., 2004. Integration of geologic information into geostatistical models. Ph.D. Dissertation, University of Alberta, Edmonton, Canada, 296 p.

Pyrcz, M.J., Boisvert, J. B., and Deutsch, C.V., 2009, Alluvsim: a Conditional Event-based Fluvial Model, Computers & Geosciences, vol. 35, no8, 1671-1685.

Reilly, C., and Gelman, A., 2007, Weighted classical variogram estimation for data with clustering, Technometrics, Vol. 49, No. 2, 184-194.

Remacre, A., 1984. L'estimation du récupérable local - Le conditionnement uniforme, Thèse de Doct.-Ing. en Sciences et Techniques Minières, Ecole des Mines de Paris.

Richmond, A., 2002. Two-point declustering for weighting data pairs in experimental variogram calculation. Computers & Geosciences 28 (2), 231–241.

Rivoirard, J., 1981, Précision de l'estimation du récupérable d'un champ de nodules. 52 p. Confidentiel.

Rivoirard, J., 1983, Remarques pratiques à propos des variances et du variogramme, Note de cours C-99, Centre de Géostatistique, Fontainebleau, France, 11 p.

Rivoirard, J., 1984 a, Looking for a kriging plan in a stockwerk deposit. In : G. Verly et al. (eds.), *Geostatistics for natural resources characterization*, Dordrecht : Reidel, Part 2, 951-963.

Rivoirard, J., 1984 b. *Le comportement des poids de krigeage*. Thèse de Docteur-Ingénieur en Sciences et techniques Minières, Option Géostatistique, ENSMP, 72 p.

Rivoirard, J., 1984 c, Une méthode d'estimation du récupérable local multivariable, Note interne non confidentielle, Centre de Géostatistique, Ecole des Mines de Paris, 8 p.

Rivoirard, J., 1985, Convergence des développements en polynômes d'Hermite. *Sciences de la terre, série Informatique géologique*, no. 24, 129-159.

Rivoirard, J., 1987 a, Computing variograms on uranium data. In G. Matheron et al. (eds.), Geostatistical case studies. Dordrecht : Reidel, p 1-22.

Rivoirard, J., 1987 b, Two key parameters when choosing the kriging neighbourhood. *Mathematical geology, 19* (8), 851-856.

Rivoirard, J., 1989, Models with orthogonal indicator residuals. In M. Armstrong (ed.), *Geostatistics*, Dordrecht : Kluwer, 1989, vol. 1, 91-107.

Rivoirard, J., 1994 a, Introduction to disjunctive kriging and non-linear geostatistics. Oxford : Clarendon Press, 181 p.

Rivoirard, J., 1994 b, Validité du modèle booléen et tests sur simulations (Second rapport). 50p. Confidentiel.

Rivoirard, J., 2000. Weighted variograms. In: Kleingeld, W.J., Krige, D.G. (Eds.), Geostatistics 2000 Cape Town. Geostatistical Association of Southern Africa, Cape Town, 145–155.

Rivoirard, J., 2001, Which models for collocated cokriging ? *Mathematical geology*, vol. 33, no. 2, 117-131.

Rivoirard, J., 2002 a, On the structural link between variables in kriging with external drift. *Mathematical geology*, vol. 34, no. 7,797-808.

Rivoirard, J., 2002 b. Analyse de la sensibilité à l'écrêtage à une coupure haute dans l'évaluation de l'or récupérable. Rapport interne, Centre de Géostatistique, Ecole des Mines de Paris, 25 p. Confidentiel.

Rivoirard, J., 2004, On some simplifications of cokriging neighbourhood. *Mathematical geology*, vol. 36, no. 8, 899-915.

Rivoirard, J., et Beucher, H., 1993, Validité du modèle booléen et tests sur simulations. 44 p. Confidentiel.

Rivoirard, J., and Bez, N., 1997, A 1D geostatistical analysis on Norwegian spring-spawning herring acoustic data in Ofotfjord (Dec. 1994). ICES CM 1997/Y:12, 9 p.

Rivoirard, J., et Bordessoule, J.-L., 1987, Etude géostatistique sur le gisement de Tréviels. 83 p. Confidentiel.

Rivoirard J., Cojan I., and Geffroy F., 2007, Controlling the location of deposits in meandering channelized reservoir models. In: Extended Abstracts Book, Petroleum Geostatistics 2007, Cascais, Portugal, European Association of Geoscientists & Engineers, Houten, The Netherlands, P19. 5 p.

Rivoirard J., Cojan I., Renard D. and Geffroy F., 2008, Advances in quantification of process-based models for meandering channelized reservoirs, VIII International Geostatistics Congress, GEOSTATS 2008, 607-616.

Rivoirard, J., Demange, C., Freulon, X., Lécureuil, A., Bellot, N., A top-cut model for deposits with very skewed grade distribution, submitted to Mathematical Geosciences.

Rivoirard, J., and Guiblin, Ph., 1997. Global estimation variance in presence of conditioning parameters. In Baafi and Schofield Eds, Geostatistics Wollongong'96, Kluwer Academic Publisher, Vol.1, 246-257.

Rivoirard, J., Renard, D., et Léger, M., 1993, Quantification de l'incertitude sur les profondeurs estimées par sismique et par puits. Rapport FSH, Centre de Géostatistique, ENSMP, 16 p. + planches. Confidentiel.

Rivoirard, J., Renard, D., Ravenne, C. and Heresim Group, 1993, Geostatistical study of Tunu gas field. Phase IV. Confidentiel.

Rivoirard, J., and Romary, T., 2011, Continuity for kriging with moving neighborhood, Mathematical geosciences 43: 469–481.

Rivoirard, J., Simmonds, J., Foote, K.G., Fernandes, P., and Bez, N., 2000. Geostatistics for estimating fish abundance. Blackwell Science, Oxford. 206 p.

Rossi, R. E., Mulla, D. J., Journel, A. G., & Franz, E. H., 1992, Geostatistical tools for modeling and interpreting ecological spatial dependence, Ecological Monographs 62(2), 277-314.

Roustant, O., Dupuy, D., and Helbert, C., 2007, Robust estimation of the variogram in computer experiments, Rapport interne, Département 3MI , Ecole des Mines de Saint-Etienne.

Samb, B. and Petitgas, P. 1997. Estimation de la précision des campagnes acoustiques au Sénégal par la méthode géostatistique transitive à une dimension. Aquatic Living Resources 10: 75-82.

Seigneurin, A., Muller, D., Galli, A., and Ravenne, C., 1993, Optimization of the well-spacing with a geostatistical model Tunu field - Mahakam area. In : Proc. Indonesian Petroleum Association, Twenty Second Annual Convention, Jakarta, Indonesia.

Simmonds, E. J., Bailey, M., Toresen, R., Couperus, B., Pedersen, J., Reid, D. G., Fernandes, P. G. & Hammer, C. 1997. 1996 ICES coordinated acoustic survey of ICES Divisions IIIa, IVa, IVb, and VIa. ICES C.M. 1997 /H:11, 13 p.

Snowden, V., Moore, C., and Kelly, D., 1994, Using Geostatistics to Assist in Optimising Grade Control Estimation at KCGM's Fimiston Open Pit, Proceedings of the 4th Large Open Pit Mining Conference, Perth, 5-9 September 1994, 117-131.

Teles, V., 1999, Construction de Réservoirs Aquifères Alluviaux par Modèle Génétique de mise en place des Alluvions. PhD Thesis, University Pierre et Marie Curie Paris VI, Mémoire des Sciences de la Terre, 99-03.

Teles, V., Marsily, G. de, Perrier, E., 1998, Sur une nouvelle approche de modélisation de la mise en place des sédiments dans une plaine alluviale pour en représenter l'hétérogénéité. CR Acad. Sci. Paris, 327 , 597–606.

Trenkel, V., 2006. Méthodes d'évaluation des ressources marines vivantes pour une exploitation durable. HDR. Université Montpellier 2.

Vann, J., and Guibal, D., 1998. Beyond Ordinary Kriging – An overview of non-linear estimation. Proceedings of Beyond Ordinary Kriging Seminar, Geostatistical Association of Australasia, Perth, Western Australia.

Verly, G., 1983. The multigaussian approach and its applications to the estimation of local reserves, Math. Geol. Vol. 15, No 2, 259-286.

Wackernagel, H., 1998. Multivariate geostatistics: an introduction with applications, 2nd ed., Springer, Berlin, 291 p.

Woillez, M. 2007. Contributions géostatistiques à la biologie halieutique, Thèse de Doctorat en Géostatistique, Ecole des Mines de Paris, 173 p.

Woillez, M., Poulard, J-C., Rivoirard, J., Petitgas, P., and Bez, N., 2007. Indices for capturing spatial patterns and their evolution in time with an application on European hake (*Merluccius merluccius*) in the Bay of Biscay. – ICES Journal of Marine Science, 64, 537-550.

Woillez, M., Rivoirard, J., and Petitgas, P., 2009, Notes on survey-based spatial indicators for monitoring fish populations. Aquat. Living Resour. 22, 155-164.

Woillez, M., Rivoirard, J., and Fernandes, P.G. 2009. Evaluating the uncertainty of abundance estimates from acoustic surveys using geostatistical simulations – ICES Journal of Marine Science, 66:1377–1383.

Xu, W., Tran, T. T., Srivastava, R. M., and Journel, A. G. 1992, Integrating seismic data in reservoir modeling: The collocated cokriging alternative. SPE paper 24742, 67th Annual Technical Conference and exhibition, 833-842.

Annexe: publications choisies

Bordessoule, J.-L., Demange, C., and Rivoirard, J., 1989, Using an orthogonal residual between ore and metal to estimate in-situ uranium resources. In M. ARMSTRONG (ed.), *Geostatistics*, Vol. 2. Dordrecht : Kluwer, 1989, 923-934.

Rivoirard, J., 1989, Models with orthogonal indicator residuals. In M. Armstrong (ed.), *Geostatistics*, Dordrecht : Kluwer, 1989, vol. 1, 91-107.

Rivoirard, J., 2004, On some simplifications of cokriging neighbourhood. *Mathematical geology*, vol. 36, no. 8, 899-915.

Woillez, M., Poulard, J-C., Rivoirard, J., Petitgas, P., and Bez, N., 2007. Indices for capturing spatial patterns and their evolution in time with an application on European hake (*Merluccius merluccius*) in the Bay of Biscay. ICES Journal of Marine Science, 64, 537-550.

Rivoirard J., Cojan I., Renard D. and Geffroy F., 2008, Advances in quantification of process-based models for meandering channelized reservoirs, VIII International Geostatistics Congress, GEOSTATS 2008, 607-616.