

ACADÉMIE DE PARIS  
UNIVERSITÉ PIERRE ET MARIE CURIE

Mémoires des Sciences de la Terre  
N° 2004.02

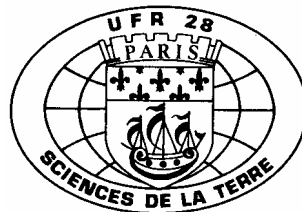
**Géostatistique et  
assimilation séquentielle de données**

Hans Wackernagel

Habilitation à diriger des recherches  
25 mai 2004

Jury

Jean-Paul	CHILÈS	Ecole des Mines de Paris	Examineur
Ghislain	DE MARSILY	Université Pierre et Marie Curie	Président
Helge	DRANGE	Université de Bergen	Rapporteur
Walid	TABBARA	Université Pierre et Marie Curie	Rapporteur
Olivier	TALAGRAND	Ecole Normale Supérieure	Rapporteur
Alain-jacques	VALLERON	Université Pierre et Marie Curie	Examineur



UFR 928 des Sciences de la Terre et Évolution des Milieux Naturels  
4, place Jussieu – 75252 Paris cedex 05

## Résumé

La géostatistique spatio-temporelle traditionnelle n'est pas en mesure de décrire adéquatement la dynamique, en général fortement non-linéaire, de processus spatio-temporels multivariés. Pour cela des modèles de transport physico-chimiques sont en général bien mieux adaptés. Cependant, étant donné que ces derniers ne maîtrisent pas totalement la complexité des processus qu'ils cherchent à décrire, soit parce qu'ils sont basés sur des hypothèses simplificatrices, soit parce que l'information servant à établir les conditions initiales et aux limites est imparfaite, il est opportun d'introduire des techniques statistiques servant à les guider pour assimiler un flot de mesures émanant de capteurs automatiques.

Des projets récents au Centre de Géostatistique de l'Ecole des Mines de Paris ont permis d'explorer l'application de ces techniques dans le domaine de l'océanographie et en pollution de l'air. Il s'est très vite avéré que la géostatistique offrait des concepts et des approches qui pouvaient contribuer à enrichir les techniques d'Assimilation Séquentielle de Données. La thèse de Laurent Bertino et des publications ultérieures ont permis de développer cette thématique porteuse, dont la présente synthèse établit un compte-rendu.

## Abstract

Traditional geostatistical space-time geostatistics is not able to take account of the generally strongly non-linear dynamics of multivariate space-time processes. To this effect physico-chemical transport models are in general more suitable. However, as the latter do not fully master the complexity of the processes they attempt to describe, either because of simplifying hypotheses or because the information serving to set up initial and boundary conditions is imperfect, it is appropriate to introduce statistical techniques in order to assimilate a flow of measurements emanating from automatic stations.

Recent projects at Centre de Géostatistique of Ecole des Mines de Paris have permitted to explore these techniques in oceanography and air pollution. Soon it became evident that geostatistics could offer concepts and approaches to enhance Sequential Data Assimilation techniques. The thesis of Laurent Bertino as well as subsequent publications have permitted to develop this fruitful theme of which the present synthesis presents an account.

## Remerciements

Je tiens d'abord à remercier tout particulièrement Ghislain de Marsily, professeur à l'Université Pierre et Marie Curie, membre de l'Académie des Sciences, pour m'avoir incité et encouragé à faire la démarche de présenter une partie de mes travaux scientifiques sous la forme d'une habilitation à diriger des recherches.

Les travaux sélectionnés ont été essentiellement le fruit de collaborations avec des partenaires européens dans le cadre de projets de recherche, et je mentionnerai surtout Hans von Storch, directeur de l'Institut de recherches côtières du GKSS à Geesthacht (Allemagne), et Geir Evensen, anciennement animateur du groupe de modélisation et assimilation de données au centre Nansen de Bergen (Norvège), et maintenant directeur d'un groupe de recherche au sein de la compagnie pétrolière Norsk Hydro.

Une grande partie des résultats ont été élaborés par des élèves de l'Ecole des Mines de Paris, et j'aimerais remercier chaleureusement Laurent Bertino, pour son enthousiasme permanent dans l'exécution des objectifs scientifiques des contrats européens Pioneer et Impact, et également Julien Sénégas, qui a dégagé et approfondi de premières idées permettant de relier la géostatistique et les méthodes d'assimilation séquentielle de données. Je tiens également à remercier mon collègue Christian Lajaunie, qui s'est intéressé plus particulièrement à la question de rapprocher l'assimilation de données et les filtres de particules pour le traitement de données d'épidémies de grippe. Je dois naturellement beaucoup au Centre de Géostatistique dans son ensemble, et en particulier à son directeur passé, Michel Schmitt, ainsi qu'à son successeur, Jean-Paul Chilès, pour l'accueil favorable réservé à ces recherches ayant a priori une orientation peu orthodoxe.

Je voudrais aussi exprimer mes sincères remerciements aux membres du jury, à commencer par Helge Drange, professeur à l'Université de Bergen, et directeur du Centre de Climatologie G.C. Rieber, qui a bien voulu venir de Norvège pour apporter son témoignage en cette occasion ; également à Walid Tabbara, professeur à l'Université Pierre et Marie Curie, pour son intérêt persistant pour la géostatistique, initié par des problèmes de plans d'expériences en compatibilité électromagnétique ; à Olivier Talagrand, directeur de recherches au Laboratoire de Météorologie Dynamique à l'Ecole Normale Supérieure, qui a toléré avec beaucoup de sympathie nos diverses incursions géostatistiques dans le domaine de l'assimilation de données ; et à Alain-jacques Valleron, professeur à l'Université Pierre et Marie Curie, membre correspondant de l'Académie des Sciences, et directeur de l'unité 444 de l'Inserm, avec laquelle nous avons eu le privilège de collaborer sur l'assimilation de données de syndrôme grippal.

# Table des matières

<b>I Synthèse</b>	<b>3</b>
<b>1 Préambule</b>	<b>4</b>
<b>2 Introduction</b>	<b>4</b>
2.1 Plan de la synthèse . . . . .	5
2.2 Présentation des travaux . . . . .	6
<b>3 L'état de l'art en géostatistique</b>	<b>7</b>
3.1 La théorie des variables régionalisées . . . . .	7
3.2 Géostatistique linéaire . . . . .	8
3.3 Géostatistique multivariable . . . . .	11
3.4 Géostatistique non-stationnaire . . . . .	12
3.5 Géostatistique non-linéaire . . . . .	13
3.6 Simulation géostatistique . . . . .	14
<b>4 Le krigeage avec des sorties de modèles numériques</b>	<b>15</b>
4.1 Données de station et modèle de transport . . . . .	15
4.2 Influence de la dérive externe . . . . .	16
4.3 Supports : données, modèle numérique, résultat . . . . .	17
4.4 Modèle de changement de support point-bloc-cellule . . . . .	18
4.5 Proportion locale de blocs au dessus d'un niveau d'alerte . . . . .	21
<b>5 Le krigeage et le filtre de Kalman</b>	<b>22</b>
5.1 Assimilation de données . . . . .	22
5.2 Filtre de Kalman . . . . .	23
5.3 Covariances des erreurs . . . . .	26
<b>6 L'assimilation séquentielle et la géostatistique</b>	<b>27</b>
6.1 Méthodes d'assimilation . . . . .	27
6.2 RRSQRT : Réduction de rang du filtre . . . . .	28
6.3 EnKF : filtre de Kalman d'Ensemble . . . . .	28
6.4 Différents apports de la géostatistique . . . . .	29

<b>7 Conclusion</b>	<b>31</b>
<b>Bibliographie</b>	<b>33</b>
<b>II Éléments biographiques</b>	<b>39</b>
8.1 Curriculum vitæ . . . . .	40
8.2 Evaluation et direction de travaux scientifiques . . . . .	40
8.3 Participation à des congrès . . . . .	43
8.4 Publications . . . . .	44
<b>III Annexes</b>	<b>48</b>
<b>A <i>Encyclopedia of Statistical Sciences</i>, 2nd edition, sous presse</b>	<b>50</b>
<b>B <i>International Journal of Climatology</i>, 14 : 77–91</b>	<b>66</b>
<b>C <i>Journal of Climate</i>, 12 : 1070-1085</b>	<b>82</b>
<b>D <i>Ecological Modelling</i>, sous presse</b>	<b>98</b>
<b>E <i>Stochastic Environmental Research and Risk Assessment</i>, 15 : 65–86</b>	<b>115</b>
<b>F <i>GeoENV III – Geostatistics for Environmental Applications</i>, 157–168, Kluwer</b>	<b>138</b>
<b>G <i>Inverse Problems</i>, 18 : 1–23</b>	<b>151</b>
<b>H <i>International Statistical Review</i>, 71 : 223–241</b>	<b>175</b>
<b>I <i>La Recherche</i>, 371 : 98–99</b>	<b>195</b>

# **Première partie**

## **Synthèse**

# 1 Préambule

Le présent mémoire en vue de l'obtention de l'habilitation à diriger des recherches se propose de rassembler quelques fruits de la recherche que j'ai effectuée depuis 1998 sur l'assimilation de données dans un cadre géostatistique. Ce travail s'est réalisé pour l'essentiel à l'occasion de deux contrats européens, PIONEER (1998–2001) et IMPACT (2000–2002). Ces deux projets, pour lesquels j'ai conçu, négocié et mené à bien le partenariat du Centre de Géostatistique de l'Ecole des Mines de Paris, tant en ce qui concerne les aspects scientifiques que les problèmes administratifs et financiers, ont occupé l'essentiel de mon temps ces cinq dernières années et m'ont permis de développer un axe de recherche nouveau au Centre de Géostatistique, à savoir l'*assimilation de données*. Cette recherche n'a naturellement pas été menée à bien de manière isolée et je tiens d'emblée à souligner les riches apports d'élèves de l'Ecole des Mines qui ont effectué leur travail de fin d'études dans les centres de recherche partenaires GKSS (Geesthacht, Allemagne) et NERSC (Bergen, Norvège) sur des thèmes reliant la géostatistique et l'assimilation de données : Julien Sénégas et Laurent Bertino (qui a ensuite produit une remarquable thèse de doctorat dans le cadre de PIONEER), ainsi que Clotilde Breuillin, Olivier Frendo, Marielle Inizan et Arnaud Hollard. Dans le cadre d'IMPACT il est important de mentionner la précieuse collaboration avec mon collègue Christian Lajaunie, qui lui aussi s'est intéressé aux techniques d'assimilation de données, ce qui a permis de lancer deux stages sur le sujet en épidémiologie de la grippe en collaboration avec l'unité 444 de l'Inserm (Bui, 2001 ; Biboud, 2002).

Dans ce document de synthèse j'ai choisi de relater cette exploration des liens existant entre la géostatistique et l'assimilation séquentielle de données, l'enjeu étant de voir les apports possibles de la première à la seconde.

## 2 Introduction

La géostatistique spatio-temporelle traditionnelle (Kyriakidis & Journel, 1999) n'est pas en mesure de décrire adéquatement la dynamique, en général fortement non-linéaire, de processus spatio-temporels multivariés. Pour cela des modèles de transport physico-chimiques sont en général bien mieux adaptés. Cependant, étant donné que ces derniers ne maîtrisent pas totalement la complexité des processus qu'ils cherchent à décrire, soit parce qu'ils sont basés sur des hypothèses simplificatrices, soit parce que l'information servant à établir les conditions initiales et aux limites est imparfaite, il est opportun d'introduire des techniques statistiques servant à les guider pour assimiler un flot de mesures émanant de capteurs automatiques.

Des projets récents au Centre de Géostatistique ont permis d'explorer ces techniques dans le domaine de l'océanographie et de la pollution de l'air. Il s'est très vite avéré que la géostatistique offrait des concepts et des approches qui pouvaient contribuer à enrichir les techniques d'Assimilation Séquentielle de Données (Sénégas et al., 2001). La thèse de Laurent Bertino (Bertino, 2001) et des publications ultérieures (Bertino et al., 2002 ; Bertino et al., 2003) ont permis de développer cette thématique très porteuse.

En parlant d'assimilation *séquentielle* (par opposition aux techniques *variationnelles*) de don-

nées, deux points de vues peuvent être adoptés. D'une part, du point de vue du concepteur de modèles numériques déterministes, l'assimilation de données peut être vue comme étant un algorithme permettant de corriger l'état du modèle au fur et à mesure que des données nouvelles arrivent. D'autre part, du point de vue du statisticien, le modèle numérique peut aider à améliorer la prédiction opérationnelle en profitant de la connaissance des relations non-linéaires entre différentes sources de données.

C'est le second point de vue que nous allons privilégier et chercher à développer en situant la Géostatistique au carrefour de tous les flux de données, qu'elles proviennent d'un réseau de capteurs, d'un modèle de transport couplé avec d'autres sources de données ou d'images de télédétection. La Géostatistique, avec ses modèles multivariés, ses méthodes d'anamorphose de distributions de probabilités et ses modèles de changement de support, ainsi que ses méthodes de simulations conditionnelles, offre un cadre intégrateur unique pour relier ces différentes informations, comprendre et modéliser leur structure statistique, élaborer des outils de prévision.

## 2.1 Plan de la synthèse

Cette synthèse nous sert à tisser un fil conducteur reliant des travaux dont la motivation remonte à la dernière décennie. Nous nous proposons de rapprocher les méthodes séquentielles d'assimilation de données et celles de la géostatistique. Dans le Chapitre 3 nous tentons de brosser un tableau succinct de l'état de l'art en géostatistique tout en définissant au passage un certain nombre de notions de base. Le Chapitre 4 propose une étude détaillée de la méthode de la dérive externe pour corriger une prévision générée par un modèle déterministe ; elle contient également une illustration de l'effet de support et de sa modélisation en géostatistique. Le Chapitre 5 explique la construction du filtre de Kalman, tel qu'il est employé en assimilation séquentielle de données, et montre que l'étape de correction correspond à un krigeage simple ; une brève présentation de la modélisation des termes d'erreurs est également proposée. Dans le Chapitre 6 nous donnons une définition plus précise de l'assimilation séquentielle et partons du filtre de Kalman étendu pour présenter les deux filtres qui ont été principalement étudiés dans nos applications, à savoir le filtre de Kalman à rang réduit (RRSQRT) et le filtre de Kalman d'ensemble (EnKF) ; après avoir constaté que le filtre d'ensemble est analogue, dans sa construction, aux simulations conditionnelles de la géostatistique nous examinons dans le détail les principaux apports possibles de la géostatistique, dont certains ont déjà pu être implémentés dans nos applications. Le Chapitre 7 résume les principales conclusions et tente, se positionnant d'un point de vue géostatistique, de dessiner des perspectives quant aux développements futurs de l'assimilation séquentielle de données.

En sus de la synthèse on trouvera en Partie II des éléments biographiques concernant le curriculum, l'évaluation et la direction de travaux scientifiques, la participation à des congrès et les publications de l'auteur.

## 2.2 Présentation des travaux

Nous reproduisons en Annexe A un article tentant de résumer succinctement l'état de l'art en géostatistique. En Annexe B nous présentons un travail dans lequel nous avons utilisé la méthode de la dérive externe, jusque-là confinée à des applications pétrolières, pour résoudre un problème de cartographie climatologique. Profitant de la relation linéaire qui existe entre température et altitude, nous nous sommes servis d'une grille régulière de valeurs topographiques de l'Ecosse comme dérive externe dans un krigeage des températures moyennes au mois de janvier. La relation linéaire avec l'altitude nous a permis de cartographier les températures même au niveau des monts Grampian (s'élevant jusqu'à 1344m) sur la base d'un jeu de stations dont la plus élevée n'était située qu'à 400m d'altitude. L'utilisation d'un voisinage glissant nous a amenés à nous intéresser à la cartographie du coefficient de la dérive externe, qui a montré que l'influence de la dérive externe est la plus forte dans les voisinages qui comportent la plus faible densité de stations.

Dans un autre travail en climatologie présenté en Annexe C, la géostatistique a été employée pour améliorer une méthode de "downscaling" afin d'estimer la pluviométrie à partir de données de station et de champs de pression atmosphérique, disponibles sur plusieurs décennies. Le krigeage n'y a pas été employé dans un espace géographique, mais dans l'espace des composantes principales de champs de pression, afin de remplacer par une combinaison linéaire la simple moyenne arithmétique utilisée jusqu'alors<sup>1</sup>. L'analyse structurale des valeurs de température dans le système de coordonnées des composantes principales les plus importantes des champs de pression a permis de comprendre dans quelles régions de la péninsule Ibérique le krigeage est plus performant : lorsque le variogramme s'avère très structuré, avec un comportement parabolique à l'origine, le krigeage est capable de donner de bien meilleures estimations que la simple moyenne arithmétique, tandis que quand il comporte un effet de pépité ses performances sont moins impressionnantes, sans parler du cas de l'effet de pépité pur, où il est tout simplement équivalent à la moyenne des données du voisinage.

Dans l'Annexe D nous présentons une étude conjointe de 19 valeurs horaires d'ozone provenant de capteurs Airparif et de valeurs correspondantes calculées sur une grille régulière par le modèle de transport Chimère. Dans une première partie, les valeurs de Chimère sont utilisées comme dérive externe dans un krigeage de blocs pour estimer des valeurs corrigées par les mesures Airparif. Dans la seconde partie, un modèle Gaussien discret est construit qui sert à estimer par conditionnement uniforme la proportion de petits blocs, à l'intérieur de chaque cellule de la grille Chimère, qui est au dessus du seuil de pollution réglementaire. Nous avons été obligés de spécifier arbitrairement la taille de ces petits blocs, car les réglementations sur les pollutions atmosphériques —très détaillées en ce qui concerne la spécification de seuils pour différents supports temporels— ne contiennent cependant pas d'indication sur le support spatial.

Le travail présenté en Annexe E a résulté de notre premier contact avec le filtre de Kalman en

---

<sup>1</sup>Une collaboration au même moment avec le Laboratoire de Signaux et Systèmes (LSS) de Supélec, suscitée par la thèse de Jean Lefèbre (dont j'ai été le rapporteur), nous a fait découvrir que nous n'étions pas les premiers à pratiquer le krigeage dans un espace non-géographique. Se basant à un article de Sacks et al. (1989) l'équipe du LSS pratiquait le krigeage d'une surface de réponse dans un contexte de plans d'expérience (Lefèbre et al., 1996 ; Walter & Pronzato, 1997).

tant qu'outil d'assimilation de données. La motivation de le publier provenait du fait que la littérature ne contenait que peu d'éléments sur l'estimation de l'erreur de modélisation déterministe et le modèle hydrodynamique Trim dans sa version 1D a servi à décortiquer les effets de la prise en compte de différents termes d'erreur dans un filtre de Kalman.

L'Annexe F contient une étude hydrodynamique de l'estuaire de l'Oder au moyen de Trim 3D et l'un de ses aspects originaux est de proposer un modèle de variogramme spatio-temporel pour l'erreur aux limites, dérivé de l'équation d'onde.

Dans l'Annexe G une étude plus détaillée de l'estuaire de l'Oder est proposée, comparant les performances de deux filtres de Kalman. Ce travail contient une autre comparaison des deux filtres de Kalman, cette fois pour un modèle écologique simplifié mettant en jeu des nutriments, du plancton et des herbivores.

L'Annexe H est constituée d'un article proposant une présentation synthétique des différentes méthodes d'assimilation séquentielles utilisées en océanographie et nous y insistons sur les aspects géostatistiques. En particulier, l'anamorphose y est présentée et appliquée dans le contexte du modèle écologique simplifié déjà mentionné ci-dessus.

Finalement, nous nous permettons de joindre également une tentative de vulgarisation, dans l'Annexe I, sous la forme d'un court article paru récemment dans *La Recherche*. L'article a d'ailleurs été suppléé d'un complément pour lecteurs plus avertis (disponible sur le site Web+ du magazine).

## 3 L'état de l'art en géostatistique

La géostatistique est une application de la théorie des fonctions aléatoires à des données localisées dans un espace géographique. Les méthodes géostatistiques, telles le krigeage, ont été initialement proposées en exploration minière et pétrolière et elles ont retrouvé leur place en statistiques il y a plus d'une décennie.

Une version anglaise du présent chapitre est reproduite en Annexe A.

### 3.1 La théorie des variables régionalisées

La géostatistique est construite autour du concept de *variable régionalisée*, c'est-à-dire une fonction  $z(\mathbf{x})$  dans un domaine  $D$  de l'espace géographique (qui inclut éventuellement un intervalle de l'axe temporel), où  $\mathbf{x}$  est un vecteur de coordonnées spatiales et  $z$  est une valeur de la quantité d'intérêt. La géostatistique peut être définie comme une application de méthodes probabilistes à des variables régionalisées. Une variable régionalisée peut être considérée comme déterministe et l'aléatoire y être introduit par l'échantillonnage : c'est l'objet de la théorie transitive qui a connu un nouvel essor dans l'estimation de stocks de poissons (Petitgas, 2001 ; Bez & Rivoirard, 2001). Toutefois, l'approche usuelle est de considérer la variable régionalisée comme une réalisation d'une fonction aléatoire  $Z(\mathbf{x})$ , c'est-à-dire une famille de variables aléatoires  $Z$  dépendant de la localisation  $\mathbf{x} \in D$ . Le problème épistémologique lié au fait qu'habituellement dans la pratique seule une réalisation partiellement échantillonnée est disponible a été résolu par

Matheron dans son livre *Estimer et Choisir* (1989) en distinguant entre des *quantités objectives* qui sont essentiellement des intégrales de la réalisation unique et des *paramètres conventionnels* qui sont associés au modèle de fonction aléatoire. Les premiers peuvent être estimés alors que les seconds doivent être choisis.

La géostatistique est née dans le domaine de la mine et de la géologie au milieu du siècle passé lorsqu'il s'agissait de d'estimer les réserves récupérables dans des gisements miniers sur la base de quelques échantillons ayant un histogramme en général très penché à gauche. De premières idées de Krige (1951) et de Wijs (1951) ont été mises dans un cadre commun par Matheron (1955, 1962, 1965, 1976a). Rapidement les méthodes géostatistiques se sont répandues dans d'autres domaines des sciences de la terre et, finalement, sont revenues par la grande porte dans les statistiques universitaires à la fin du siècle, où elles font désormais l'objet de nombreuses recherches (Cressie, 1993 ; Stein, 1999).

La géostatistique est classiquement subdivisée en géostatistique linéaire et multivariable, géostatistique non-linéaire, simulations géostatistiques. La présentation se fera dans cet ordre.

## 3.2 Géostatistique linéaire

En géostatistique on explore la structure de la variabilité spatiale à l'aide du *variogramme*  $\gamma(\mathbf{h})$ , qui est défini à partir de la variance des incréments  $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$ , où  $\mathbf{h}$  est le vecteur reliant deux points dans  $D$ . Sous l'hypothèse que les incréments sont stationnaires, c'est-à-dire que la variance des incréments est invariante pour toute translation du vecteur  $\mathbf{h}$ , on a

$$\text{var}(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) = 2\gamma(\mathbf{h}). \quad (3.1)$$

On suppose habituellement que l'espérance des accroissement est stationnaire et nulle, de sorte que

$$\gamma(\mathbf{h}) = \frac{1}{2} E[(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))^2]. \quad (3.2)$$

Des fonctions aléatoires dont les incréments sont stationnaires d'ordre deux et de moyenne nulle sont réputés satisfaire l'*hypothèse intrinsèque* (Matheron, 1965 ; Chilès & Delfiner, 1999), ce qui par abus de langage a été qualifié de *stationnarité intrinsèque* dans la littérature statistique (Cressie, 1993). Le variogramme n'est pas borné et est pour cela préférable à la fonction de covariance  $C(\mathbf{h}) = E[(Z(\mathbf{x} + \mathbf{h}) - \mu) \cdot (Z(\mathbf{x}) - \mu)]$ , laquelle est définie pour des fonctions aléatoires stationnaires d'ordre deux de moyenne  $\mu$ . Pour des fonctions aléatoires stationnaires d'ordre deux le variogramme est borné et peut être calculé à partir de la fonction de covariance,

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}). \quad (3.3)$$

Le variogramme est une fonction *de type négatif conditionnel*, à savoir que

$$\sum_{\alpha=0}^n \sum_{\beta=0}^n w_{\alpha} w_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \leq 0 \quad (3.4)$$

pour tout ensemble de  $n+1$  coefficients de somme nulle,

$$\sum_{\alpha=0}^n w_{\alpha} = 0. \quad (3.5)$$

Des combinaisons linéaires de variables aléatoires  $Z(\mathbf{x}_{\alpha})$  qui satisfont à la contrainte (3.5) sont qualifiées d'*autorisées* et la propriété (3.4) est alors évidente quand on calcule la variance d'une combinaison linéaire autorisée.

Ainsi  $-\gamma(\mathbf{h})$  est de type positif conditionnel, tandis qu'une fonction de covariance est *de type positif*,

$$\sum_{\alpha=0}^n \sum_{\beta=0}^n w_{\alpha} w_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \geq 0$$

pour tout jeu de  $n+1$  coefficients  $w_{\alpha}$  ; il n'y a pas de condition sur la somme des poids.

Un variogramme est *isotrope*, si  $\gamma(\mathbf{h})$  dépend seulement de la longueur  $|\mathbf{h}|$  et non de l'orientation du vecteur  $\mathbf{h}$ . Un moyen simple d'obtenir un variogramme anisotrope à partir d'un variogramme isotrope est d'effectuer une transformation linéaire des coordonnées du vecteur  $\mathbf{h}$  ; cela génère ce que l'on qualifie d'*anisotropie géométrique*.

Les fonctions de variogramme sont définies dans le cas isotrope. Une classe simple de modèles non bornés est la classe des modèles puissance

$$\gamma(\mathbf{h}) = b |\mathbf{h}|^{\nu} \quad \text{avec } b > 0, 0 < \nu < 2,$$

où  $\nu$  est un paramètre de forme produisant un aspect convexe, linéaire ou concave, respectivement pour des valeurs de  $\nu > 1, \nu = 1, \nu < 1$ .

Des variogrammes bornés sont obtenus à partir de fonctions de covariance par la formule (3.3). Une classe de fonctions de covariances est celle dite de *Matérn*, construite à l'aide de fonctions de Bessel  $\mathcal{K}_{\nu}$  de la troisième espèce

$$C(\mathbf{h}) = b \left( \frac{|\mathbf{h}|}{a} \right)^{\nu} \mathcal{K}_{\nu} \left( \frac{|\mathbf{h}|}{a} \right) \quad \text{avec } b > 0, \nu > 0.$$

où  $a$  est un paramètre d'échelle et  $\nu$  le paramètre de forme, qui mesure le degré de différentiabilité de la fonction aléatoire. Pour  $\nu = 1/2$  la fonction de covariance est exponentielle et a un aspect linéaire au voisinage de l'origine. En général le comportement au voisinage de l'origine est en  $|\mathbf{h}|^{2\nu}$  pour  $\nu$  non entier et en  $|\mathbf{h}|^{2\nu} \log |\mathbf{h}|$  pour  $\nu$  entier (pour une discussion détaillée de cette classe de modèles cf Matheron, 1965 ; Yaglom, 1986 ; Stein, 1999 ; Chilès & Delfiner, 1999).

Pour un jeu de  $n$  données  $z(\mathbf{x}_{\alpha})$  il est recommandé de procéder à une analyse exploratoire incluant une représentation de la *nuée variographique* (pour  $n$  de l'ordre de quelques centaines), c'est-à-dire un diagramme des différences quadratiques  $(z(\mathbf{x}_{\alpha}) - z(\mathbf{x}_{\beta}))^2/2$  reportées par rapport aux vecteurs  $\mathbf{h}_{\alpha\beta} = \mathbf{x}_{\alpha} - \mathbf{x}_{\beta}$  pour chaque couple de points (Chauvet, 1982 ; Haslett et al., 1991). Le variogramme *expérimental* est obtenu en regroupant les vecteurs  $\mathbf{h}_{\alpha\beta}$  en classes de distance (et éventuellement d'angle)  $\mathcal{H}_k$  et calculant les valeurs moyennes pour chaque classe,

$$\hat{\gamma}(\mathcal{H}_k) = \frac{1}{2|\mathcal{H}_k|} \sum_{\mathbf{h}_{\alpha\beta} \in \mathcal{H}_k} (z(\mathbf{x}_{\alpha}) - z(\mathbf{x}_{\beta}))^2,$$

où  $|\mathcal{H}_k|$  est le cardinal de la classe. Un estimateur hautement robuste du variogramme est suggéré dans Genton (1998), où des références antérieures sur le sujet peuvent être trouvées.

La séquence de valeurs expérimentales de variogramme pour différentes classes de distance n'est pas une fonction de type négatif conditionnel et il est nécessaire de l'ajuster par une fonction paramétrique de variogramme. Souvent cette ajustement est effectué "à la main", mais évidemment des procédures pour effectuer cette opération par moindres carrés (Cressie, 1985), par maximum de vraisemblance (Mardia & Marshall, 1984 ; Zimmerman & Zimmerman, 1991) ou dans un cadre Bayésien (Handcock & Stein, 1993 ; Diggle et al., 1998) ont aussi été proposées.

Munis à ce stade d'un modèle de variogramme valide on peut se pencher sur la question de son utilisation pour prédire une valeur  $\hat{z}(\mathbf{x}_0)$  en un point arbitraire  $\mathbf{x}_0$  du domaine  $D$ . Nous nous plaçons dans le cadre intrinsèque et utilisons un prédicteur linéaire,

$$\hat{Z}(\mathbf{x}_0) = \sum_{\alpha=1}^n w_{\alpha} Z(\mathbf{x}_{\alpha}),$$

pour lequel les pondérateurs sont de somme unité,

$$\sum_{\alpha=1}^n w_{\alpha} = 1, \quad (3.6)$$

pour travailler avec une combinaison linéaire autorisée. La contrainte (3.6) est en effet équivalente à la condition (3.5) avec  $w_0 = -1$ , de sorte que l'on pourra utiliser le variogramme pour calculer la variance de la combinaison linéaire autorisée.

L'erreur de prédiction est sans biais,

$$E(\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)) = \sum_{\alpha=1}^n w_{\alpha} E(Z(\mathbf{x}_{\alpha}) - Z(\mathbf{x}_0)) = 0.$$

La variance de notre problème de prédiction s'écrit

$$\text{var}(\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)) = - \sum_{\alpha=1}^n \sum_{\beta=1}^n w_{\alpha} w_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + 2 \sum_{\alpha=1}^n w_{\alpha} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0),$$

sachant que  $\gamma(\mathbf{x}_0 - \mathbf{x}_0)$  vaut zéro par définition (3.2).

La minimisation de la variance d'estimation sous la contrainte (3.6) par la méthode de Lagrange avec son paramètre  $\lambda$  mène aux équations du *krigeage ordinaire*

$$\begin{cases} \sum_{\beta=1}^n w_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \lambda = \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \forall \alpha \\ \sum_{\beta=1}^n w_{\beta} = 1 \end{cases} \quad (3.7)$$

produisant la variance au minimum

$$\sigma_{OK}^2 = \lambda + \sum_{\alpha=1}^n w_{\alpha} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0).$$

Dans le cadre de fonctions aléatoires stationnaires d'ordre deux on peut utiliser le krigeage pour estimer la moyenne  $\hat{\mu}$  à partir des données autocorrélées au moyen du système

$$\begin{cases} \sum_{\beta=1}^n w_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \lambda = 0 & \forall \alpha \\ \sum_{\beta=1}^n w_{\beta} = 1 \end{cases}$$

avec la variance d'estimation minimale

$$\sigma_{KM}^2 = \lambda.$$

Dans le cadre stationnaire d'ordre deux le krigeage ordinaire (3.7) peut être qualifié de *krigeage à moyenne inconnue*, car le modèle de fonction aléatoire est  $Z(\mathbf{x}) = \mu + Y(\mathbf{x})$ , avec  $Y(\mathbf{x})$  une fonction aléatoire de moyenne nulle.

### 3.3 Géostatistique multivariable

Il y a deux manières de formuler la généralisation du variogramme pour modéliser la corrélation croisée entre des incréments : le *variogramme croisé* (Matheron, 1965) et le *pseudo-variogramme croisé* (Myers, 1991). Le variogramme croisé est défini comme étant la covariance croisée des incréments de deux fonctions aléatoires  $Z_i(\mathbf{x})$  et  $Z_j(\mathbf{x})$ ,

$$\begin{aligned} \gamma_{ij}(\mathbf{h}) &= \frac{1}{2} \text{cov}((Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})) (Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x}))) \\ &= \frac{1}{2} E((Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})) (Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x}))) \end{aligned}$$

où l'on suppose la stationnarité conjointe des incréments, tandis que le pseudo-variogramme croisé  $\pi_{ij}(\mathbf{h})$  est basé sur la variance des incréments croisés

$$\begin{aligned} \pi_{ij}(\mathbf{h}) &= \frac{1}{2} \text{var}(Z_i(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})) \\ &= \frac{1}{2} E((Z_i(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x}))^2) \end{aligned}$$

où l'on postule la stationnarité d'incréments croisés de moyenne nulle. Les mérites et les inconvénients des deux approches ont conduit à des débats dont (Cressie & Wikle, 1998) est l'un des

derniers épisodes. Pour une couple de fonctions aléatoires conjointement stationnaires d'ordre deux de fonction de covariance

$$C_{ij}(\mathbf{h}) = E((Z_i(\mathbf{x} + \mathbf{h}) - \mu_i)(Z_j(\mathbf{x}) - \mu_j)) \quad (3.8)$$

les deux formulations du variogramme croisé s'écrivent

$$\begin{aligned} \gamma_{ij}(\mathbf{h}) &= C_{ij}(0) - \frac{1}{2}(C_{ij}(-\mathbf{h}) + C_{ij}(\mathbf{h})), \\ \pi_{ij}(\mathbf{h}) &= \frac{1}{2}(C_{ii}(0) + C_{jj}(0)) - C_{ij}(\mathbf{h}). \end{aligned}$$

La première formulation donne une fonction paire et ne permet donc pas de modéliser des décalages de corrélation entre deux variables. La seconde formulation le permet, mais par contre il s'agit d'une fonction positive qui ne reproduit plus les corrélations de signe négatif. Seule la fonction de covariance croisée  $C_{ij}(\mathbf{h})$  permet à la fois de modéliser les corrélations différées et les corrélations négatives; cependant on ne bénéficie plus du cadre plus souple des fonctions aléatoires intrinsèques, auquel le variogramme se réfère.

Des méthodes pour ajuster des matrices de fonctions  $C_{ij}(\mathbf{h})$ ,  $\gamma_{ij}(\mathbf{h})$  ou  $\pi_{ij}(\mathbf{h})$  à des fonctions expérimentales sont décrites dans (Goulard & Voltz, 1992; Grzebyk & Wackernagel, 1994; Ver Hoef & Barry, 1998; Gelfand et al., 2002).

En possession d'un modèle valide pour un ensemble de fonctions aléatoires la valeur d'une variable particulière peut être estimée par *cokrigeage* à partir d'échantillons. De multiples configurations de voisinage de cokrigeage sont possibles, suivant que la localisation des échantillons est commune ou non à plusieurs variables; certaines d'entre elles entraînent des simplifications du problème de cokrigeage, ce qui est important tant en ce qui concerne la taille qu'en ce qui relève de la stabilité numérique du système de cokrigeage (Rivoirard, 2001; Wackernagel, 2003).

### 3.4 Géostatistique non-stationnaire

Dans certaines situations la variable d'intérêt primaire, disons  $Z(\mathbf{x})$ , est échantillonnée en seulement quelques endroits, tandis qu'une variable secondaire  $s(\mathbf{x})$  est disponible sur une maille dense sur l'ensemble du domaine d'étude. Des exemples pour  $s(\mathbf{x})$  sont les données sismiques en exploration pétrolière (Chilès & Delfiner, 1999), l'altitude lorsque l'on cherche à cartographier la température (Hudson & Wackernagel, 1994), ou encore des sorties de modèles déterministes qui prédisent la variable régionalisée d'intérêt sur la base de processus physiques décrits par des équations aux dérivées partielles (Lajaunie et al., 2001). Sous hypothèse d'une relation linéaire entre les deux variables,  $E(Z(\mathbf{x})) = a + b s(\mathbf{x})$ , on modélise  $Z(\mathbf{x})$  avec une moyenne  $s(\mathbf{x})$  non-

stationnaire qualifiée de *dérive externe*. Les équations correspondantes sont alors

$$\left\{ \begin{array}{l} \sum_{\beta=1}^n w_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \lambda_0 + \lambda_1 s(\mathbf{x}_{\alpha}) = \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) \quad \forall \alpha \\ \sum_{\beta=1}^n w_{\beta} = 1 \\ \sum_{\beta=1}^n w_{\beta} s(x_{\beta}) = s(x_0) \end{array} \right. \quad (3.9)$$

Les liens avec le cokrigeage sont expliqués dans Rivoirard (2002), tandis que dans Royle & Berliner (1999) une généralisation de la méthode est proposée dans le cadre d'un modèle Bayésien hiérarchique.

Au lieu d'une dérive externe il est également possible d'utiliser une dérive polynomiale (linéaire ou quadratique) et l'on parle alors de *krigeage universel* (Matheron, 1970). Le variogramme des résidus entre la variable primaire et la dérive sont problématiques à inférer comme on peut le relire dans les textes traitant du krigeage universel. Pour cette raison, Matheron (1973) a développé pour les dérivées exponentielles-polynômes une théorie qui résoud les problèmes d'inférence par le recours à des classes plus générales de fonctions de covariance, les *covariances généralisées*, dont le variogramme est un cas particulier. Ce cadre inclus évidemment les dérivées trigonométriques (Séguret & Huchon, 1990). L'équivalence entre le krigeage non-stationnaire et les splines a été étudiée (Matheron, 1981 ; Wahba, 1990b) et a donné lieu à des discussions (Wahba, 1990a ; Cressie, 1990).

Les données spatio-temporelles ont été modélisées avec différents types de modèles de covariance spatio-temporels (Kyriakidis & Journel, 1999 ; Gneiting, 2002). Les liens entre le krigeage et le filtre de Kalman ont été reconnus (Mardia et al., 1998 ; Wikle & Cressie, 1999). Le filtre de Kalman et d'autres méthodes d'assimilation de données sont particulièrement intéressantes quand un réseau de stations automatiques et un modèle numérique décrivant la dynamique temporelle, généralement non-linéaire, sont disponibles (Bertino et al., 2002 ; Bertino et al., 2003).

### 3.5 Géostatistique non-linéaire

Les méthodes de prédiction linéaire sont conçues pour estimer des valeurs de  $Z$  en des points ou pour des intégrales  $\mu(\mathbf{x})$  sur le domaine. Cependant, lorsqu'il s'agit d'estimer la proportion de valeurs au dessus d'un seuil pour des unités spatiales données, les méthodes non-linéaires sont préférables.

Un seuil (valeur limite) spécifié dans une réglementation sur l'environnement, par exemple, est défini par rapport à un intervalle de temps, une unité de surface ou un volume, qualifiés de *support*, lequel est souvent d'une taille différente du support des échantillons. Nous sommes alors confrontés à un problème de *changement de support* qui consiste en l'anticipation de la loi des valeurs sur le support de sélection quand on ne connaît que la loi des échantillons sur

support ponctuel. La classification, au dessus ou en dessous d'un seuil, des valeurs prédites dépend également d'un *effet d'information*, dû à l'échantillonnage incomplet du domaine spatial ou temporel, et l'ampleur de cet effet va dépendre du type d'estimateur utilisé.

L'estimation non-linéaire est généralement effectuée en utilisant l'une des formulations du *krigeage disjonctif*, qui nécessite une connaissance des lois bivariées point-point et point-bloc. Les modèles de changement de support sont construits avec des *modèles isofactoriels* de lois bivariées qui ont été établis par Matheron en s'inspirant de modèles analogues utilisés en analyse des correspondances continues et en mécanique quantique (Matheron, 1976b ; Matheron, 1984). Les résultats sont présentés en utilisant différents types de *courbes de sélectivité*, qui ressemblent à des représentations graphiques usuelles en économétrie et en analyse de données de survie, et qui sont particulièrement commodes pour comparer l'impact du changement de support par rapport à des seuils donnés (Rivoirard, 1994 ; Chilès & Delfiner, 1999 ; Wackernagel, 2003).

### 3.6 Simulation géostatistique

Armé d'un modèle stochastique spatial de fonctions ou d'ensembles aléatoires, de processus ponctuels ou de populations d'objets, il existe une foule de techniques et algorithmes pour générer des réalisations. La difficulté majeure est de générer des réalisations qui coïncident aux points de données avec le phénomène régionalisé étudié.

Certains algorithmes génèrent directement des simulations de fonctions aléatoires conditionnellement aux données, soit par des méthodes séquentielles (Deutsch & Journel, 1992 ; Gómez-Hernández & Cassiraga, 1994), soit sur la base d'une décomposition de Cholesky de la matrice de covariance (Davis, 1987).

L'alternative est de générer d'abord des simulations non-conditionnelles, dont la compatibilité des propriétés statistiques avec le modèle spatial stochastique est vérifiée, et de les conditionner dans un second temps : dans le cas de fonctions aléatoires Gaussiennes ce conditionnement se fait par krigeage, sur la base de données transformées si nécessaire. Pour  $Y_S(\mathbf{x})$  une fonction aléatoire Gaussienne simulée non-conditionnellement on a trivialement

$$Y_S(\mathbf{x}_0) = Y_S^*(\mathbf{x}_0) + (Y_S(\mathbf{x}_0) - Y_S^*(\mathbf{x}_0)),$$

où  $Y_S^*(\mathbf{x}_0)$  est une estimation par krigeage simple à partir des valeurs de  $Y_S$  en  $n$  points de données  $\mathbf{x}_\alpha$ . L'astuce (Journel, 1974) consiste à remplacer le premier terme de droite par la valeur  $Y^*(\mathbf{x}_0)$  krigée à partir des données  $Y(\mathbf{x}_\alpha)$ , ce qui donne la valeur  $Y_{CS}$  de la simulation conditionnelle,

$$Y_{CS}(\mathbf{x}_0) = Y^*(\mathbf{x}_0) + Y_S(\mathbf{x}_0) - Y_S^*(\mathbf{x}_0).$$

Sachant que le krigeage simple est  $Y^*(\mathbf{x}_0) = (\mathbf{C}^{-1} \mathbf{c}_0)^\top \mathbf{y}$ , où  $\mathbf{C}$  est la matrice des covariances  $C(\mathbf{x}_\alpha - \mathbf{x}_\beta)$  entre les points de données et  $\mathbf{c}_0$  est le vecteur des covariances  $C(\mathbf{x}_\alpha - \mathbf{x}_0)$ , la simulation conditionnelle se calcule par

$$Y_{CS}(\mathbf{x}_0) = Y_S(\mathbf{x}_0) + (\mathbf{C}^{-1} \mathbf{c}_0)^\top (\mathbf{y} - \mathbf{y}_S). \quad (3.10)$$

Une description détaillée de la grande variété des méthodes de simulation géostatistique peut être trouvée dans les ouvrages de Chilès & Delfiner (1999), Lantuéjoul (2001). Le problème particulier de simuler des faciès géologiques et des lithotypes en utilisant des Gaussiennes tronquées est développé dans Armstrong et al. (2003). La simulation de réseaux de fractures, quant à elle, est décrite dans Chilès & de Marsily (1993).

## 4 Le krigeage avec des sorties de modèles numériques

Nous proposons un cas d'étude simple, dans lequel on examine d'abord la question de l'interpolation de données en incorporant des sorties de modèles numériques. Puis on abordera le problème de se trouver en présence de trois supports : celui des données et celui des cellules de la grille du modèle numérique, sachant que les résultats devront se référer à un troisième support, différent des deux premiers.

On trouvera la version anglaise de cette étude en Annexe D.

### 4.1 Données de station et modèle de transport

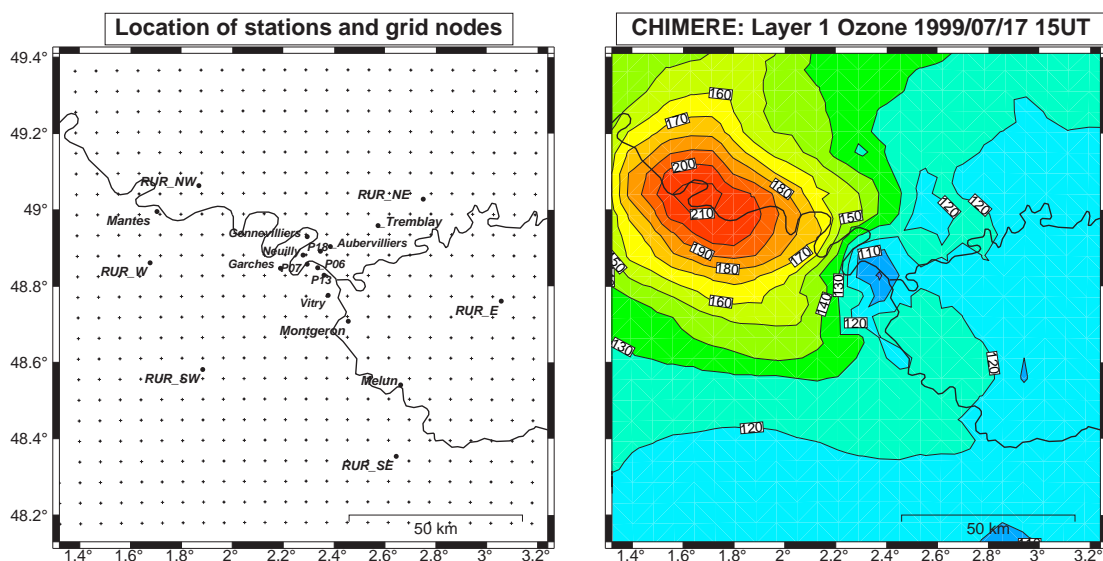


FIG. 4.1 – *A gauche* : implantation de 19 stations Airparif mesurant l’ozone et des  $25 \times 25$  noeuds de la grille Chimere. *A droite* : prédiction du modèle de transport Chimere pour le 17 juillet 1999 à 15 heures UTC.

La **Figure 4.1** montre un exemple typique de données de station et de sortie de modèle. Sur le graphique de gauche on voit la localisation en Ile de France d’une grille couvrant un domaine

de  $150 \times 150 \text{ km}^2$  avec des cellules de taille  $6 \times 6 \text{ km}^2$  et les implantations de 19 stations Airparif mesurant l'ozone, dont 6 stations qualifiées de "rurales". Sur le graphique de droite on voit la prédiction du modèle Chimère pour le 17 juillet 1999 à 15 heures UTC, c'est-à-dire 17 heures de l'heure d'été française. On remarque que le panache d'ozone est situé au NW de Paris en cette journée caractérisée par un temps stagnant (vitesse du vent à Roissy : 3 m/s).

Les données Airparif proviennent de *station de fond*, implantées à l'écart de routes ou de sites industriels importants. Chimère est un modèle de transport chimique simplifié, décrivant 200 processus relatifs à 80 substances gazeuses mis au point à l'Institut Pierre-Simon de Laplace (Vautard et al., 2001 ; Schmidt et al., 2001) et la sortie dont nous disposons provient de la version multi-échelle (Blond, 2002). Les prédictions Chimère ont été obtenues sur la base de prédictions météorologiques du Centre Européen de Prédiction Météorologique à Moyen Terme (ECMWF, à Reading, Angleterre) et d'un cadastre d'émissions établi avec des données de 1998.

## 4.2 Influence de la dérive externe

Dans un krigeage avec dérive externe le coefficient multiplicatif  $b$  donne l'ordre de grandeur de l'influence de la dérive. C'est particulièrement intéressant lorsque l'on utilise un voisinage glissant comme dans Hudson & Wackernagel (1994) : on remarque alors que le coefficient devient plus fort dans les régions où la densité de données est plus faible ; l'estimation repose plus fortement sur la dérive externe  $s(\mathbf{x})$  lorsqu'il y a moins de données disponibles sur la variable primaire  $Z(\mathbf{x})$  dans le voisinage.

Dans Lajaunie et al. (2001) la question a été analysée plus finement sur la base d'une modélisation du résidu  $Y(\mathbf{x})$  entre la variable d'intérêt  $Z(\mathbf{x})$  et la dérive externe  $s(\mathbf{x})$  multipliée par un coefficient  $b$ ,

$$Y(\mathbf{x}) = Z(\mathbf{x}) - b s(\mathbf{x}), \quad (4.1)$$

où  $Y(\mathbf{x})$  est une fonction intrinsèque d'ordre  $k$  de covariance généralisée  $K$ , et dont la dérive (non externe) est constituée à partir d'un ensemble de fonctions de base  $\{f_0, f_1, \dots, f_L\}$  qui sont des monômes au plus d'ordre  $k$ , c'est-à-dire à 1D on a  $L = k$  et les fonctions de base sont :  $f_0 = 1$ ,  $f_1 = x$ ,  $\dots$ ,  $f_k = x^k$ . Si le coefficient  $b$  était connu, on pourrait estimer directement  $Y^*(\mathbf{x}_0) = \mathbf{w}^\top (\mathbf{z} - b \mathbf{s})$  par le système de krigeage universel :

$$\begin{cases} \mathbf{K} \mathbf{w} + \mathbf{F} \mathbf{l} = \mathbf{k}_0 \\ \mathbf{F}^\top \mathbf{w} = \mathbf{f}_0 \end{cases} \quad (4.2)$$

où  $\mathbf{l}$  est le vecteur des multiplicateurs de Lagrange et  $\mathbf{F}$  est la matrice  $n \times (L+1)$  des fonctions de base.

Pour estimer le coefficient de la dérive externe à partir des données d'un voisinage on définit l'estimateur  $b^* = \mathbf{w}_b^\top \mathbf{z}$  pour lequel les poids  $\mathbf{w}_b$  sont déterminés par un système dont le membre gauche est du même type (au nombre de monômes près) que celui de (3.9) tandis que seul le

dernier élément du membre droit est non nul :

$$\begin{cases} \mathbf{K} \mathbf{w}_b + \mathbf{F} \mathbf{l} + \lambda_s \mathbf{s} & = 0 \\ \mathbf{F}^\top \mathbf{w}_b & = 0 \\ \mathbf{s}^\top \mathbf{w}_b & = 1 \end{cases}$$

où  $\lambda_s$  est un multiplicateur de Lagrange.

L'estimateur de  $Z(\mathbf{x})$  dans le cadre du modèle (4.1) de krigeage universel avec dérive externe se décompose en deux termes :

$$Z^*(\mathbf{x}_0) = \mathbf{w}^\top \mathbf{z} + b^* (s(\mathbf{x}_0) - \mathbf{w}^\top \mathbf{s}).$$

Le second terme est le produit de :

- $b^*$  qui dépend des données  $\mathbf{z}$  et caractérise l'importance de l'influence de la dérive externe ; un coefficient nul indique qu'une dérive proportionnelle à  $s(\mathbf{x})$  n'est pas appropriée.
- et de la différence  $s(\mathbf{x}_0) - \mathbf{w}^\top \mathbf{s}$  qui dépend uniquement de la covariance  $K$  dans le système (4.2) et non des données  $z(\mathbf{x}_\alpha)$  ; lorsque cette différence est faible (par exemple pour  $s(\mathbf{x})$  constant dans le voisinage), la dérive n'aura pas d'influence dans les faits même si  $b^*$  est important.

Le coefficient  $b^*$  a été calculé avec deux modèles de covariance différents pour les prédictions horaires Chimère pour les trois journées du 16-18 juillet 1999 et les mesures Airparif correspondantes (Lajaunie et al., 2001). Outre des coefficients nuls, des coefficients négatifs ont pu apparaître (en milieu de nuit), ce qui s'interprète comme étant une contradiction entre données et prédiction Chimère, cette dernière prédisant le contraire de ce qui a été mesuré. Pour l'exemple que nous traitons (17 juillet à 15h UTC) le coefficient  $b^*$  s'est avéré important pour les deux modèles de covariance, et positif.

Nous ne donnons pas tous les détails d'une étude que l'on trouve dans Wackernagel (2002) et nous ne montrons que le résultat du krigeage avec dérive externe sur la **Figure 4.2**, sachant que l'on a calculé des valeurs moyennes sur des cellules (ce qui implique l'utilisation de covariances moyennes point-cellules dans le membre droit du krigeage). L'abscisse des deux diagrammes se réfère à la prédiction Chimère ; le diagramme de gauche compare les valeurs de Chimère avec les mesures des 19 stations et l'on remarque qu'elles sont au dessus de la bisectrice, sauf pour quatre stations (coloriées en bleu) qui se trouvent au SE de l'Ile-de-France. Le diagramme de droite oppose les valeurs de Chimère aux valeurs obtenues par krigeage avec dérive externe : ces dernières sont en moyenne de  $13 \mu\text{g}/\text{m}^3$  plus élevées que les valeurs Chimère et nous rajouterons que la forme du nuage sera plus ou moins bombée suivant le modèle de variogramme utilisé.

### 4.3 Supports : données, modèle numérique, résultat

En pollution de l'air de nombreuses réglementations existent, dont Garcia et al. (2001) ont fait un inventaire. Si l'on a bien pensé au support temporel en spécifiant des seuils maximaux différents pour des moyennes sur 1h, 4h, 8h ou 24h, l'on a cependant pas, à notre connaissance,

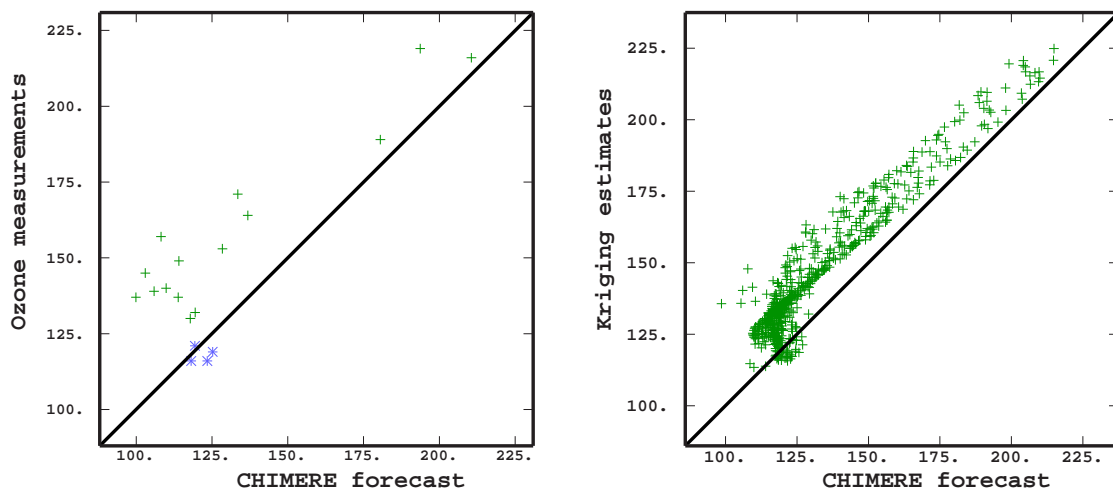


FIG. 4.2 – *A gauche* : les valeurs de Chimère (abscisse) et les mesures des 19 stations (ordonnée). *A droite* : les valeurs de Chimère (abscisse) et les valeurs obtenues par krigeage avec dérive externe (ordonnée).

spécifié le support spatial auquel un seuil de pollution est sensé se référer<sup>1</sup>. Un support ponctuel n'a pas beaucoup de sens car la pollution peut varier fortement d'un point à un autre, immédiatement voisin. Un support de  $6 \times 6 \text{ km}^2$  comme celui de Chimère, quant-à-lui, peut ne pas sembler assez local puisque cela concerne une population plus importante que celle d'un quartier ou d'un village. Nous avons ainsi décidé de nous référer à un support de l'ordre du kilomètre carré, qui nous semble adapté au problème de seuil d'alerte — même si, pour l'instant, les réglementations ne tiennent pas compte de la notion de support spatial.

Le problème que nous examinons comporte donc trois supports : le support (comparativement) ponctuel des stations de mesure Airparif, le support  $6 \times 6 \text{ km}^2$  des cellules Chimère et un support de  $1 \times 1 \text{ km}^2$  qui est l'unité spatiale de référence du seuil de pollution, unité que nous qualifierons de *bloc*.

## 4.4 Modèle de changement de support point-bloc-cellule

Le modèle Gaussien discret, classique en exploration minière, va nous servir à aborder un problème à trois supports tel qu'il est représenté schématiquement sur la **Figure 4.3**, où nous avons des mesures en des points, des valeurs de modèle numérique pour des cellules elles-mêmes partitionnées en des blocs.

On considère que  $\underline{x}$  est un point aléatoire dans un bloc  $v$ . La relation de *Cartier* (Lantuéjoul,

<sup>1</sup>Nous avons d'ailleurs constaté le même type d'oubli dans les réglementations sur l'exposition à des produits chimiques ou au bruit sur le lieu de travail (Wackernagel et al., 1998 ; Vincent et al., 1999) : dans l'application de méthodes statistiques on a tendance à escamoter les problèmes de support et c'est l'application de la géostatistique qui en fait prendre conscience.

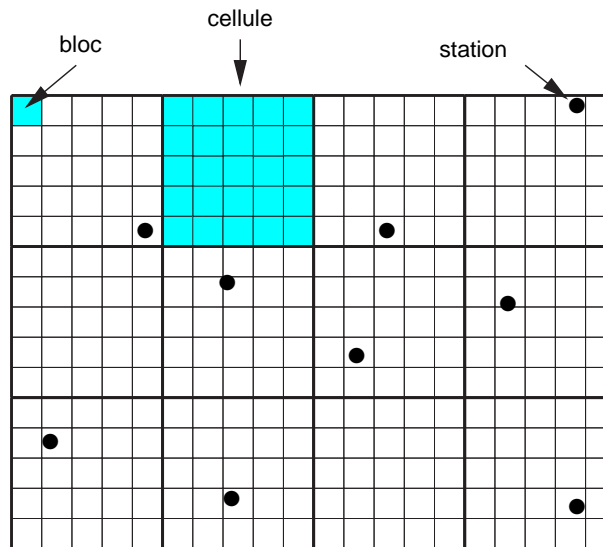


FIG. 4.3 – Trois supports : stations en des *points* à l'intérieur de *cellules* partitionnées en *blocs*.

1988) nous apprend que l'espérance de  $Z(\underline{\mathbf{x}})$  conditionnellement à la valeur du bloc  $v$  est égale à la moyenne de  $Z$  sur le bloc :

$$E[Z(\underline{\mathbf{x}}) \mid Z(v)] = Z(v).$$

Pour une anamorphose Gaussienne,

$$Z(\underline{\mathbf{x}}) = \varphi(Y(\underline{\mathbf{x}})) = \sum_{k=0}^{\infty} \frac{\varphi_k}{k!} H_k(Y(\underline{\mathbf{x}})),$$

où  $H_k(Y(\underline{\mathbf{x}}))$  sont des polynômes d'Hermite, et avec une anamorphose  $\varphi_v(Y(v))$  de blocs correspondante, la relation se réécrit :

$$E(\varphi(Y(\underline{\mathbf{x}})) \mid Z(v)) = E(\varphi(Y(\underline{\mathbf{x}})) \mid \varphi_v(Y(v))) = \varphi_v(Y(v)).$$

L'anamorphose Gaussienne des blocs peut être exprimée avec les coefficients  $\varphi_k$  (ajustés aux données) de l'anamorphose ponctuelle en introduisant un coefficient positif  $r$  de corrélation point-bloc :

$$\varphi_v(Y(v)) = E(\varphi(Y(\underline{\mathbf{x}})) \mid Y(v)) = \sum_{k=0}^{\infty} \frac{\varphi_k}{k!} E(H_k(Y(v)) \mid Y(v)) = \sum_{k=0}^{\infty} \frac{\varphi_k}{k!} r^k H_k(Y(v)),$$

La corrélation point-bloc est déterminée à partir de la variance de dispersion des blocs,

$$\text{var}(Z(v)) = \text{var}(\varphi_v(Y(v))) = \sum_{k=1}^{\infty} \frac{\varphi_k^2}{k!} (r)^{2k},$$

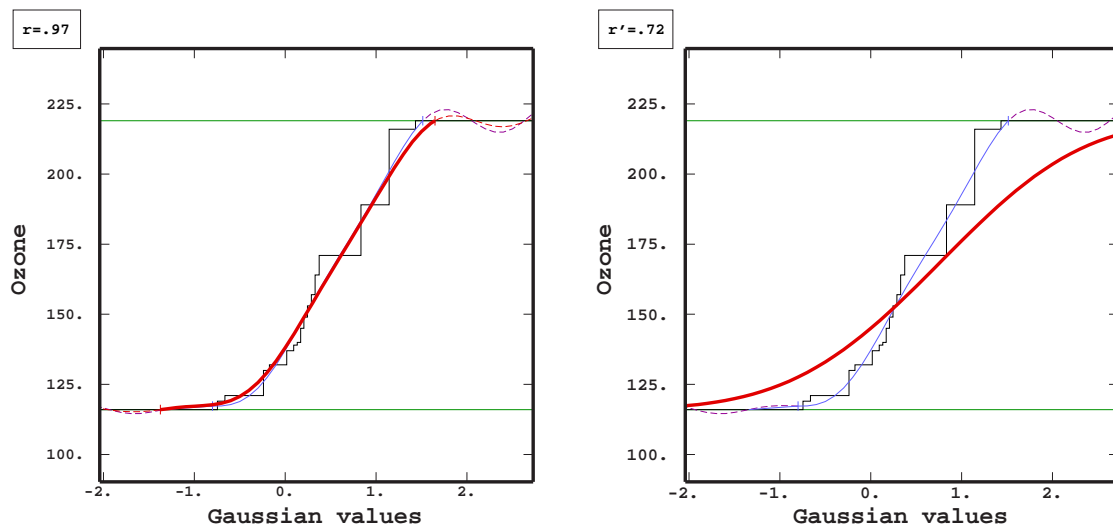


FIG. 4.4 – *A gauche* : anamorphose ponctuelle (en bleu) et anamorphose des blocs (en rouge). *A droite* : anamorphose ponctuelle (en bleu) et anamorphose des cellules (en rouge).

qui est elle-même calculée avec le modèle de variogramme. De la même manière on peut calculer un coefficient  $r'$  de corrélation *point-cellule* et l'on obtient également un coefficient *bloc-cellule* en formant le rapport  $r_{vV} = r'/r$ .

Dans le cas d'étude de l'ozone en Ile-de-France un modèle de variogramme comportant un effet-de-pépite de 50 et modèle cubique de portée 70 km avec un palier de 831 a été ajusté. La moitié des 19 stations étant regroupées au centre du domaine, l'autocorrélation des données biaise les moments et l'on a procédé à une "declusterisation" (Goovaerts, 1997) au moyen d'une fenêtre de  $40 \times 40 \text{ km}^2$  ce qui a conduit à une augmentation de la variance de 881 à 1228, sachant que l'on a également normalisé le palier total du variogramme à cette nouvelle valeur de la variance. L'anamorphose ponctuelle a été tronquée en conservant les 30 premiers polynômes d'Hermite. Le coefficient corrélation point-bloc vaut .97 comparé à la valeur de .72 du coefficient point-cellule. Les courbes (en rouge) des anamorphoses des blocs (à gauche) et des cellules (à droite) sont montrées sur la **Figure 4.4**; on distingue sur ces graphiques également l'anamorphose empirique ponctuelle (en noir) avec son développement en 30 polynômes d'Hermite (en bleu).

Sur la gauche de la **Figure 4.5** on peut voir les histogrammes des blocs (en bleu) et des cellules (en rouge) tels qu'ils sont calculés par le modèle de changement de support. Sur le graphique de droite on a représenté les proportions de blocs (bleu) et de cellules (en rouge) au-dessus de valeurs limites d'ozone. On est frappé, dans l'examen de ces courbes de proportion, par l'ampleur de l'effet de support et les résultats très différents obtenus pour des seuils bas ou élevés.

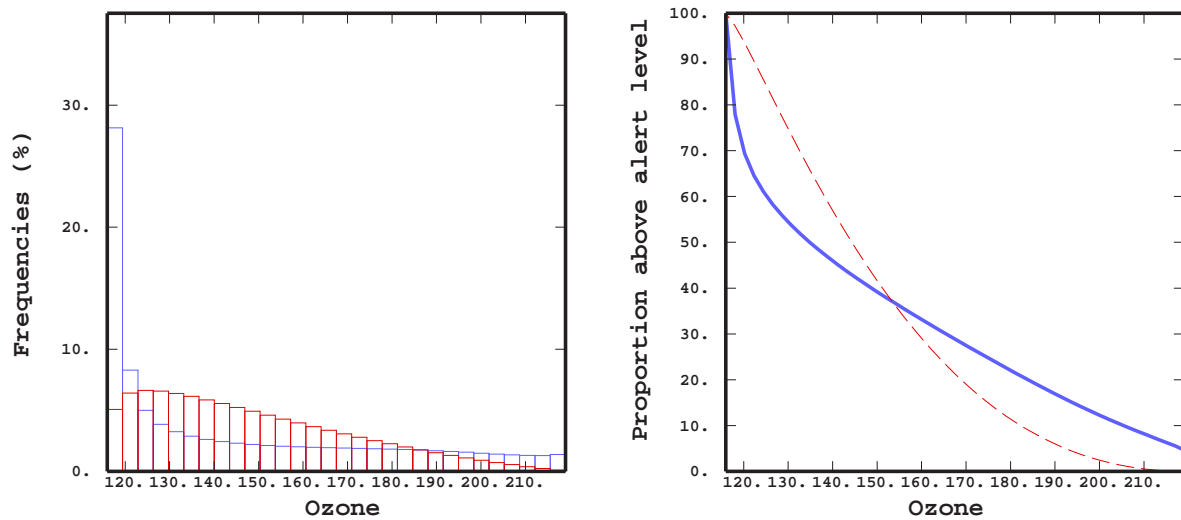


FIG. 4.5 – *A gauche* : histogrammes des blocs (en bleu) et des cellules (en rouge). *A droite* : courbes de proportions de blocs (bleu) et de cellules (en rouge) au-dessus de valeurs limites d’ozone.

## 4.5 Proportion locale de blocs au dessus d’un niveau d’alerte

Le modèle point-bloc-cellule étant pleinement spécifié on peut appliquer des techniques d’estimation de la géostatistique non-linéaire pour calculer des statistiques de blocs localement à l’échelle des cellules. Nous utiliserons le *conditionnement uniforme* (Guibal & Remacre, 1984) qui consiste à estimer l’espérance conditionnelle d’une fonction non-linéaire des blocs par rapport à des valeurs de cellule connues. Par exemple, on pourra calculer la proportion de blocs au-dessus d’un seuil  $z_c$  connaissant la valeur  $z(V_0)$  d’une cellule  $V_0$  donnée :

$$\begin{aligned} E[\mathbf{1}_{Z(\underline{v}) \geq z_c} \mid Z(V_0) = z(V_0)] &= P(Y(\underline{v}) \geq y_c \mid Y(V_0) = y(V_0)) \\ &= 1 - G\left(\frac{y_c - r_{vV} y(V_0)}{\sqrt{1 - r_{vV}^2}}\right), \end{aligned}$$

où la variable conditionnelle  $Y(\underline{v}) \mid y(V_0)$  suit une distribution normale de moyenne  $r_{vV} y(V_0)$  et de variance  $1 - r_{vV}^2$ .  $G$  est la fonction de répartition gaussienne.

Le modèle de changement de support et en particulier les graphiques de la **Figure 4.5** fournissent une appréciation globale sans tenir compte de l’image spatiale détaillée fournie par la prédiction Chimère. Le conditionnement uniforme nous permet de les introduire pour conditionner localement la statistique à estimer, en l’occurrence la proportion de blocs.

Les seuils de pollution pour des moyennes horaires d’ozone varient considérablement dans les réglementations de différents pays du monde (Garcia et al., 2001). Pour ce cas d’école nous avons choisi de considérer le seuil d’alerte suisse de  $120 \mu\text{g}/\text{m}^3$  et le seuil européen de  $180 \mu\text{g}/\text{m}^3$  (directive 92/72/CEE du 21.09.1992). Nous avons appliqué la technique du conditionnement

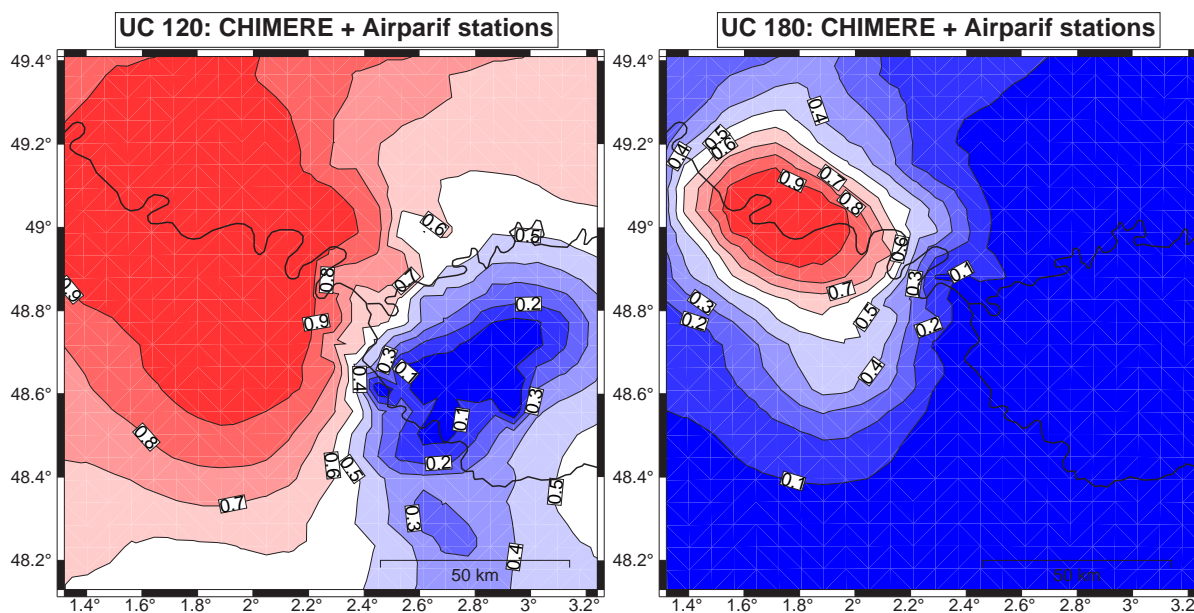


FIG. 4.6 – Proportion de blocs  $1 \times 1 \text{ km}^2$  au-dessus du seuil suisse (à gauche) et européen (à droite) dans des cellules  $6 \times 6 \text{ km}^2$ .

uniforme en conditionnant par les valeurs moyennes de cellules obtenues précédemment par un krigeage utilisant la prédiction Chimère comme dérive externe. Les cartes (interpolées) de proportions estimées de blocs  $1 \times 1 \text{ km}^2$  au-dessus des seuils suisse ou européens dans des cellules  $6 \times 6 \text{ km}^2$  sont montrées sur la **Figure 4.6**. Elles quantifient le risque de dépassement de seuil : sur la base du seuil suisse on alerterait toute la population d’Ile-de-France, tandis qu’avec le seuil européen on pourrait se contenter de ne prévenir que la population au Nord-Ouest de Paris.

## 5 Le krigeage et le filtre de Kalman

Ce chapitre montre d’abord que l’étape de correction du filtre de Kalman est un krigeage simple et s’intéresse ensuite aux améliorations que l’interprétation géostatistique permet d’apporter au niveau de l’erreur de modélisation du système par des équations déterministes. Il s’agit d’un résumé partiel d’un article (Sénégas et al., 2001) que l’on trouvera dans l’Annexe E.

### 5.1 Assimilation de données

L’assimilation de données est utilisée dans le cadre de systèmes de prévision opérationnels qui sont constitués :

- d’un réseau de stations automatiques fournissant des mesures fréquentes dans le temps,

- de modèles numériques déterministes décrivant la physique, la chimie ou l'écologie sur la base d'équations aux dérivées partielles. Ces modèles donnent une description de la dynamique temporelle d'un système et permettent une description spatiale fine,
- d'une procédure d'assimilation de données.

Nous détaillons une première méthode basée sur le filtre de Kalman et qui donne plusieurs points d'entrée aux techniques géostatistiques.

## 5.2 Filtre de Kalman

Ce filtre est dû à R.E. Kalman (1960), qui a proposé une solution récursive du problème de filtrage linéaire. Le filtre de Kalman a été introduit en océanographie dans les années '80 et '90 comme méthode d'assimilation de données en tant qu'alternative *séquentielle* aux méthodes dites *variationnelles*.

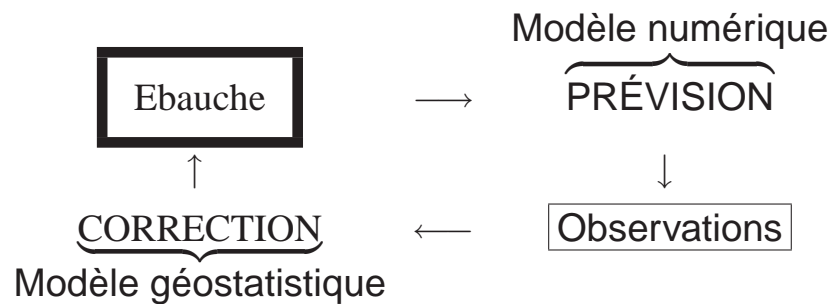


FIG. 5.1 – Assimilation de données : cycle de prévision / correction

Le filtre de Kalman est un algorithme récursif et le schéma sur la **Figure 5.1** montre le déroulement d'un cycle d'assimilation de données. A partir d'une ébauche décrivant l'état du système à un instant donné, le modèle numérique établit une prévision pour un instant suivant. Cette prévision est comparée aux données mesurées à cet instant-là par les capteurs automatiques et donne lieu à une correction de l'état du système dans le cadre d'un modèle géostatistique. L'état corrigé sert alors d'ébauche pour réaliser la prochaine prévision.

### Etape de correction

On souhaite estimer l'état  $\mathbf{y}$  du système sur une grille de  $m$  points en un instant donné à partir de  $n$  observations  $\mathbf{z}^o$  et de  $m$  prévisions  $\mathbf{y}^f$ . Les prévisions sont entachées d'erreurs  $\mathbf{u}^f$ ,

$$\mathbf{y}^f = \mathbf{y} + \mathbf{u}^f,$$

tandis que les observations sont perturbées par des erreurs  $\mathbf{u}^o$ ,

$$\mathbf{z}^o = \mathbf{H}\mathbf{y} + \mathbf{u}^o.$$

et elles sont reliées à l'état par une matrice  $\mathbf{H}$  de taille  $n \times m$ . Les erreurs sont supposées de moyennes nulles,

$$E[\mathbf{u}^f] = 0 \quad E[\mathbf{u}^o] = 0,$$

et leurs matrices de covariance s'écrivent :

$$E[\mathbf{u}^f(\mathbf{u}^f)^\top] = \mathbf{C}^f \quad E[\mathbf{u}^o(\mathbf{u}^o)^\top] = \mathbf{C}^o.$$

Etant donné que les erreurs de prévision et d'observation ne sont pas de même nature, on peut les supposer non-corrélées :

$$E[\mathbf{u}^o(\mathbf{H}\mathbf{u}^f)^\top] = 0.$$

On forme l'estimateur de krigeage simple pour l'ensemble de la grille :

$$\mathbf{u}^* = \mathbf{W}(\mathbf{H}\mathbf{y}^f - \mathbf{z}^o)$$

avec une matrice de pondérateurs  $\mathbf{W}$  de taille  $m \times n$ . Le vecteur  $\mathbf{H}\mathbf{y}^f - \mathbf{z}^o$  contient les *innovations*. Pour déterminer les pondérateurs on calcule la variance d'estimation :

$$\text{var}(U_x^* - U_x^f) = \sum_{\alpha\beta} w_{\alpha x} C_{\alpha\beta} w_{\beta x} - 2 \sum_{\alpha} w_{\alpha x} C_{\alpha x} + C_{xx}.$$

Dans cette expression les covariances entre les points d'observation sont

$$\begin{aligned} C_{\alpha\beta} &= \text{cov}(U_\alpha^f - U_\alpha^o, U_\beta^f - U_\beta^o) \\ &= (\mathbf{H}\mathbf{C}^f\mathbf{H}^\top + \mathbf{C}^o)_{\alpha\beta} \end{aligned}$$

tandis que les covariances avec les noeuds de la grille s'expriment

$$\begin{aligned} C_{\alpha x} &= \text{cov}(U_\alpha^f - U_\alpha^o, U_x^f) \\ &= (\mathbf{H}\mathbf{C}^f)_{\alpha x} \end{aligned}$$

et aux noeuds de la grille on a

$$\begin{aligned} C_{xx} &= \text{var}(U_x^f) \\ &= \mathbf{C}_{xx}^f. \end{aligned}$$

La minimisation de la variance d'estimation donne des équations de krigeage simple,

$$(\mathbf{C}_{\alpha\beta})(\mathbf{W}_{\beta x}) = (\mathbf{C}_{\alpha x}),$$

et l'erreur de krigeage vaut

$$\mathbf{u}^* = (\mathbf{C}_{\alpha x})^\top (\mathbf{C}_{\alpha\beta})^{-1} (\mathbf{H}\mathbf{y}^f - \mathbf{z}^o).$$

En fin de compte, la correction de l'état s'écrit :

$$\begin{aligned} \mathbf{y}^* &= \mathbf{y}^f - \mathbf{u}^* \\ &= \mathbf{y}^f + \underbrace{\mathbf{C}^f \mathbf{H}^\top (\mathbf{H} \mathbf{C}^f \mathbf{H}^\top + \mathbf{C}^o)^{-1}}_{\mathbf{K}} (\mathbf{z}^o - \mathbf{H} \mathbf{y}^f) \end{aligned}$$

où la matrice  $\mathbf{K}$  multipliant les innovations est le *gain de Kalman*. Celui-ci permet de calculer l'erreur associée à l'estimateur  $\mathbf{y}^*$  à partir de l'erreur de prévision,

$$\mathbf{C}_t^* = (\mathbf{I}_m - \mathbf{K}_t \mathbf{H}) \mathbf{C}^f.$$

Naturellement, à l'étape initiale, l'erreur de prévision prend une valeur arbitraire, qui est cependant actualisée à chaque itération.

## Etape de prévision

On affecte l'indice  $t-1$  à l'ébauche  $\mathbf{y}^*$  résultant de l'étape de correction. Celle-ci est utilisée pour construire une nouvelle prévision à l'instant  $t$ ,

$$\mathbf{y}_t^f = \mathcal{M}_t(\mathbf{y}_{t-1}^*, \mathbf{b}_t),$$

au moyen du modèle déterministe  $\mathcal{M}_t$  avec des conditions aux limites  $\mathbf{b}_t$ . Pour la propagation de l'erreur on se sert de l'opérateur linéaire tangent  $\mathbf{F}_t$  autour de l'état courant et cela donne l'expression suivante de l'erreur de prévision :

$$\mathbf{C}_t^f = \mathbf{F}_t \mathbf{C}_{t-1}^* \mathbf{F}_t^\top + \mathbf{Q}_t,$$

où  $\mathbf{Q}_t$  est la matrice de covariance du bruit du système  $\mathbf{q}_t$ .

## Equations du filtre de Kalman

Le système d'équations du filtre de Kalman, tel qu'il a été détaillé plus haut, est constitué de cinq équations,

$$\mathbf{y}_t^f = \mathcal{M}_t(\mathbf{y}_{t-1}^*, \mathbf{b}_t) \quad (5.1)$$

$$\mathbf{C}_t^f = \mathbf{F}_t \mathbf{C}_{t-1}^* \mathbf{F}_t^\top + \mathbf{Q}_t \quad (5.2)$$

$$\mathbf{K}_t = \mathbf{C}_t^f \mathbf{H}^\top (\mathbf{H} \mathbf{C}_t^f \mathbf{H}^\top + \mathbf{C}^o)^{-1} \quad (5.3)$$

$$\mathbf{y}_t^* = \mathbf{y}_t^f + \mathbf{K}_t (\mathbf{z}_t^o - \mathbf{H} \mathbf{y}_t^f) \quad (5.4)$$

$$\mathbf{C}_t^* = (\mathbf{I}_m - \mathbf{K}_t \mathbf{H}) \mathbf{C}_t^f \quad (5.5)$$

dont les deux premières correspondent à la propagation de l'instant  $t-1$  vers  $t$ , tandis que les trois dernières représentent la correction à l'instant  $t$ .

### 5.3 Covariances des erreurs

On peut raisonnablement supposer que les erreurs d'observation se réduisent à des erreurs de mesure et ne sont donc pas autocorrélées. La valeur de l'effet de pépète des erreurs de mesure servira à les calibrer.

Le bruit  $\mathbf{q}_t$  du système, quant à lui, peut être réduit à celui  $\tilde{\mathbf{q}}_t$  des quelques variables les plus sensibles du système, à savoir les plus importantes parmi :

- les conditions limites,
- les paramètres du modèle,
- les paramètres d'implémentation (discrétisation).

On introduit l'opérateur  $\mathbf{G}_t$ , obtenu par un développement limité au premier ordre, qui relie l'erreur  $\tilde{\mathbf{q}}_t$  des conditions limites au bruit du système,

$$\mathbf{q}_t = \mathbf{G}_t \tilde{\mathbf{q}}_t,$$

et la matrice de covariance correspondante est alors :

$$\mathbf{Q}_t = \mathbf{G}_t \tilde{\mathbf{Q}}_t \mathbf{G}_t^\top.$$

Différents auteurs, dont Cañizares (1999), qui se réfère explicitement à la géostatistique, ont tenu compte de l'autocorrélation spatiale de l'erreur sur les variables du système et l'ont modélisée avec des fonctions de covariance.

Dans une application hydrodynamique, cf Wolf et al. (2001) reproduit en Annexe F, le faible nombre de stations automatiques ne permettait pas de calibrer la fonction de covariance de l'erreur liée aux conditions aux limites pour la hauteur d'eau. En s'appuyant sur l'équation d'onde plane un modèle spatio-temporel de variogramme a pu être établi, qui permettait de déduire le variogramme spatial à partir du variogramme temporel.

Dans Sénégas et al. (2001) une autre approche est proposée et le modèle dynamique est utilisé pour calculer la covariance du bruit du système. Soit  $v$  une variable du système,  $\mathbf{y}_{t+1} = \mathcal{M}(\mathbf{y}_t, v)$ , et  $q_v$  l'incertitude associée, de sorte que l'on peut écrire

$$\mathcal{M}(\mathbf{y}_t, v + q_v) \approx \mathcal{M}(\mathbf{y}_t, v) + q_v \frac{\partial \mathcal{M}}{\partial q_v}(\mathbf{y}_t, v).$$

L'erreur du modèle peut être calculée en approximant la dérivée de  $\mathcal{M}$  selon  $q_v$  par rapport à l'état  $\mathbf{y}_t$ ,

$$q_t = \frac{1}{\varepsilon} (\mathcal{M}(\mathbf{y}_t, v + \varepsilon q_v) - \mathcal{M}(\mathbf{y}_t, v)) \quad \text{avec} \quad |\varepsilon| \ll 1.$$

Une combinaison des deux approches permet de décomposer la covariance de l'erreur de modélisation en deux termes :

$$\mathbf{Q}_t = \mathbf{G}_t \tilde{\mathbf{Q}}_t \mathbf{G}_t^\top + \mathbf{Q}_t^{\text{stat}}.$$

Le premier terme décrit la covariance du bruit associé aux paramètres physiques, tandis que le second terme prend en compte le bruit qui n'est pas directement lié à la physique du système. Une

étude détaillée de ces effets en insistant sur l'interprétation géostatistique de l'analyse structurale a été effectuée avec un modèle hydrodynamique à 1D dans Sénégas et al. (2001), cf Annexe E.

## 6 L'assimilation séquentielle et la géostatistique

Nous présentons deux méthodes d'assimilation séquentielle utilisées en océanographie, où le vecteur d'état est de très grande dimension, vu que l'on y étudie en 3D des domaines de grande taille avec une discrétisation souvent aussi fine que le permet la capacité des calculateurs actuels. Ces méthodes visent à produire des algorithmes moins coûteux que le filtre de Kalman étendu qui est également mentionné. Les aspects géostatistiques sont examinés par la suite.

Dans le projet PIONEER nous avons utilisé ces méthodes pour la modélisation écologique de l'estuaire de l'Oder, cf Bertino et al. (2002) en Annexe G. Ce chapitre se réfère essentiellement à l'article Bertino et al. (2003) reproduit en Annexe H et l'on pourra trouver de plus amples développements dans la thèse de Laurent Bertino (2001).

### 6.1 Méthodes d'assimilation

Les méthodes d'assimilation de données se décomposent en deux grandes familles, à savoir les méthodes variationnelles et les méthodes séquentielles. Les méthodes *variationnelles* sont issues de la théorie du contrôle optimal et les plus connues sont la méthode de l'adjoint (Le Dimet & Talagrand, 1986) et celle des représentants (Bennett, 1992).

Les méthodes *séquentielles* sont basées sur des algorithmes récursifs. Du point de vue statistique, on les décrit avantageusement dans le cadre de modèles à espace d'état et plus particulièrement comme une chaîne de Markov à variable cachée. On y suppose que la séquence des états est markovienne, à savoir que l'état  $\mathbf{y}_{t+1}$  dépend uniquement de l'état précédent  $\mathbf{y}_t$ . Ce sont les méthodes d'assimilation séquentielle que nous examinons plus particulièrement dans ce document, sur la base du modèle non-linéaire,

$$\begin{aligned}\mathbf{y}_t &= \mathcal{M}_{t-1}(\mathbf{y}_{t-1}, \mathbf{q}_t) \\ \mathbf{z}_t &= \mathbf{H} \mathbf{y}_t + \mathbf{u}_t^o.\end{aligned}\tag{6.1}$$

où  $\mathbf{q}_t$  est le vecteur des erreurs sur les paramètres du modèle dynamique.

Dans le cadre du *filtre de Kalman étendu* (Jazwinski, 1970) on procède à un développement limité du modèle numérique à chaque pas de temps, dont on se sert pour propager la matrice de covariance des erreurs. Soit  $\partial \mathcal{M}_{t-1} / \partial \mathbf{y} | \mathbf{y}_{n-1}^*$  le modèle linéaire tangent au point  $\mathbf{y}_{n-1}^*$  et l'on considère de même  $\partial \mathcal{M}_{t-1} / \partial \mathbf{q} | \mathbf{y}_{n-1}^*$ . Le développement au premier ordre donne les équations de prévision du filtre de Kalman étendu,

$$\mathbf{y}_t^f = \mathcal{M}_{t-1}(\mathbf{y}_{t-1}^*, 0)$$

$$\mathbf{C}_t^f = \left( \frac{\partial \mathcal{M}}{\partial \mathbf{y}} \Big|_{\mathbf{y}_{t-1}^*} \right) \mathbf{C}_{t-1}^* \left( \frac{\partial \mathcal{M}}{\partial \mathbf{y}} \Big|_{\mathbf{y}_{t-1}^*} \right)^\top + \left( \frac{\partial \mathcal{M}}{\partial \mathbf{q}} \Big|_{\mathbf{y}_{t-1}^*} \right) \mathbf{Q}_t \left( \frac{\partial \mathcal{M}}{\partial \mathbf{q}} \Big|_{\mathbf{y}_{t-1}^*} \right)^\top. \quad (6.2)$$

Vu la non-linéarité de la dynamique  $\mathcal{M}$  ce filtre donne lieu à un biais non négligeable dans l'estimation de l'état. La solution au problème du biais sera d'adopter une approche de Monte-Carlo.

Le biais pourrait certes être amoindri en étendant le développement au second ordre, mais dans ce cas la complexité du filtre devient telle, qu'il n'est plus intéressant pour les applications. De surcroît, les tentatives allant dans ce sens ont buté sur une explosion de l'erreur d'estimation pour les régimes instables du modèle numérique (Evensen, 1992).

## 6.2 RRSQRT : Réduction de rang du filtre

Les applications du filtre de Kalman avec des modèle numériques discrétisés sur un grand nombre  $m$  de noeuds se heurtent très vite au problème de la très grande dimension du vecteur d'état et en conséquence de la matrice  $\mathbf{C}_t^f$  dans l'équation (5.2). Cette matrice intervient dans le calcul du gain de Kalman (5.3), puis dans celui la covariance  $\mathbf{C}_t^*$  de l'estimateur dans (5.5). On a recours à des filtres suboptimaux réduisant le rang de ces matrices comme par exemple le *Reduced Rank Square Root Filter* (RRSQRT) de Martin Verlaan (Verlaan & Heemink, 1997 ; Verlaan, 1998). Dans cet algorithme on se sert de la "racine carrée"  $\mathbf{S}$  obtenue par décomposition en valeurs et vecteurs propres d'une matrice d'erreur  $\mathbf{C}$ , c'est-à-dire  $\mathbf{C} = \mathbf{S}\mathbf{S}^\top$ . La matrice  $\mathbf{S}$  dans la représentation en "racine carrée" de l'équation (6.2) peut être écrite comme la juxtaposition de deux matrices,

$$\mathbf{S}_t^f = \left[ \left( \frac{\partial \mathcal{M}}{\partial \mathbf{y}} \Big|_{\mathbf{y}_{t-1}^*} \right) \mathbf{S}_{t-1}^*, \left( \frac{\partial \mathcal{M}}{\partial \mathbf{q}} \Big|_{\mathbf{y}_{t-1}^*} \right) \mathbf{Q}_t^{1/2} \right]. \quad (6.3)$$

Dans la mise en oeuvre de l'algorithme RRSQRT on ne retient que les  $p$  valeurs propres les plus fortes, ce qui a pour avantage de réduire considérablement le rang du problème, car  $p$  pourra être choisi bien plus petit que  $m$ . Une astuce est de se servir du fait que les valeurs propres de  $\mathbf{S}\mathbf{S}^\top$  sont identiques à celles de  $\mathbf{S}^\top\mathbf{S}$  ce qui permet de les calculer sur une matrice  $n \times n$  au lieu de  $m \times m$ . On se servira donc dans la pratique d'approximations des matrices d'erreur :  $\mathbf{C}^f \approx \mathbf{S}^f \mathbf{S}^{f\top}$  et  $\mathbf{C}^* \approx \mathbf{S}^* \mathbf{S}^{*\top}$ .

## 6.3 EnKF : filtre de Kalman d'Ensemble

Le *filtre de Kalman d'Ensemble* (EnKF) dû à Geir Evensen (Evensen, 1994 ; Burgers et al., 1998) se sert d'une méthode de Monte-Carlo, qui a l'avantage de ne pas nécessiter une linéarisation du modèle numérique de prévision  $\mathcal{M}$ . On génère à chaque pas de temps un ensemble de  $N$  prévisions

$$\left\{ \mathbf{y}_t^{f,i} = \mathcal{M}_{t-1}(\mathbf{y}_{t-1}^{*,i}, \mathbf{q}_t^i); i = 1 \dots N \right\},$$

initialisées à partir d'un ensemble d'états  $\{\mathbf{q}_0^i\}$ . Les réalisations  $\mathbf{y}_t^i$  sont propagées en utilisant des erreurs de modèle  $\mathbf{q}_t^i$  simulées à chaque itération. En termes géostatistiques on peut interpréter cette première étape comme étant une *simulation non-conditionnelle* à l'instant  $t-1$  générant  $N$  réalisations d'une fonction aléatoire non stationnaire pour l'instant  $t$ . La prévision moyenne  $\mathbf{y}_t^f$  et la matrice de covariance  $\mathbf{C}_t^f$  sont calculées directement sur cet ensemble de réalisations.

La seconde étape est alors le *conditionnement* de ces réalisations par krigeage, sur la base de  $n$  observations collectées à l'instant  $t$ ,

$$\left\{ \mathbf{y}_t^{*,i} = \mathbf{y}_t^{f,i} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{H}\mathbf{y}_t^{f,i} + \mathbf{u}_t^{o,i}); i = 1 \dots N \right\},$$

où les erreurs d'observation sont simulées selon une loi gaussienne  $\mathcal{N}(0, \mathbf{C}^o)$ . Les deux premiers moments de cet ensemble de réalisations tendent vers  $\mathbf{y}_t^*$  et  $\mathbf{C}_t^*$ , tout comme la moyenne d'un grand nombre de simulations conditionnelles redonne la solution du krigeage en géostatistique. Les détails d'implémentation algorithmique du filtre de Kalman d'Ensemble sont discutés dans (Evensen, 2003).

## 6.4 Différents apports de la géostatistique

Nous avons vu que l'étape de correction du filtre de Kalman implique un krigeage et que le filtre de Kalman d'Ensemble adopte une démarche comparable à celle des simulations conditionnelles depuis longtemps pratiquée en géostatistique. En particulier, nous avons expliqué en section 5.3 comment la géostatistique a été employée pour ajuster la corrélation spatiale de l'erreur de modélisation par différents auteurs.

### Conditions d'universalité

L'étape de correction des filtres de Kalman présentés dans cette synthèse implique un krigeage simple. Dans sa thèse, Bertino (2001) a examiné la possibilité d'ajouter des conditions d'universalité à ce krigeage afin de filtrer des biais multiplicatifs ou additifs.

Dans les applications d'assimilation séquentielle auxquelles nous avons eu affaire jusqu'à présent, le nombre de stations était cependant insuffisant pour rendre cette généralisation efficace ; par exemple, dans le cas de l'estuaire de l'Oder, il n'y avait que 5 stations automatiques à disposition et l'adjonction de conditions d'universalité détériorait alors les résultats.

### Anamorphose de variables non-gaussiennes

Les méthodes d'assimilation séquentielle proposées font des hypothèses de distribution Gaussienne. Or il se trouve que dans les applications les distributions sont souvent fortement asymétriques. La géostatistique non-linéaire propose des méthodes d'anamorphose, ainsi que des méthodes d'estimation ou de simulation conditionnelle. Dans le cas de l'assimilation séquentielle il faut transformer les variables d'état et les observations, et veiller à ce que cela soit cohérent avec la relation d'observation pour chaque variable. En principe, il faut également s'intéresser au

comportement des lois bivariées spatiales, lorsque l'on effectue des transformations sur les lois marginales, comme dans toute étude de géostatistique non-linéaire. Il faut également regarder les lois bivariées entre les différentes variables, dont l'anamorphose est effectuée séparément, sachant que même en géostatistique non-linéaire on a très peu d'expérience dans ce domaine.

L'anamorphose peut être implémentée de manière particulièrement simple dans le cadre du filtre de Kalman d'Ensemble. Dans Bertino et al. (2003), cf Annexe H, une anamorphose lognormale est appliquée à trois variables (nutriments, phytoplancton, herbivores) dans le contexte d'un modèle écologique simple du cycle annuel d'une colonne d'eau. Ce modèle a été calibré de manière à reproduire les conditions de l'océan au large de la Terre Neuve dans l'Atlantique. L'assimilation avec anamorphose prévoit avec une erreur bien plus faible que le filtre d'ensemble standard sans anamorphose ; en particulier, la floraison printanière, principale non-linéarité au niveau de la dynamique, ne perturbe pas le filtre avec anamorphose, contrairement au filtre standard, qui est responsable d'une succession de "faux départs" de la floraison.

## Modélisation de l'effet de support

Nous avons insisté sur l'effet de support au Chapitre 4 et y avons fourni une illustration du changement de l'allure de la distribution lorsque l'on change de support sur la **Figure 4.5**. Il est important de tenir compte en assimilation de données de la différence de support entre les valeurs mesurées, de support souvent quasi-ponctuel, et les variables d'état, dont le support est celui des cellules du modèle de prévision  $\mathcal{M}$ . Des résultats classiques en géostatistique ont été transposés dans le contexte de l'assimilation de données dans Lajaunie & Wackernagel (2000).

Pour une variable d'état et d'observation Gaussiennes la correction de support se résume à une correction affine des variances, puisque dans l'hypothèse d'une absence de biais le premier moment des observations est identique à celui de la variable d'état de par la relation de Cartier.

Si l'on se place dans le cadre d'un modèle lognormal avec hypothèse de permanence de lognormalité pour les différents supports impliqués, le changement de support est gouverné par un coefficient de changement de support dont il suffit de faire l'inférence.

L'anamorphose Gaussienne généralise l'approche lognormale au sens que l'on transforme une distribution quelconque vers une distribution Gaussienne. On pourra appliquer le modèle Gaussien discret décrit au Chapitre 4 dans le cadre de l'assimilation séquentielle et il suffit de faire l'inférence d'un coefficient de changement de support, si l'on fait l'hypothèse de lois spatiales bi-Gaussiennes pour les variables anamorphosées.

On peut envisager d'autres modélisations de l'effet de support, comme par exemple dans le cadre d'un modèle Gamma, étudié en géostatistique par Hu (1988).

Finalement, on peut aussi renoncer à un modèle explicite de changement de support et avoir recours à des simulations géostatistiques ponctuelles, dont les valeurs spatialement moyennées permettent de passer directement à des valeurs sur un plus grand support. En considérant un ensemble de réalisations des lois conditionnelles empiriques peuvent être alors aisément calculées.

La modélisation du changement de support n'a pas encore été étudiée finement et expérimentée dans des systèmes de prévision opérationnels. Cela est dû à ce que la sensibilisation aux implications de l'effet de support n'a pas encore eu lieu et reste cantonnée aux domaines déjà

investis par les méthodes géostatistiques. Par ailleurs, il faut avouer que la géostatistique non-linéaire nécessite d'être soigneusement adaptée aux applications en assimilation de donnée pour être performante.

## 7 Conclusion

### Les acquis

Dans ce mémoire nous avons cherché à relater quelques premières incursions géostatistiques dans le monde de l'assimilation de données. Si elles ont déjà donné quelques résultats encourageants, on ne peut pas encore parler de percée, car il reste encore de grands thèmes à aborder, tel le changement de support entre observations et variables d'état, qui ont été à peine effleurés.

Avec un peu moins de modestie on peut avancer que la géostatistique a probablement un rôle unifiant, voir unificateur, à jouer. Elle dispose d'un cadre bien défini, la théorie des variables régionalisées avec différentes interprétations probabilistes, et elle a accumulé une riche expérience dans l'élaboration d'outils, tels le krigeage et les simulations conditionnelles, que l'on voit transparaître dans les techniques et algorithmes souvent un peu heuristiques de l'assimilation de données. La géostatistique permet de les situer dans un cadre plus large et de les enrichir de ses concepts.

Nous avons évoqué dans cette synthèse la possibilité d'introduire

- des *conditions d'universalité* : les quelques applications dont nous disposons ne comportaient pas un nombre suffisant de stations pour le faire avec succès ; des résultats encourageants allant dans ce sens ont par contre été obtenus par Blond (2002), Blond et al. (2003),
- l'*anamorphose* de distributions, qui a tout de suite fait ses preuves dans un cas d'étude avec un modèle écologique simplifié,
- le *changement de support*, dont la nécessité de le rendre explicite paraît évidente, ce que nous avons illustré par le cas d'étude du Chapitre 4, mais dont il n'existe pour l'instant pas d'exemple d'implémentation dans une assimilation séquentielle.

### Perspectives

Les impulsions des projets PIONEER et IMPACT se sont concrétisées par un éventail de publications, que nous avons en partie reproduites dans les Annexes, et qui contribuent à consolider une interprétation géostatistique de l'assimilation de données. Les deux méthodes séquentielles que nous avons évoquées sont d'un usage très répandu en océanographie et commencent à être envisagées pour les applications météorologiques<sup>1</sup>.

<sup>1</sup>Voir l'article de Mackenzie (2003), certes un peu provocateur, illustré avec l'exemple de la tempête du 26 décembre 1999 en France.

Ces deux méthodes, le filtre de rang réduit et le filtre d'ensemble, ont en commun de procéder dans les itérations à une propagation non-linéaire (lorsque la dynamique est non-linéaire), puis à une correction linéaire par krigeage. Il existe des modèles plus généraux pour lesquels la correction n'est pas linéaire et qui sont également présentés dans Bertino et al. (2003). Il s'agit d'algorithmes de Monte Carlo donnant lieu à une assimilation séquentielle par *échantillonnage d'importance*, de préférence avec rééchantillonnage (SIR), tels qu'ils sont décrits dans Doucet et al. (2001). Ces algorithmes, et de manière générale les algorithmes de Monte-Carlo basés sur des chaînes de Markov (Gilks et al., 1996), ont provoqué une véritable révolution des statistiques Bayésiennes dans les années '90, leur ouvrant (enfin) la voie des applications. Leur attrait principal dans notre contexte est qu'elles sont affranchies de toute hypothèse Gaussienne. Leur inconvénient est qu'elles nécessitent des calculs très intensifs qui seront réhilitoires dans les applications opérant avec des modèles numériques définis sur des grilles de très grandes taille comme on les rencontre en océanographie et météorologie. Ces méthodes se répandent également en géostatistique et représentent certainement la voie dans laquelle les efforts majeurs sont investis actuellement.

# Références

- Armstrong, M. & P. A. Dowd (Eds.) (1994). *Geostatistical Simulation*. Amsterdam : Kluwer.
- Armstrong, M., A. G. Galli, G. Le Loc'h, F. Geffroy, & R. Eschard (2003). *Plurigaussian Simulations in Geosciences*. Berlin : Springer-Verlag.
- Bear, J., G. de Marsily, & C. F. Tsang (Eds.) (1993). *Flow and contaminant transport in fractured rock*. San Diego : Academic Press.
- Bennett, A. F. (1992). *Inverse Methods in Physical Oceanography*. Cambridge : Cambridge University Press. 27
- Bertino, L. (2001). *Assimilation de Données pour la Prédiction de Paramètres Hydrodynamiques et Ecologiques : Cas de la Lagune de l'Oder*. Doctoral thesis, Ecole des Mines de Paris, Fontainebleau. <http://pastel.paristech.org>. 4
- Bertino, L., G. Evensen, & H. Wackernagel (2002). Combining geostatistics and Kalman filtering for data assimilation in an estuarine system. *Inverse Problems* 18, 1–23. 4, 13
- Bertino, L., G. Evensen, & H. Wackernagel (2003). Sequential data assimilation techniques in oceanography. *International Statistical Review* 71, 223–241. 4, 13
- Bez, N. & J. Rivoirard (2001). Transitive geostatistics to characterise spatial aggregation with diffuse limits : an application on mackerel ichthyoplankton. *Fisheries Research* 50, 41–58. 7
- Biboud, E. (2002). Modélisation des épidémies de grippe. Technical Report S-436, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau. 4
- Blond, N. (2002). *Assimilation de Données Photochimiques et Prévission de la Pollution Troposphérique*. Doctoral thesis, Ecole Polytechnique. 16
- Blond, N., L. Bel, & R. Vautard (2003). Three-dimensional ozone data analysis with an air quality model over Paris area. *J. Geophysical Research* 108(D23), 4744.
- Bui, V. A. P. (2001). Modélisation par chaînes de Markov cachées d'une épidémie de grippe. Technical Report S-420, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau. 4
- Burgers, G., P. J. van Leeuwen, & G. Evensen (1998). On the analysis scheme in the ensemble Kalman filter. *Monthly Weather Review* 126, 1719–1724. 28
- Cañizares, R. (1999). *On the Application of Data Assimilation in Regional Coastal Models*. Ph. D. thesis, TU Delft, Rotterdam.
- Chauvet, P. (1982). The variogram cloud. In T. B. Johnson & R. J. Barnes (Eds.), *17<sup>th</sup> AP-COM*, New York, pp. 757–764. Society of Mining Engineers. 9

- Chilès, J. P. & G. de Marsily (1993). Stochastic models of fracture systems and their use in flow and transport modelling. See Bear, de Marsily, & Tsang (1993), pp. 169–236.
- Chilès, J. P. & P. Delfiner (1999). *Geostatistics : Modeling Spatial Uncertainty*. New York : Wiley. 8, 9, 12, 14
- Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology* 17, 563–586. 10
- Cressie, N. (1990). Reply to letter by G. Wahba. *The American Statistician* 44, 256–258. 13
- Cressie, N. (1993). *Statistics for Spatial Data* (revised ed.). New York : Wiley. 8
- Cressie, N. & C. K. Wikle (1998). The variance-based cross-variogram : you can add apples and oranges. *Mathematical Geology* 30, 789–800. 11
- Davis, M. W. (1987). Production of conditional simulations via the lu triangular decomposition of the covariance matrix. *Mathematical Geology* 19, 91–98. 14
- de Wijs, H. J. (1951). Statistics of ore distribution, part I : frequency distribution of assay values. *Geologie en Mijnbouw* 13, 365–375.
- Deutsch, C. & A. Journel (1992). *GSLIB : Geostatistical Software Library and User's Guide*. New York : Oxford University Press. 14
- Diggle, P. J., J. A. Tawn, & R. A. Moyeed (1998). Model-based geostatistics (with discussion). *Applied Statistics* 47, 299–350. 10
- Doucet, A., N. de Freitas, & N. Gordon (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. New York : Springer-Verlag.
- Evensen, G. (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophysical Research* 97(C11), 17905–17924. 28
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophysical Research* 99(C5), 10143–10162. 28
- Evensen, G. (2003). The Ensemble Kalman filter : Theoretical formulation and practical implementation. *Ocean Dynamics*, in press. 29
- Garcia, J., J. Colosio, & P. Jamet (2001). *Les indices de qualité de l'air : élaboration, usages et comparaisons internationales*. Paris : Les Presses de l'Ecole des Mines. 21
- Gelfand, A. E., A. M. Schmidt, & C. F. Sirmans (2002). Multivariate spatial processes models : Conditional and unconditional bayesian approaches using coregionalization. Technical report, University of Connecticut. 12
- Genton, M. G. (1998). Highly robust variogram estimation. *Mathematical Geology* 30, 213–221.
- Gilks, W. R., S. Richardson, & D. J. Spiegelhalter (Eds.) (1996). *Markov chain Monte Carlo in Practice*. London : Chapman & Hall. 32
- Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. American Statistical Association* 97, 590–600. 13
- Gómez-Hernández, J. & E. Cassiraga (1994). Theory and practice of sequential simulation. See Armstrong & Dowd (1994), pp. 111–124. 14

- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford : Oxford University Press. 20
- Goulard, M. & M. Voltz (1992). Linear coregionalization model : tools for estimation and choice of multivariate variograms. *Mathematical Geology* 24, 269–286. 12
- Grzebyk, M. & H. Wackernagel (1994). Multivariate analysis and spatial/temporal scales : real and complex models. In *Proceedings of XVIIth International Biometrics Conference*, Volume 1, Hamilton, Ontario, pp. 19–33. 12
- Guarascio, M., M. David, & C. Huijbregts (Eds.) (1976). *Advanced Geostatistics in the Mining Industry*, Volume C24 of *NATO ASI Series*. Dordrecht : Reidel.
- Guibal, D. & A. Remacre (1984). Local estimation of the recoverable reserves : comparing various methods with the reality on a porphyry copper deposit. See Verly, David, & Journel (1984), pp. 435–448. 21
- Handcock, M. S. & M. L. Stein (1993). A Bayesian analysis of kriging. *Technometrics* 35, 403–410. 10
- Haslett, J., R. Bradley, P. S. Craig, G. Wills, & A. R. Unwin (1991). Dynamic graphics for exploring spatial data, with application to locating global and local anomalies. *The American Statistician* 45, 234–242. 9
- Hu, L. Y. (1988). *Mise en oeuvre du modèle gamma pour l'estimation des distributions spatiales*. Doctoral thesis, Ecole des Mines de Paris, Fontainebleau.
- Hudson, G. & H. Wackernagel (1994). Mapping temperature using kriging with external drift : theory and an example from Scotland. *International J. Climatology* 14, 77–91. 12
- Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. New York : Academic Press. 27
- Journel, A. G. (1974). Geostatistics for conditional simulation of orebodies. *Economic Geology* 69, 673–687. 14
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.-T. ASME* 82(D), 35–45.
- Krige, D. G. (1951). A statistical approach to some mine valuation and allied problems on the Witwatersrand. *J. Chem. Metal. Min. Soc. South Africa* 52, 119–139.
- Kyriakidis, P. C. & A. G. Journel (1999). Geostatistical space-time models : a review. *Mathematical Geology* 31, 651–684. 4, 13
- Lajaunie, C. & H. Wackernagel (2000). Geostatistical approaches to change of support problems : Theoretical framework. IMPACT Project Deliverable Nr 19, Technical Report N–30/01/G, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- Lajaunie, C., H. Wackernagel, & L. Bertino (2001). Geostatistical normalization : Case studies. IMPACT Project Deliverable Nr 3, Technical Report N-31/01/G, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau. 12, 17
- Lantuéjoul, C. (1988). Some stereological and statistical consequences derived from Cartier's formula. *J. Microscopy* 151, 265–276. 19
- Lantuéjoul, C. (2001). *Geostatistical Simulation : Models and Algorithms*. Berlin : Springer-Verlag.

- Le Dimet, F. X. & O. Talagrand (1986). Variational algorithms for analysis and assimilation of meteorological observations : Theoretical aspects. *Tellus* 38A, 97–110. 27
- Lefèbvre, J., H. Roussel, E. Walter, D. Lecointe, & W. Tabbara (1996). Prediction from wrong models : the kriging approach. *Antennas & Propagation Magazine IEEE* 38, 35–45. 6
- Mackenzie, D. (2003, October). Ensemble Kalman filters bring weather models up to date. *SIAM News* 36(8). <http://www.siam.org/siamnews/10-03/tococt03.htm>.
- Mardia, K. V., C. Goodall, E. Redfern, & F. J. Alonso (1998). The kriged Kalman filter (with discussion). *Test* 7, 217–285. 13
- Mardia, K. V. & R. J. Marshall (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* 71, 135–146. 10
- Matheron, G. (1955). Application des méthodes statistiques à l'évaluation des gisements. *Annales des Mines* 144 (12), 50–75.
- Matheron, G. (1962). *Traité de Géostatistique Appliquée*, Volume 1. Paris : Technip.
- Matheron, G. (1965). *Les Variables Régionalisées et leur Estimation*. Paris : Masson. 8, 9, 11
- Matheron, G. (1970). *La Théorie des Variables Régionalisées et ses Applications*. Number 5 in Les Cahiers du Centre de Morphologie Mathématique. Fontainebleau : Ecole des Mines de Paris. 13
- Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability* 5, 439–468.
- Matheron, G. (1976a). Les concepts de base et l'évolution de la géostatistique minière. See Guarascio, David, & Huijbregts (1976), pp. 3–10.
- Matheron, G. (1976b). A simple substitute for conditional expectation : the disjunctive kriging. See Guarascio, David, & Huijbregts (1976), pp. 221–236. 14
- Matheron, G. (1981). Splines and kriging : their formal equivalence. In D. F. Merriam (Ed.), *Down-to-Earth Statistics : Solutions Looking for Geological Problems*, Volume 8 of *Syracuse University Geology Contribution*, New York, pp. 77–95. Syracuse University. 13
- Matheron, G. (1984). Isofactorial models and change of support. See Verly, David, & Journel (1984), pp. 449–467. 14
- Matheron, G. (1989). *Estimating and Choosing*. Berlin : Springer-Verlag.
- Monestiez, P., D. Allard, & R. Froidevaux (Eds.) (2001). *GeoENV III – Geostatistics for Environmental Applications*. Amsterdam : Kluwer.
- Myers, D. E. (1991). Pseudo-cross variograms, positive-definiteness, and cokriging. *Mathematical Geology* 23, 805–816. 11
- Petitgas, P. (2001). Geostatistics in fisheries survey design and stock assessment : models, variances and applications. *Fish and Fisheries* 2, 231–249. 7
- Rivoirard, J. (1994). *Introduction to Disjunctive Kriging and Non-Linear Geostatistics*. Oxford : Oxford University Press. 14
- Rivoirard, J. (2001). Which models for collocated cokriging ? *Mathematical Geology* 33, 117–131. 12
- Rivoirard, J. (2002). On the structural link between variables in kriging with external drift. *Mathematical Geology* 34, 797–808.

- Royle, J. A. & L. M. Berliner (1999). A hierarchical approach to multivariate spatial modeling and prediction. *J. Agricultural, Biological and Environmental Statistics* 4, 29–56.
- Sacks, J., W. J. Welch, T. J. Mitchell, & H. P. Wynn (1989). Design and analysis of computer experiments. *Statistical Science* 4, 409–435.
- Schmidt, H., C. Derognat, R. Vautard, & M. Beekmann (2001). A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in western Europe. *Atmospheric Environment* 35, 6277–6297. 16
- Séguret, S. & P. Huchon (1990). Trigonometric kriging : a new method for removing the diurnal variation from geomagnetic data. *J. Geophysical Research* 32(B13), 21.383–21.397. 13
- Sénégas, J., H. Wackernagel, W. Rosenthal, & T. Wolf (2001). Error covariance modeling in sequential data assimilation. *Stochastic Environmental Research and Risk Assessment* 15, 65–86. 4, 22
- Stein, M. L. (1999). *Interpolation of Spatial Data : Some Theory for Kriging*. New York : Springer-Verlag. 8, 9
- Vautard, R., M. Beekmann, J. Roux, & D. Gombert (2001). Validation of a deterministic forecasting system for the ozone concentrations over the Paris area. *Atmospheric Environment* 35, 2449–2461. 16
- Ver Hoef, J. M. & R. P. Barry (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *J. Statistical Planning and Inference* 69, 275–294. 12
- Verlaan, M. (1998). *Efficient Kalman filtering algorithms for hydrodynamic models*. Ph. D. thesis, TU Delft, Delft, <http://ta.twi.tudelft.nl/users/verlaan/artikelen/thesis.ps.gz>. 28
- Verlaan, M. & A. W. Heemink (1997). Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics* 11(5), 349–368. 28
- Verly, G., M. David, & A. G. Journel (Eds.) (1984). *Geostatistics for Natural Resources Characterization*, Volume C-122 of *NATO ASI Series*. Dordrecht : Reidel.
- Vincent, R., M. Grzebyk, H. Wackernagel, & C. Lajaunie (1999). Application de la géostatistique à l'hygiène industrielle : Evaluation d'un cas d'exposition professionnelle au trichloroéthylène. *Cahiers de Notes Documentaires* 174(ND 2094-174-99), 5–13. 18
- Wackernagel, H. (2002). Geostatistical normalization of air pollution transport model output and station data using Isatis. IMPACT Project Deliverable Nr 4, Technical Report N–20/02/G, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- Wackernagel, H. (2003). *Multivariate Geostatistics : an Introduction with Applications* (3rd ed.). Berlin : Springer-Verlag. 12, 14
- Wackernagel, H., C. Lajaunie, L. Thiéry, & M. Grzebyk (1998). Evaluation de l'exposition sonore en milieu professionnel : Application de méthodes géostatistiques à l'estimation du Leq et conséquences sur les stratégies de mesure. Technical Report MAV–NT-373/LT, Institut National de Recherche et de Sécurité, Vandoeuvre-Les-Nancy. 18
- Wahba, G. (1990a). Letter to the editor. *The American Statistician* 44, 255–256. 13
- Wahba, G. (1990b). *Spline Models for Observational Data*. Philadelphia : Society for Industrial and Applied Mathematics. 13

- Walter, E. & L. Pronzato (1997). *Identification of Parametric Models from Experimental Data*. Berlin : Springer-Verlag. 6
- Wikle, C. K. & N. Cressie (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika* 86, 815–829. 13
- Wolf, T., J. Sénégas, L. Bertino, & H. Wackernagel (2001). Application of data assimilation to three-dimensional hydrodynamics : the case of the Odra lagoon. See Monestiez, Allard, & Froidevaux (2001), pp. 157–168.
- Yaglom, A. M. (1986). *Correlation Theory of Stationary and Related Random Functions*. Berlin : Springer-Verlag. 9
- Zimmerman, D. L. & M. B. Zimmerman (1991). A comparison of spatial variogram estimators and corresponding ordinary kriging predictors. *Technometrics* 33, 77–91. 10

# **Deuxième partie**

## **Éléments biographiques**

## 8.1 Curriculum vitae

**Nom :** Wackernagel

**Prénoms :** Hans Heinrich

**Date et lieu de naissance :** 17 Mai 1955 à Genève, Suisse

**Nationalités :** allemande et suisse

**Adresse professionnelle :** Ecole des Mines de Paris, Centre de Géostatistique, 35 rue Saint Honoré, 77305 Fontainebleau, Tél : 01 64 69 47 60, Fax : 01 64 69 47 05, Courriel : hans.wackernagel@ensmp.fr

**Situation de famille :** marié, 4 enfants

**Formation universitaire :**

- *Vordiplom in Geologie-Paleontologie*, Freie Universität, Berlin, Allemagne, 1979
- *Diplôme de Formation Spécialisée en Géostatistique*, Ecole des Mines de Paris, 1981
- *Doctorat de 3e Cycle en Sciences et Techniques Minières*, Ecole des Mines de Paris, 1985

**Carrière scientifique et technique :**

**Tuteur** (*studentische Hilfskraft*), Freie Universität, Berlin, Septembre 1979 à Août 1980

**Chercheur** (*wissenschaftlicher Mitarbeiter*, Deutsche Forschungsgemeinschaft), Freie Universität, Berlin, Septembre 1981 à Juin 1982

**Ingénieur** chez BEB Erdgas & Erdöl, Hannover, Juillet à Septembre 1982

**Enseignant-Chercheur** (*Ingénieur de Recherche*, ARMINES), Ecole des Mines de Paris, depuis Août 1983

**Reponsable administratif** du DEA et Cycle de Formation Spécialisée en Géostatistique, Septembre 1987 à Juin 1990

## 8.2 Evaluation et direction de travaux scientifiques

**Participation à des comités d'édition de revues scientifiques**

- *Editeur associé* de la revue *Metron* (depuis 2003).
- *Editeur associé* de la revue *Mathematical Geology* (de 1995 à 1997).
- *Conseiller éditorial* de la revue *Environmental and ecological statistics* (depuis 2003).
- *Conseiller éditorial* de la revue *Computers & Geosciences* (depuis 1989).

**Direction de thèses de Doctorat**

- Laurent Bertino (2001) Assimilation de données pour la prédiction de paramètres hydrodynamiques et écologiques : cas de la Lagune de l'Oder. Doctorat de Géostatistique, Ecole des Mines de Paris.

- Michel Grzebyk (1993) Ajustement d'une corégionalisation stationnaire. Doctorat de Géostatistique, Ecole des Mines de Paris.

### **Participation à des jurys de thèse de Doctorat**

- Nadège Blond (2002) Assimilation de données photochimiques et prévision de la pollution troposphérique. Ecole Polytechnique (*Examineur*).
- Laurent Bertino (2001) Assimilation de données pour la prédiction de paramètres hydrodynamiques et écologiques : cas de la Lagune de l'Oder. Ecole des Mines de Paris (*Examineur*).
- Véronique Rannou (2001) Apport des probabilités et des statistiques à la prédiction des perturbations induites par un téléphone portable sur des câbles. Supélec/Université de Paris VI (*Rapporteur*).
- Gérard Biau (2000) Méthodes itératives en estimation fonctionnelle et systèmes dynamiques. Université Montpellier II (*Examineur*).
- Stéphanie Deshayes (2000) Prédiction et analyse des perturbations subies par un dispositif électronique placé dans un environnement électromagnétique sévère. Application à l'immunité des systèmes de communication. Supélec/Université de Paris VI (*Rapporteur*).
- Jasha Oosterbaan-Eritzpokhoff (2000) Utilisation des méthodes d'analyse exploratoire de données environnementales pour l'établissement de descripteurs macroscopiques de la dynamique des termes-sources polluants dans la géosphère. Ecole des Mines de Paris (*Examineur*).
- Gilles Guillot (1998) Modélisation Statistique des Champs de Pluie Sahéliens. Application à leur désagrégation spatiale et temporelle. Université Joseph Fourier - Grenoble I (*Rapporteur*).
- Jean Lefèbvre (1997) Apport des statistiques aux études de phénomènes de couplage en compatibilité électromagnétique. Supélec/Université de Paris VI (*Rapporteur*).
- Hocine Bourenane (1997) Etude des lois de distribution spatiale des sols de Petite Beauce. INRA/Université d'Orléans (*Examineur*).
- Michel Grzebyk (1993) Ajustement d'une corégionalisation stationnaire. Ecole des Mines de Paris (*Examineur*).
- Sylvain N'Kanza (1993) Géostatistique et statistique spatiale : applications à l'analyse structurelle des images et à l'optimisation d'un plan de sondage agricole en télédétection satellitaire. Ecole des Hautes Etudes en Sciences Sociales (*Rapporteur*).
- Pierre Goovaerts (1992) Multivariate geostatistical tools for studying scale-dependent correlation structures and describing space-time variations. Université Catholique de Louvain (*Examineur*).

### **Encadrement de mémoires de DEA et de Magistère**

- Guillot G (1994/95) Eléments pour l'estimation des lignes de grain d'Afrique de l'Ouest aux faibles pas de temps (ORSTOM, Grenoble). DEA de Géostatistique (INPL, Nancy).

- Métivier F (1990) Introduction à la géostatistique ; application à des données géochimiques (Příbram, Tchécoslovaquie). Magistère de Sciences de la Terre (Ecole Normale Supérieure, Paris).
- Soulaïmani M (1987/88) Simulation conditionnelle du taux de carbone organique sur une coupe en roche mère dans le Yorkshire (Institut Français du Pétrole, Rueil-Malmaison). DEA de Géostatistique (INPL, Nancy).
- Worm R (1985/86) Etude en géostatistique multivariée d'un réservoir de gaz naturel (Texaco, Hamburg). DEA de Géostatistique (INPL, Nancy).
- Jaquet O (1984/85) Etude géostatistique multivariée d'un champ de gaz (BEB Erdgas & Erdöl, Hannover). DEA de Géostatistique (INPL, Nancy).

### **Encadrement de mémoires de 3e année de l'Ecole des Mines de Paris**

- Hollard A (2002/03) How to improve the EnKF using an anamorphosis operation (NERSC, Bergen).
- Kanaev T (2002/03) Prévision de l'évolution temporelle de paramètres limnologiques du Lac de Constance (Institut für Seenforschung, Langenargen).
- Esposito E (2001/02) Statistical investigation about the radar backscatter intensity of sea clutter (Oceanwaves GmbH, Lüneburg).
- Inizan M (2001/02) Validation géostatistique des résultats d'un modèle écologique à l'aide de données de mesures (NERSC, Bergen).
- Frendo O (2000/01) Climate/fish stock interactions : observing, modeling, assessing (NERSC, Bergen).
- Breuillin C (1999/00) Data assimilation methods applied to marine ecosystem models (NERSC, Bergen).
- Sénégas J (1998/99) Hydrodynamical modeling and data assimilation within the Odra estuary (GKSS, Geesthacht).
- Royer C (1998/99) Krigage avec carte thématique et classification spatiale (INRA, Avignon).
- Balabdaoui F (1998/99) L'évolution spatio-temporelle du taux de fécondité en Inde (CNRS-Musée de l'Homme, Paris).
- Bertino L (1997/98) Short-term prediction of nutrient load in the Elbe river (GKSS, Geesthacht).
- Venard C (1997/98) Multivariate and spatial analysis of chemical elements of the river Elbe (GKSS, Geesthacht).
- Biau G (1996/97) Estimation of precipitation by analog kriging from the empirical orthogonal functions of a pressure field (GKSS, Geesthacht).

## Encadrement de mémoires du Cycle de Formation Spécialisée en Géostatistique

- Guillaume B (2001/2) Analyse exploratoire des données de concentrations horaires d ozone en région Ile-de-France : comportement temporel et méthodes géostatistiques.
- Yordanov M (1994/95) Geostatistical study of contaminated land.
- Boykova A (1993/94) Analyse exploratoire des variables Cd, Mo et pH pour des données de pollution.
- Deme A (1993/94) Comparaison géostatistique de données de forage, de trous de tirs et de production.
- Lindner S (1991/92) Etude géostatistique multivariable de données radiométriques et géochimiques (CNRS/Institut de Géologie, Strasbourg).
- Hudson G (1990/91) Kriging the temperature in Scotland using the external drift method (interpolation of climate for land evaluation).
- Baltaci R (1989/90) Estimation linéaire d'une indicatrice de l'arsenic en prospection géochimique multi-éléments.
- N'Kanza S (1987/88) Comparaison de trois méthodes statistiques appliquées à des images de microsonde. Mastère de Géostatistique (EMP).
- Truong XL (1987/88) Variogrammes et krigeages sur un gisement d'or du Colorado.
- Murthy PSN (1986/87) Comparaison entre krigeage et cokrigeage, et optimisation de carrière sur un gisement de fer en Inde (NMDC, Hyderabad).
- Butenuth C (1986/87) Analyse d'un jeu de données géochimiques multiéléments du Harz.

## 8.3 Participation à des congrès

### Organisation de sessions (depuis 2001)

- Session *Statistiques Spatiales, Géométrie Stochastique et Traitement d'Images* pour les Journées de Stochastique (19-22 mars 2002, Magdebourg, Allemagne) de la section "Stochastique" de la Société de Mathématiques Allemande.
- Session *Géostatistique Multivariable et Assimilation de Données* pour la Conférence 2002 (15-20 septembre 2002, Berlin, Allemagne) de Association Internationale de Géologie Mathématique.
- Session *Statistiques, Modèles et Assimilation de Données* pour la 8e Conférence Internationale de Climatologie Statistique (12-16 mars 2001, Lüneburg, Allemagne).

### Conférences invitées (2003)

- Conférence invitée sur la *Géostatistique non-linéaire pour les applications environnementales* lors de 6. Conférence SPRUCE des Statistiques pour l'Environnement (15-19 juin 2003, Lund, Suède).
- Conférence invitée sur l'*Assimilation séquentielle de données en océanographie* lors de la

- Conférence Internationale de Statistiques pour l'Environnement et la Santé (16-18 juillet 2003, Saint Jacques de Compostelle, Espagne) de l'Institut International de Statistiques.
- Conférence invitée sur les *Modèles de la géostatistique et le krigeage* lors du 13e Symposium sur l'Identification des Systèmes (27-29 août 2003, Rotterdam, Pays-Bas) de la Fédération Internationale de Contrôle Automatique.

## 8.4 Publications

### Monographie

- Wackernagel H (1995) *Multivariate Geostatistics : an Introduction with Applications*. Springer-Verlag, Berlin, 256p (2. édition 1998, 291p ; 3. édition 2003, 387p).  
Traduction Japonaise de la 2. édition : Morikita Shuppan, Tokyo, 2003, 266p.

### Revues

- Bertino L & Wackernagel H (2004) Prévoir l'état de l'océan. *La Recherche*, 371, 98–99.
- Bertino L, Evensen G & Wackernagel H (2003) Sequential Data Assimilation techniques in oceanography. *International Statistical Review*, 71, 223–241.
- Bertino L, Evensen G & Wackernagel H (2002) Combining geostatistics and Kalman filtering for data assimilation in an estuarine system. *Inverse Problems*, 18, 1–23.
- Özer Ü, Thévenin I, Wackernagel H, Mesut A & Yörükoğlu M (2001) Etude géostatistique multivariable de la qualité du lignite d'Afsin-Elbistan, projet de Çöllolar (Turquie). *Bulletin of Engineering Geology and the Environment*, 60, 241–251.
- Sénégas J, Wackernagel H, Rosenthal W & Wolf T (2001) Error covariance modeling in sequential data assimilation. *Stochastic Environmental Research and Risk Assessment*, 15, 65–86.
- Biau G, Zorita E, von Storch H & Wackernagel H (1999) Estimation of precipitation by kriging in the EOF space of the sea level pressure field. *Journal of Climate*, 12, 1070-1085.
- Wackernagel H (1994) Cokriging versus kriging in regionalized multivariate data analysis. *Geoderma*, 62, 83–92.
- Hudson G & Wackernagel H (1994) Mapping temperature using kriging with external drift : theory and an example from Scotland. *International Journal of Climatology*, 14, 77–91.
- Rouhani S & Wackernagel H (1990) Multivariate geostatistical approach to space-time data analysis. *Water Resources Research*, 26, 585–591.
- Wackernagel H (1989) Description of a computer program for analysing multivariate spatially distributed data. *Computers & Geosciences*, 15, 593–598.
- Wackernagel H & Butenuth C (1989) Caractérisation d'anomalies géochimiques par la géostatistique multivariable. *Journal of Geochemical Exploration*, 32, 437–444.

## Ouvrages collectifs

- Wackernagel H, Bertino L, Sierra JP & Gonzalez del Rio J (2002) Multivariate kriging for interpolating data from different sources. In : Anderson CW, Barnett V, Chatwin P & El Shaarawi AH (eds) *Quantitative Methods for Current Environmental Issues*, 57-75, Springer-Verlag, London.
- Wackernagel H (2001) Multivariate geostatistics. In : A. El-Shaarawi, W. Piegorisch (eds) *Encyclopedia of Environmetrics*, vol. 3, 1344-1347, Wiley, Chichester.
- Grimvall A, Wackernagel H & Lajaunie C (2001) Normalisation of environmental quality data. In : L.M. Hilty, P.W. Gilgen (eds) *Sustainability in the Information Society*, 583-590, Metropolis-Verlag, Marburg.
- Venard C, Aulinger A, Lajaunie C & Wackernagel H (2001) Geostatistical evaluation of a monitoring strategy for water quality assessment. In : Monestiez P, Allard D & Froidevaux R (eds) *geoENV III : Geostatistics for Environmental Applications*, 149-156, Kluwer, Amsterdam.
- Wolf T, Sénégas J, Bertino L & Wackernagel H (2001) Application of data assimilation to three-dimensional hydrodynamics : the case of the Odra lagoon. In : Monestiez P, Allard D & Froidevaux R (eds) *geoENV III : Geostatistics for Environmental Applications*, 157-168, Kluwer, Amsterdam.
- Bertino L, Wackernagel H & von Storch H (2001) Data assimilation and geostatistics in ecological modelling. In : Monestiez P, Allard D & Froidevaux R (eds) *geoENV III : Geostatistics for Environmental Applications*, 501-502, Kluwer, Amsterdam.
- Wackernagel H & Schmitt M (2001) Statistical interpolation models. In : von Storch H & Floeiser G (eds) *Models in Environmental Research*, 185-201, Springer-Verlag, Berlin.
- Bertino L, Sénégas J, Wackernagel H & von Storch H (2000) Data assimilation for hydrodynamical modeling of the Odra lagoon. In : Kleingeld WJ & Krige DG (eds) *Geostats 2000 : Cape Town, Proceedings of the 6th International Geostatistics Congress*, 449-458, De Beers, Cape Town.
- Wackernagel H, Thiéry L & Grzebyk M (1999) The Larsen model from a de Wijsian perspective. In : Gomez-Hernandez J, Soares A & Froidevaux R (eds) *geoENV II : Geostatistics for Environmental Applications*, 125-135, Kluwer, Amsterdam.
- Lajaunie C, Wackernagel H, Thiéry L & Grzebyk M (1999) Sampling multiphase noise exposure time series. In : Gomez-Hernandez J, Soares A & Froidevaux R (eds) *geoENV II : Geostatistics for Environmental Applications*, 101-112, Kluwer, Amsterdam.
- Wackernagel H, Lajaunie C, Thiéry L, Vincent R, Grzebyk M (1997) Applying geostatistics to exposure monitoring data in industrial hygiene. In : Soares A, Gomez-Hernandez J & Froidevaux R (eds) *geoENV I : Geostatistics for Environmental Applications*, 463-476, Kluwer, Amsterdam.
- Wackernagel H & Sanguinetti H (1993) Gold prospecting with factorial cokriging in the Limousin, France. In : Davis JC & Herzfeld UC (eds) *Computers in Geology : 25 years of progress*. *Studies in Mathematical Geology*, Vol. 5, 33-43, Oxford University Press, Oxford.

- Wackernagel H, Petitgas P & Touffait Y (1989) Overview of methods for coregionalization analysis. In : Armstrong M (ed) *Geostatistics*, Vol. I, 409–420, Kluwer, Amsterdam.
- Galli A & Wackernagel H (1988) Multivariate geostatistical methods for spatial data analysis. In : E Diday (ed) *Data Analysis and Informatics 5*, 273–280, Elsevier (North-Holland), Amsterdam.
- Wackernagel H (1988) Geostatistical techniques for interpreting multivariate spatial information. In : CF Chung et al (eds) *Quantitative Analysis of Mineral and Energy Resources*, 393–409, NATO-ASI Series C 223, Reidel, Dordrecht.
- Wackernagel H, Webster R & Oliver MA (1988) A geostatistical method for segmenting multivariate sequences of soil data. In : HH Bock (ed) *Classification and Related Methods of Data Analysis*, 641–650, Elsevier, Amsterdam.

### Actes de conférences et séries

- Vincent R, Grzebyk M, Wackernagel H & Lajaunie C (1999) Application de la géostatistique à l'hygiène industrielle : évaluation d'un cas d'exposition professionnelle au trichloroéthylène. INRS : Cahiers de Notes Documentaires - Hygiène et Sécurité du Travail, No 174, 5-13.
- Thiéry L, Wackernagel H & Lajaunie C (1996) Geostatistical analysis of time series of short  $L_{Aeq}$  values. Proceedings of Inter-Noise96, 30.7–2.8.96, Liverpool (UK), 2065–2068.
- Wackernagel H & Grzebyk M (1995) Linear models for spatial or temporal multivariate data. Proceedings of the 6th International Meeting on Statistical Climatology, 19–23 June 1995, Galway (Ireland), 427–429.
- Grzebyk M & Wackernagel H (1994) Multivariate analysis and spatial/temporal scales : real and complex models. Proceedings of XVIIth International Biometrics Conference, 8–12 August 1994, Hamilton (Ontario, Canada), Volume 1, 19–33.
- Lindner S & Wackernagel H (1993) Statistische Definition eines Lateritpanzer-Index für SPOT / Landsat-Bilder durch Redundanzanalyse mit bodengeochemischen Daten. Peschel G (ed) *Beiträge zur Mathematischen Geologie und Geoinformatik*, 5, 69–73, Sven-von-Loga Verlag, Köln.
- Wackernagel H (1993) Fondements de la géostatistique. In : Cicolella A (ed) *Estimation des Nuisances : Approches Statistiques*, Notes Scientifiques et Techniques de l'INRS, vol. 104, 1–14, Nancy.
- Wackernagel H & Hudson G (1992) Kriging mit externer Drift am Beispiel von Temperaturmessungen aus Schottland. In : Peschel G (ed) *Beiträge zur Mathematischen Geologie und Geoinformatik*, Bd 3, 15–24, Sven-von-Loga Verlag.
- Wackernagel H (1991) Changement de maille et changement de support. In : Mullon C (ed) *SEMINFOR 4 : Le Transfert d'Echelle*, 159–171, Editions de l'ORSTOM, Paris.
- Armstrong M & Wackernagel H (1988) The influence of the covariance function on the kriged estimator. *Sciences de la Terre*, Série Informatique, 27, 245–262.
- Wackernagel H (1985) The inference of the linear model of the coregionalization in the case of a geochemical data set. *Sciences de la Terre*, Série Informatique, 24, 81–93.

Prissang R, Wackernagel H (1980) Modelle zur globalen Vorratsschätzung. In : Burger H et al. (eds) *Aspekte der Zukunftsforschung in den Geowissenschaften*. Manuskripte des Geographischen Instituts der Freien Universität Berlin, 4, 183–201.

## Polycopié de cours de DEA

Wackernagel H (1993) *Cours de Géostatistique Multivariable*. Cours C-146 (4. édition), Centre de Géostatistique, Ecole des Mines de Paris, 80p. <ftp://cg.ensmp.fr/pub/hans/cmv.pdf>

## Rapports publiés sur Internet

Wackernagel H, Lajaunie C, Thiéry L & Grzebyk M (1998) Evaluation de l'exposition sonore en milieu professionnel : application de méthodes géostatistiques à l'estimation du Leq et conséquences sur les stratégies de mesure. Note technique MAV - NT-373/LT, Juin 1998, INRS, Vandoeuvre, 135 pages. [http://cg.ensmp.fr/~hans/indhyg\\_absnst.html](http://cg.ensmp.fr/~hans/indhyg_absnst.html)

Bertino L & Wackernagel H (2000) The Oderhaff salinity data : their use for data assimilation. N-9/00/G, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau. <http://pioneer.geogr.ku.dk>

Lajaunie C & Wackernagel H (2000) Geostatistical Normalization : Theoretical Framework. IMPACT Deliverable No 2. Rapport N-29/01/G, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau. <http://www.mai.liu.se/impact>

Lajaunie C & Wackernagel H (2000) Geostatistical Approaches to Change of Support Problems : Theoretical Framework. IMPACT Deliverable No 19. <http://www.mai.liu.se/impact>

Lajaunie C, Wackernagel H & Bertino L (2001) Geostatistical Normalization : Case Studies. IMPACT Deliverable No 3. <http://www.mai.liu.se/impact>

Bertino L & Wackernagel H (2002) Case Studies of Change-of-support Problems. IMPACT Deliverable No 20. <http://www.mai.liu.se/impact>

Wackernagel H (2002) Geostatistical normalization of air pollution transport model output and station data using Isatis. IMPACT Deliverable No 4. <http://www.mai.liu.se/impact>

## Mémoires

Wackernagel, H (1985) *L'Inférence d'un Modèle Linéaire en Géostatistique Multivariable*. Thèse de Doctorat de 3<sup>e</sup> Cycle en Sciences et Techniques Minières, Ecole des Mines de Paris, 100p.

Wackernagel H (1983) *Etude de la faisabilité macroéconomique d'une grande mine de phosphate*. Cours Postgrade sur les Pays en Développement, Institut de Génie Rural, Ecole Polytechnique Fédérale de Lausanne, 70p.

Wackernagel H (1981) *Etude géostatistique d'une couche de charbon*. Cycle de Formation Spécialisée en Géostatistique, Publication S-112, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, 43p.

# **Troisième partie**

## **Annexes**

# Liste des Articles

Les articles ne sont pas inclus et peuvent être obtenus séparément.  
Voir aussi : [www.cg.ensmp.fr/~hans/publications.html](http://www.cg.ensmp.fr/~hans/publications.html)

<b>A</b>	<i>Encyclopedia of Statistical Sciences</i> , 2nd edition, sous presse	<b>50</b>
<b>B</b>	<i>International Journal of Climatology</i> , 14 : 77–91	<b>66</b>
<b>C</b>	<i>Journal of Climate</i> , 12 : 1070-1085	<b>82</b>
<b>D</b>	<i>Ecological Modelling</i> , sous presse	<b>98</b>
<b>E</b>	<i>Stochastic Environmental Research and Risk Assessment</i> , 15 : 65–86	<b>115</b>
<b>F</b>	<i>GeoENV III – Geostatistics for Environmental Applications</i> , 157–168, Kluwer	<b>138</b>
<b>G</b>	<i>Inverse Problems</i> , 18 : 1–23	<b>151</b>
<b>H</b>	<i>International Statistical Review</i> , 71 : 223–241	<b>175</b>
<b>I</b>	<i>La Recherche</i> 371 : 98–99	<b>195</b>

## **Annexe A**

# ***Encyclopedia of Statistical Sciences, 2nd edition, sous presse***

Wackernagel H (2004)  
Geostatistics

## **Annexe B**

### ***International Journal of Climatology*, 14 : 77–91**

Gordon Hudson & Hans Wackernagel (1994)

Mapping temperature using kriging with external drift : theory and an example from Scotland.

## **Annexe C**

### ***Journal of Climate*, 12 : 1070-1085**

Gérard Biau, Eduardo Zorita, Hans von Storch & Hans Wackernagel (1999)  
Estimation of precipitation by kriging in the EOF space of the sea level pressure field.

## **Annexe D**

### ***Ecological Modelling, sous presse***

Wackernagel H, Lajaunie C, Blond N, Roth C & Vautard R (2003)  
Geostatistical risk mapping with chemical transport model output and ozone station data.

## **Annexe E**

### ***Stochastic Environmental Research and Risk Assessment, 15 : 65–86***

Julien S n gas, Hans Wackernagel, Wolfgang Rosenthal & Thomas Wolf (2001)  
Error covariance modeling in sequential data assimilation.

## **Annexe F**

### ***GeoENV III – Geostatistics for Environmental Applications, 157–168, Kluwer***

Thomas Wolf, Julien Sénégas, Laurent Bertino, Hans Wackernagel & (2001)  
Application of data assimilation to three-dimensional hydrodynamics :  
the case of the Odra lagoon.

# **Annexe G**

## ***Inverse Problems, 18 : 1–23***

Laurent Bertino, Geir Evensen & Hans Wackernagel (2002)

Combining geostatistics and Kalman filtering for data assimilation in an estuarine system.

## **Annexe H**

### ***International Statistical Review, 71 : 223–241***

Laurent Bertino, Geir Evensen & Hans Wackernagel (2003)  
Sequential Data Assimilation techniques in oceanography.

# **Annexe I**

## ***La Recherche, 371 : 98–99***

Laurent Bertino & Hans Wackernagel (2004)  
Prévoir l'état de l'océan  
(avec un complément WEB+ sur le site : [www.larecherche.fr](http://www.larecherche.fr)).