UNIVERSITÉ MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ MONTPELLIER II

Discipline : Mathématiques appliquées
Formation doctorale : Biostatistique
Ecole doctorale : Information, Structures, Systèmes

présentée et soutenue publiquement

par

Edith GABRIEL
le jeudi 02 décembre 2004

Titre :

DÉTECTION DE ZONES DE CHANGEMENT ABRUPT DANS DES DONNÉES SPATIALES

ET APPLICATION À L'AGRICULTURE DE PRÉCISION

JURY

Denis Allard,	Co-directeur de thèse - Chargé de recherche INRA Avignon
Jean-Marc Azaïs,	Rapporteur - Professeur Université Toulouse III
Gilles Ducharme,	Président - Professeur Université Montpellier II
Xavier Guyon,	Rapporteur - Professeur émérite Université Paris I
Bruno Mary,	Examinateur - Directeur de recherche INRA Laon
Rachid Senoussi,	Directeur de thèse - Directeur de recherche INRA Avignon

Remerciements

J'adresse mes sincères remerciements,

A mon encadrant, Denis Allard, non seulement pour m'avoir accordé sa confiance durant ces trois années, pour la qualité de ses conseils et pour la pertinence de ses remarques, mais aussi pour avoir su me donner des ailes.

A Martine Guérif, ma co-directrice et interlocutrice agronome, pour avoir soulevé bon nombre de problèmes.

A Joël Chadœuf, pour sa perspicacité et dont les nombreuses critiques ont permis d'enrichir ce travail.

A Rachid Senoussi et toute l'équipe de biométrie pour s'être montrés accueillants et présents à tout moment, avec un clin d'œil à Malika pour sa maîtrise de la gestion de l'informatique de l'unité.

A Jean-Noël Bacro, pour avoir fait naître un des problèmes originaux de cette thèse.

A mon jury de thèse, Bruno Mary pour m'avoir donné l'opportunité d'illustrer ce travail, Jean-Marc Azaïs et Xavier Guyon, pour avoir accepté de rapporter sur cette thèse, et plus particulièrement à Gilles Ducharme qui m'a guidée vers le monde de la recherche.

A l'équipe technique de l'unité d'Agronomie de Laon-Reims-Mons pour le travail fourni sur le terrain et au laboratoire : Gonzague Alavoine, Florence Barrois, Daniel Boitez, Olivier Delfosse, Eric Gréhan, Christian Herre, Frédéric Mahu, Francis Millon et Eric Venet.

A tous les thésards de l'équipe de biométrie avignonaise, Edwige, Emilie, Gaël, Nicolas, Rémy, Samuel, mais surtout à Sabrina, qui m'a toujours soutenue et écoutée.

A mes parents, pour leur encouragement durant ces longues années d'études.

Et enfin, à Fabrice pour l'amour, le soutien et la patience dont il fait preuve.

Table des matières

Introduction **9**

I Problématique **15**

Introduction 17

1 Comment détecter des ZCAs ? **19**

1.1 Le wombling et ses dérivées 19

1.2 Méthodes statistiques 27

1.3 Problèmes résultants 31

2 Champs aléatoires **35**

2.1 Stationnarité 36

2.2 Ergodicité 38

2.3 Continuité et différentiabilité 39

2.4 Champ aléatoire conditionnel 41

2.5 Champs gaussiens 42

2.6 Champs de χ^2 44

3 Géostatistique **47**

3.1 Krigeage 47

3.2 Propriétés du prédicteur de krigeage 51

3.3 Modèles de covariance 54

4 Ensembles d'excursion de champs aléatoires **59**

4.1 Nombre de composantes connexes 59

4.2 Taille d'une composante connexe 63

4.3 Ensembles d'excursion de champs de χ^2 64

II Détection de Zones de Changement Abrupt 69

Introduction . . . 71

5 Détection locale des ZCAs 75

5.1 Notations . . . 75

5.2 Test local de changement abrupt . . . 77

5.3 Illustration . . . 78

6 Détection globale des ZCAs 81

6.1 Courbure du champ $T(\cdot)$ . . . 81

6.2 Champ conditionnel . . . 87

6.3 Distribution asymptotique de la taille d'une ZCA . . . 89

6.4 Test de significativité . . . 91

6.5 Illustration . . . 93

7 Implémentation 95

7.1 Discrétisation du plan par une grille . . . 95

7.2 Détermination de α . . . 97

7.3 Estimation de la covariance sous l'alternative . . . 107

8 Discussion 111

8.1 Espérance inconnue . . . 111

8.2 Modèles de covariance plus généraux . . . 113

8.3 Comparaison avec d'autres méthodes . . . 114

8.4 Amélioration de l'algorithme . . . 115

8.5 Schéma d'échantillonnage . . . 117

III Puissance du test de détection 119

Introduction . . . 121

9 Calcul de la puissance pour un champ sur $\mathbb{R}$ 123

9.1 Calcul de la puissance . . . 123

9.2 Pour un modèle de covariance exponentiel . . . 125

9.3 Pour un modèle de covariance sphérique . . . 127

10 Calcul de la puissance pour un champ spatial sur $\mathbb{R}^2$ 133
10.1 Spécification de l'hypothèse alternative . . . 133
10.2 Puissance ponctuelle . . . 135
10.3 Non indépendance des tests locaux . . . 138
10.4 Puissance conditionnelle locale . . . 141
Discussion . . . 151

IV Validation de la méthode 153
Introduction . . . 155

11 Puissance 157
11.1 Puissance globale . . . 157
11.2 Effet de la densité de points . . . 159

12 Sensibilité 161
12.1 Par rapport à l'estimation de la covariance . . . 161
12.2 Par rapport à la discrétisation . . . 163
12.3 Par rapport au choix de l'interpolateur spatial . . . 164
Discussion . . . 169

V Application aux données d'agriculture de précision 171
Introduction . . . 173

13 Présentation des données 175
13.1 Plan d'échantillonnage . . . 175
13.2 Variables dynamiques . . . 177
13.3 Variables permanentes . . . 181

14 Résultats 183
14.1 Variables dynamiques . . . 183
14.2 Variables permanentes . . . 188
14.3 Validation de la méthode sur les données . . . 188
14.4 Effet de la densité locale de l'échantillon . . . 191
Discussion . . . 199

Conclusion **201**

14.5 Bilan 201

14.6 Perspectives 203

VI Annexes **207**

A Décomposition du champ $T(\cdot)$ **209**

B Propriétés du champ $U(\cdot)$ **211**

C Propriétés du champ $T(\cdot)$ **219**

Bibliographie **223**

Liste des figures **231**

Liste des tables **234**

Introduction

L'évolution de l'agriculture lors des dernières décennies a conduit à des tailles de parcelles de plus en plus grandes. Cet accroissement, associé à la variabilité de la nature des sols, induit une hétérogénéité intra-parcellaire de plus en plus forte. Une conduite homogène sur de telles parcelles mène à des performances très variables en terme de production, mais aussi à des effets néfastes pour l'environnement. L'agriculture de précision offre la perspective de prendre en compte cette hétérogénéité pour gérer les interventions techniques afin d'optimiser la production et de limiter les nuisances environnementales. De nombreuses techniques existent aujourd'hui pour cartographier cette hétérogénéité, comme le système de positionnement par satellite, la télédétection aéroportée ou satellitale ou encore les capteurs embarqués sur divers engins. L'adaptation du savoir-faire agronomique à ce flux d'informations peut permettre de proposer une modulation spatiale des interventions culturales. L'Institut National de la Recherche Agronomique (INRA) et les instituts techniques, tel que le Centre Technique Interprofessionnel des Oléagineux Métropolitains (CETIOM), ont mis en place des programmes de recherche dans ce domaine, appliqués aux grandes cultures, en se focalisant en premier lieu sur les questions de fertilisation azotée dont les enjeux sont importants aussi bien en terme de facteur de production que le risque pour l'environnement par pollution des eaux.

La modulation spatiale des interventions culturales nécessite deux étapes. La première consiste à définir des unités spatiales élémentaires et la deuxième à appliquer un niveau de pratique agricole (intrant, ...) adapté aux caractéristiques de chaque unité. La première étape peut être abordée soit en établissant une cartographie exhaustive des propriétés de la parcelle avec une résolution spatiale fine, soit en définissant des zones homogènes. L'INRA vise, en particulier, à mettre en place des techniques de modulation intra-parcellaire de la fertilisation azotée du blé en fonction des caractéristiques locales de la parcelle [Guérif *et al.* (2001), Houlès (2004)]. La gestion locale de la fertilisation azotée nécessite d'avoir une description spatiale du milieu et des couverts végétaux et ainsi d'obtenir une spatialisation des variables d'état sol/plante. Ces variables sont introduites dans un modèle de culture

qui fournit une aide à la préconisation azotée spatialisée. L'introduction de l'information dans un modèle de culture peut se faire point à point en travaillant à la résolution minimale possible, *i.e.* celle imposée par les machines. Ce procédé ne tient pas compte de la structure spatiale de la parcelle. Il consiste dans un premier temps à obtenir une description des variables selon une maille arbitraire, puis à transcrire ces variables sur un support commun (grille régulière) et enfin à introduire l'information dans le modèle de culture avec une résolution fine (de la taille d'un pixel de la grille commune). Cependant travailler à une telle résolution implique des difficultés techniques considérables. En particulier il est nécessaire de connaître les variables du sol avec une précision suffisante en tout point (ou aux nœuds d'une grille ayant la résolution choisie) ce qui représente une tâche difficile et onéreuse. De plus, la carte de préconisation issue du modèle de culture risque de présenter des discontinuités importantes, difficiles (voire impossibles) à gérer par les machines. Les difficultés techniques rencontrées dans la modulation continue peuvent être écartées en considérant des zones homogènes. Les zones homogènes peuvent être définies par des zones de potentiel de production, en faisant l'hypothèse qu'elles nécessitent des niveaux de pratique agricole homogènes et différenciées entre elles. Ces zones sont généralement définies par des méthodes de classification, voir Lawson et Denison (2002) pour un panarama complet de cette problématique. Les méthodes de classification peuvent être appliquées soit à des caractéristiques du sol [Shatar et McBratney (2001), Castrignanò *et al.* (2003)], soit à des données pluriannuelles de cartographie des rendements [Gilliot *et al.* (2004)], soit à des données multiples (sol, rendement, ...) [Whelan et McBratney (2003)]. Dans Shatar et McBratney (2001) un algorithme de clustering dérivé des nuées dynamiques et prenant en compte les distances géographiques entre les points d'échantillonnage est proposé. Bien que prometteuse, cette approche n'est pas construite sur des bases théoriques solides dans sa version actuelle et présente certains inconvénients. Par exemple, le nombre de zones doit être spécifié à l'avance au lieu d'être le résultat de l'analyse de données. De plus, les germes de départ de l'algorithme ont une influence très importante sur les résultats de la classification.

La définition des limites entre zones homogènes reste un problème dans ces méthodes et les conséquences, en terme d'imprécision sur les traitements réalisés, peuvent être importantes, particulièrement quand il existe des changements abrupts entre deux zones. Nous proposons une approche différente des méthodes de classification, reposant sur l'estimation de zones où la variable étudiée varie brusquement. De telles zones constituent potentiellement des frontières entre zones homogènes. Ces lieux sont appelés Zones de Changement Abrupt (ZCAs). La variable est modélisée par un champ aléatoire $Z(\cdot)$.

Détecter des ZCAs revient à rejeter l'hypothèse nulle de stationnarité de la moyenne du champ aléatoire en posant comme alternative la présence de discontinuités de la moyenne. La méthode que nous proposons est complémentaire aux méthodes de classification, dans le sens où les clusters issus de la classification ne définissent pas toujours des Zones de Changement Abrupt le long de leur frontière, et inversément les ZCAs ne définissent pas nécessairement des groupes spatialement bien séparés. Notre approche met en évidence les structures existantes, ce qui associé à d'autres informations (comme la pédologie et l'évolution temporelle de certaines variables) peut aussi permettre de déterminer un zonage.

La détection de Zones de Changement Abrupt est également un problème de la biologie, Womble (1951) cherchait à déterminer des frontières entre populations. Les 5% des valeurs les plus élevées du gradient des variables d'étude (fréquences de gènes) définissaient les ZCAs. Gleyze *et al.* (2001) ont proposé une méthode statistique pour tester la significativité des ZCAs détectées. Bien qu'originale de par son concept, cette méthode s'est avérée être un échec en terme d'applicabilité. La méthode que nous proposons prend en compte les structures spatiales de la variable dans la définition des ZCAs, ce qui faisait défaut dans les méthodes issues de la procédure de Womble (1951). Elle reprend le concept de tester la significativité des ZCAs détectées et se base sur les ensembles d'excursion de champs aléatoires pour y parvenir. Nous nous sommes inspirés de travaux réalisés en analyse d'images médicales, plus particulièrement des travaux de Worsley (1994) et Cao (1999). Dans ces travaux, les images obtenues par tomographie par émission de positrons (PET) et par résonnance magnétique fonctionnelle (fMRI) ont été analysées afin d'identifier les zones du cerveau réagissant à des stimulations. Ces images sont modélisées comme des champs aléatoires gaussiens et l'ensemble d'excursion est un ensemble de points où le champ dépasse un seuil donné. Leur démarche nécessite un test local de détection et un test global de significativité des réactions. Le test global est basé sur la taille de la plus grande composante connexe de l'ensemble d'excursion. De nombreux résultats que nous développons dans ce travail généralisent ceux de Cao (1999).

Dans la première partie nous explicitons le problème de la détection de ZCAs. Les travaux de Womble (1951) et les approches qui en sont issues y sont détaillés en premier lieu. Des approches basées sur des fondements plus statistiques [Hall *et al.* (2001), Chaudhuri et Marron (1999), Godtliebsen *et al.* (2002)] sont ensuite décrites. Comme la méthode se place dans le cadre général des champs aléatoires [Adler (1981)] et de la géostatistique [Chilès et Delfiner (1999), Cressie (1993)], nous introduisons des pro-

priétés de continuité et de différentiabilité des champs aléatoires, ainsi qu'une méthode géostatistique de prédiction, le krigeage, utilisée pour interpoler la variable considérée sur le domaine d'étude. Nous énonçons ensuite quelques propriétés sur les ensembles d'excursion et les principaux résultats de Worsley (1994) et de Cao (1999).

La seconde partie est consacrée au développement de la méthode de détection de ZCAs. La méthode est basée sur les propriétés locales de l'estimation du gradient du champ aléatoire modélisant la variable d'étude. Dans un premier temps le gradient est interpolé de façon optimale par krigeage. La variable est supposée spatialisée et faiblement échantillonnée, aussi l'estimation du gradient du champ est non stationnaire. Nous définissons, sous l'hypothèse de gaussianité du champ aléatoire, un test local dont la statistique est reliée à l'estimation du gradient. Ce test permet de détecter les points où l'hypothèse nulle locale de stationnarité est rejetée. Cet ensemble de points correspond à un ensemble d'excursion de la statistique au dessus d'un seuil donné et définit les ZCAs potentielles. Nous agrégeons ensuite les tests locaux en un test global en utilisant la distribution asymptotique de la taille des ZCAs potentielles sous l'hypothèse nulle, afin de tester la significativité des ZCAs potentielles. La méthode requiert donc deux niveaux de significativité : un niveau global, $1-\eta$, que nous fixons, et un niveau local, $1-\alpha$, que nous déterminons. En pratique la méthode est appliquée aux nœuds d'une grille régulière superposée sur le domaine d'étude. Aussi, le niveau $1-\alpha$ doit tenir compte de η, de la discrétisation, du schéma d'échantillonnage et de la fonction de covariance du champ aléatoire. La fonction de covariance est en théorie supposée connue et nécessite d'être estimée en pratique. Les tests étant effectués sous l'hypothèse nulle d'absence de ZCAs, nous suggérons une procédure itérative permettant d'estimer les ZCAs et le modèle de covariance à partir des données. Cette procédure permet de corriger le biais introduit dans l'estimation de la variance en présence d'une ZCA.

La qualité de l'estimation du gradient étant liée à la densité d'échantillonnage et la variable pouvant être irrégulièrement échantillonnée, nous devons regarder s'il est pertinent d'appliquer la méthode sur l'ensemble du domaine ou s'il existe des zones sur lesquelles la méthode ne peut être appliquée. Une réponse à ce problème repose sur le calcul de la puissance du test local de détection de changement abrupt. Dans la troisème partie, la puissance est calculée en chaque point du domaine d'étude. Chacun de ces points est le centre d'une fenêtre mobile et appartient à une discontinuité. La cartographie de la puissance est ainsi possible. Les zones de faible puissance définissent les lieux où l'échantillonnage n'est pas approprié à l'estimation des ZCAs.

Dans la quatrième partie nous validons la méthode sur des simulations. La puissance

globale de la méthode a été évaluée pour des échantillons de points de taille variable. L'influence de la discrétisation sur la détection des ZCAs a ensuite été analysée. Le modèle de covariance nécessitant d'être estimé en pratique, des exercices de simulations ont permis de discuter de la sensibilité de la méthode par rapport à une erreur commise dans l'estimation des paramètres de la fonction de covariance (forme paramétrique et portée).

L'application de la méthode à des données d'agriculture de précision est présentée dans la cinquième partie. Les caractéristiques permanentes (limon, calcaire, argile, ...) et non permanentes (teneur en eau et en azote minéral) sont analysées sur une parcelle du nord de la France.

Première partie

Problématique

Introduction

La détection de Zones de Changement Abrupt est un problème rencontré dans de nombreux domaines tels que la biologie, l'écologie ou l'agriculture, lorsqu'il est intéressant et nécessaire de cartographier les zones où une variable d'étude présente des changements abrupts. Prenons l'exemple de l'étude des populations en biologie et en écologie. Les changements abrupts dans les fréquences d'allèles peuvent être liés aux frontières entre différentes populations. Le problème d'estimation des ZCAs a été introduit par Womble (1951). Les ZCAs étaient définies comme l'ensemble des points où le gradient d'une variable calculé sur une grille d'interpolation varie le plus, par exemple les 5% valeurs les plus élevées. Le principal défaut de cette méthode est que, dans tous les cas, 5% des valeurs sont conservées. Cette méthode a été améliorée dans Barbujani *et al.* (1989) où la direction du gradient fut incluse dans la définition des ZCAs. Elle a été généralisée au cas multivariable dans Bocquet-Appel et Bacro (1994). Gleyze *et al.* (2001) ont proposé de tester la significativité des ZCAs détectées, *i.e.* de rejeter l'hypothèse nulle d'absence de structure spatiale. L'inconvénient dans leur méthode est qu'elle ne peut être appliquée en pratique car elle repose sur une hypothèse très restrictive ne reflétant pas les phénomènes biologiques qui nous intéressent. Nous analysons dans le chapitre 1 l'évolution de la méthode introduite par Womble (1951).

L'estimation de ZCAs est un problème lié à l'estimation de frontières, la détection de ruptures ou de contours. La détection de ruptures joue un rôle important en traitement d'images et en traitement du signal, aussi existe-t-il une vaste littérature à ce sujet. De nombreuses méthodes incluent les champs de Markov comme Geman et Geman (1984), Besag (1986) ou Besag *et al.* (1995). D'autres sont basées sur la théorie des ondelettes [Wang (1998)] ou sur des méthodes non paramétriques [Qiu (1998)].

Le problème de l'estimation de courbes de discontinuités de champs aléatoires dans le plan pour des échantillons irréguliers a été beaucoup moins abordé. Hall *et al.* (2001) ont développé une méthode de *tracking*, basée sur les approximations locales de la vraisemblance qu'une discontinuité existe en un point donné du plan. Cette méthode est détaillée

dans le chapitre 1. Dans ce chapitre, nous nous intéressons également à une méthode, SiZer, et à ses extensions. SiZer permet de tester la significativité des structures détectées dans des processus unidimensionnels [Chaudhuri et Marron (1999)] ou spatiaux [Godtliebsen *et al.* (2002)].

La méthode que nous proposons dans la deuxième partie de ce travail se situe dans le cadre des champs aléatoires [Adler (1981) ; Yaglom (1986)] et de la géostatistique [Cressie (1993) ; Chilès et Delfiner (1999)]. Les données géoréférencées $(Z_1, \ldots, Z_n)$ peuvent, dans ce cadre, être modélisées par un champ aléatoire $Z(\cdot)$. Il s'agit d'un ensemble de quantités aléatoires $Z(\mathbf{x})$, indexé sur $\mathcal{D} \subset \mathbb{R}^d$,

$$Z(\cdot) = \left\{ Z(\mathbf{x}) : \mathbf{x} = \left(x^1, \ldots, x^d\right)' \in \mathcal{D} \right\}. \tag{1}$$

Nous allons rappeler dans le chapitre 2 des généralités sur les champs aléatoires ainsi que quelques propriétés de continuité et de différentiabilité de champs aléatoires stationnaires.

Comme nous travaillons à partir d'un échantillon de points pouvant être d'assez faible densité et irrégulièrement échantillonné, nous devons interpoler le champ aléatoire sur le domaine. L'interpolation est faite à l'aide d'un outil issu de la géostatistique, le krigeage, dont nous exposons le principe dans le chapitre 3.

Une des originalités de la méthode de détection de Zones de Changement Abrupt que nous proposons est qu'elle permet de tester la significativité des ZCAs détectées. Elle est fondée sur les ensembles d'excursion de champs de χ^2 non stationnaires. Dans Worsley (1994) et Cao (1999), une méthode reposant sur les ensembles d'excursion de champs aléatoires est mise en place dans un contexte d'analyse d'images médicales. Ces travaux reposent en particulier sur la détection de zones du cerveau, modélisées par un ensemble d'excursion de champ de χ^2 standard, réagissant à des stimulations. Leurs principaux résultats sont énoncés dans le chapitre 4. La principale différence entre la méthode de détection de ZCAs que nous proposons et celle de Worsley (1994) et Cao (1999) est que nous travaillons à partir d'une variable spatialisée, alors qu'ils travaillent à partir d'images. Aussi, le champ interpolé de la méthode de détection de ZCAs est non stationnaire. De nombreux résultats que nous développons dans ce travail généralisent ceux de Cao (1999) au cas non stationnaire dans le cas d'un champ spatial sur $\mathbb{R}^2$.

Chapitre 1

Comment détecter des ZCAs ?

1.1 Le wombling et ses dérivées

Womble (1951) a proposé une méthode permettant de détecter des barrières biologiques. Sa méthode est basée sur la moyenne de la valeur absolue de la dérivée des fonctions décrivant les variations biologiques dans l'espace en différents sites. Nous allons décrire cette méthode et suivre son évolution.

Le wombling

La procédure de Womble (1951), le *wombling*, consiste à détecter des Zones de Changement Abrupt au travers de la norme du gradient d'une variable spatiale Z, pour laquelle un nombre fini de données est mesuré sur une surface géographique donnée.
Le wombling se résume de la façon suivante. Les données sont dans un premier temps interpolées sur une grille régulière. La norme du gradient est ensuite calculée au centre $\mathbf{x}$ de chaque pixel de la grille. Le carré formé par les nœuds de la grille entourant le point $\mathbf{x} = (x^1, x^2)'$ en lequel le gradient est calculé, est noté $ABCD$ (figure 1.1). Womble (1951)

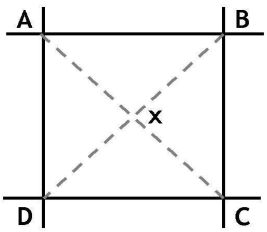


FIG. 1.1: *Représentation d'un pixel de la grille d'interpolation.*

appelle *fonction systémique* la norme du gradient $\mathcal{W}$, obtenu et défini à chaque pixel en utilisant l'interpolation au premier ordre des pentes le long de chaque coordonnée avec les quatre nœuds les plus proches :

$$\mathcal{W} = \begin{pmatrix} \mathcal{W}_V \\ \mathcal{W}_H \end{pmatrix} = \begin{pmatrix} (Z_A - Z_D)/2 + (Z_B - Z_C)/2 \\ (Z_A - Z_B)/2 + (Z_D - Z_C)/2 \end{pmatrix}.$$

Les Zones de Changement Abrupt sont définies par l'ensemble des pixels connectés de la manière suivante : à partir des quantiles empiriques à 90 et 95 %, q^*_{90} et q^*_{95}, calculés sur les valeurs de la fonction systémique, il est possible de définir deux familles de pixels :

- les q_{95}-pixels, qui sont les pixels où les valeurs de la fonction systémique sont supérieures à q^*_{95},
- les q_{90}-pixels, qui sont les pixels où les valeurs de la fonction systémique sont comprises entre q^*_{90} et q^*_{95}.

Les q_{95}-pixels voisins sont liés entre eux, de même que les q_{95}-pixels séparés par au plus un q_{90}-pixel (un tel pixel est appelé un *pont*). Chaque ensemble différent de q_{95}-pixels reliés et leurs ponts définit une barrière, *i.e.* une ZCA.
Afin de synthétiser plusieurs variables en une variable composite, Womble (1951) propose de calculer la moyenne des valeurs absolues des dérivées de chaque variable et d'opérer de la même façon que pour une seule variable pour détecter les barrières.

Womble (1951) a développé sa méthode pour des fréquences de gènes et pour des mesures morphologiques. La génétique étant depuis des décennies un thème de recherche très important, le wombling a suscité beaucoup d'intérêt pour de nombreux biologistes, qui ont contribué à l'évolution de cette méthode. En effet, le wombling a de sérieuses limites. Le calcul de la fonction systémique ne prend pas en compte les corrélations entre variables et ne considère pas l'effet de lissage dû à l'interpolation ce qui engendre des biais importants dans le calcul de la fonction systémique. De plus, étant basée sur une sélection par quantile, la procédure de Womble détecte toujours des ZCAs, ce qui ne permet pas de juger de la significativité des résultats. Ces limites ont pu être étudiées dans divers travaux comme nous allons le voir dans ce qui suit.

Adams, 1970

Adams (1970) soulève deux inconvénients dans le wombling. Le premier est d'ordre numérique : le wombling s'avère lourd en terme de calcul de la fonction systémique. Le

second est d'ordre scientifique : comment pondérer les différentes variables intervenant dans le calcul de la fonction systémique ?

Jusqu'en 1970, peu de biologistes ont utilisé le wombling car il était laborieux de calculer la fonction systémique. Avec l'arrivée des méthodes modernes de calcul, cet handicap a pu être éliminé et Adams (1970) a proposé un algorithme de calcul. Son algorithme introduit l'utilisation de la matrice des points de la grille carrée, de taille $n \times n$, générée par un programme de contour.
La variable Z est interpolée sur la grille régulière par krigeage (voir section 3.1). Les nœuds A, B, C et D, délimitant un pixel de la grille correspondent aux valeurs interpolées de la variable. Le gradient, en leur centre $\mathbf{x}$ est un vecteur W. Ce vecteur se décompose en $(W_1, W_2)'$ avec $W_1 = (Z_A - Z_C)/\sqrt{2}$ le long de l'axe AC et $W_2 = (Z_B - Z_D)/\sqrt{2}$ le long de l'axe BD (figure 1.2).

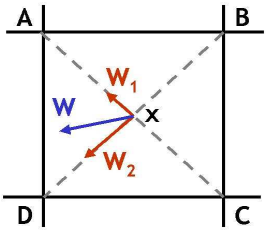


FIG. 1.2: *Décomposition de W.*

En définissant W à la place de $\mathcal{W}$, Adams (1970) simplifie les calculs et la norme du gradient est inchangée : $\|\mathcal{W}\| = \|W\|$.

Le second problème évoqué par Adams (1970) est le suivant : lors de la recherche de barrières dans le cas multivariable, est-ce que considérer une variable synthétique constitue une approche pertinente ? et quels poids attribuer aux différentes variables lors du calcul de la fonction systémique ? Il n'y a pas de solution évidente. Il propose de pondérer plus lourdement les variables ayant des écarts significatifs (test de Student-Newman-Keuls) ou d'assigner des poids proportionnels à la variabilité de chaque variable, mais laisse cette question ouverte.

Barbujani, Oden, Sokal, 1989

Dans Womble (1951), les barrières, *i.e.* les ZCAs, sont définies par les régions où la norme du gradient est élevée. Barbujani *et al.* (1989) ont étendu ce concept en intégrant la direction du gradient. Ils proposent un algorithme permettant de calculer ces normes et ces directions ainsi qu'une méthode de détection des barrières basée sur le wombling.

Barbujani *et al.* (1989) considèrent les deux matrices suivantes : la matrice de la norme du gradient calculée en chaque point de la grille (M_g) et la matrice des directions du gradient (M_d). Pour les calculer, le gradient $\widetilde{W}$ est déterminé au point $\mathbf{x} = (x^1, x^2)'$ centre du carré $ABCD$ (figure 1.3*a*) :

$$\widetilde{W} = \left(\frac{\partial Z(\mathbf{x})}{\partial x^1}, \frac{\partial Z(\mathbf{x})}{\partial x^2} \right)',$$

avec $Z(\mathbf{x}) = Z_A(1-x^1)(1-x^2) + Z_B x^1(1-x^2) + Z_C x^1 x^2 + Z_D(1-x^1)x^2$. Les dérivées partielles de $Z(\mathbf{x})$ s'écrivent :

$$\begin{array}{rcl} \frac{\partial Z(x)}{\partial x^1} & = & Z_B - Z_A + x^2(Z_A - Z_B + Z_C - Z_D), \\ \frac{\partial Z(x)}{\partial x^2} & = & Z_D - Z_A + x^1(Z_A - Z_B + Z_C - Z_D). \end{array} \tag{1.1}$$

Ainsi, pour $A = (0,1)'$, $B = (1,1)'$, $C = (1,0)'$, $D = (0,0)'$, au centre $\mathbf{x} = (1/2,\, 1/2)'$ du carré $ABCD$ les gradients calculés selon la procédure de Womble (1951), de Adams (1970) et de Barbujani *et al.* (1989) ont la même norme :

$$\|\mathcal{W}\| = \|W\| = \|\widetilde{W}\|.$$

Lorsque $ABCD$ est petit, la norme ainsi calculée est assez proche de celle du gradient réel. Sa direction est :

$$\theta = \arctan\left(\frac{\partial Z(\mathbf{x})}{\partial x^2} \Big/ \frac{\partial Z(\mathbf{x})}{\partial x^1} \right). \tag{1.2}$$

Ceci est illustré dans la figure 1.3*b*.

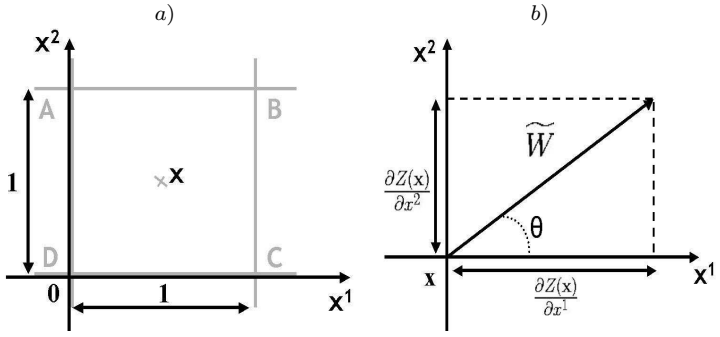


FIG. 1.3: *a*) *Notations, b*) *Décomposition de* $\widetilde{W}$.

Il est alors possible, grâce aux équations (1.1) et (1.2), de calculer les matrices M_g et M_d. Dans certains exemples la simple inspection de ces matrices permet d'inférer l'emplacement des barrières. Cependant, dans la majorité des cas, il est nécessaire de mettre en

commun l'information contenue dans les deux matrices et donc de donner une nouvelle définition des barrières.

L'information contenue dans les matrices M_g et M_d peut être synthétisée afin d'identifier les ZCAs. Pour se faire, une première approche consiste à tracer une série de barres sur le plan, la longueur d'une barre étant proportionnelle à l'élément de M_g correspondant et son orientation à l'élément de M_d. Une approche alternative est basée sur le principe qu'une ZCA est caractérisée par des gradients élevés sur un voisinage de points.

Un nouveau critère a été défini pour relier les pixels dans les frontières. Il impose que les connexions soient faites seulement lorsque la différence entre les directions des gradients voisins est inférieure à une valeur arbitraire (Barbujani *et al.* (1989) ont choisi 30 degrés). Les deux critères localisent les ZCAs. Le problème est que ces critères mettent en évidence toujours la même proportion de pixels quelle que soit la variable étudiée. En effet, la procédure de Womble (1951) détecte toujours des ZCAs parce qu'elles proviennent des q_{95} et q_{90}-pixels dont le nombre est constant en proportion quelles que soient les valeurs du gradient.

Pour ce qui est du problème d'obtenir les normes et les directions lorsque Z est constitué de p variables, Barbujani *et al.* (1989) considèrent les p gradients associés : $\widetilde{W}_i = (\widetilde{W}_{i,1}, \widetilde{W}_{i,2})$, $i = 1, ..., p$. La fonction systémique est alors définie par la norme moyenne :

$$\overline{\|\widetilde{W}\|} = \frac{1}{p} \sum_{i=1}^{p} \|\widetilde{W}_i\|.$$

Afin d'obtenir la direction moyenne, l'angle est calculé à partir des gradients moyens :

$$(w_1, w_2) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{\|\widetilde{W}_i\|} (\widetilde{W}_{i,1}, \widetilde{W}_{i,2}) \text{ et } \theta = \arctan\left(w_1/w_2\right).$$

En ce qui concerne le problème de pondération soulevé par Adams (1970), ils suggèrent de s'adapter selon les cas. Par exemple, lors d'une étude en génétique, un chercheur pourra préférer donner des poids égaux à chaque locus étudié plutôt qu'à chaque allèle, ou encore pondérer par l'inverse de l'écart-type comme le suggérait Adams (1970).

Un inconvénient supplémentaire, dans les méthodes précédemment citées pour le cas multivariable, est qu'en moyennant les normes des gradients l'information "changement abrupt" peut être noyée si elle porte sur une seule variable.

Bocquet-Appel et Bacro, 1994

Selon Bocquet-Appel et Bacro (1994), le wombling amélioré par Barbujani *et al.* (1989) est inadapté lorsque les variables sont corrélées. En effet, caractériser les changements abrupts via la procédure de sommation fournit une surestimation de ces changements, *i.e.* une surestimation de la valeur de la fonction systémique, du fait de la redondance de l'information contenue dans ces variables. Le problème est que les variables corrélées ont le même poids dans la fonction systémique que celles qui ne le sont pas. Si les deux catégories de variables (corrélées et non corrélées) sont générées par le même environnement (géographique par exemple) et/ou par un processus produisant un ou plusieurs schémas communs à toutes les variables, l'équi-pondération est sans conséquence car c'est cette variable sous-jacente qui nous intéresse et qui sera renforcée. Mais si les deux catégories sont générées par des processus différents, il y a un risque d'identifier de fausses Zones de Changement Abrupt. Dans ce cas une analyse par variable pourrait être préférable. Bocquet-Appel et Bacro (1994) ont préféré introduire, dans la procédure de Womble (1951), une fonction systémique qui prend en compte l'information redondante entre les variables. De plus, toutes les variables n'ont probablement pas la même contribution sur les barrières observées, ce qui soulève le problème de l'influence individuelle des variables sur les barrières.

Bocquet-Appel et Bacro (1994) ont calculé une matrice d'information entre les variables (matrice de covariance) et interpolé les variables sur une grille à l'aide d'un interpolateur spatial. Le champ multivarié, $Z(\mathbf{x})$, et les valeurs de ses dérivées partielles sont calculées au centre de chaque pixel de la grille d'interpolation. L'idée est d'écrire le carré de la fonction systémique comme étant la somme de deux formes quadratiques :

$$\|\breve{W}(\mathbf{x})\|^2 = \left(\frac{\partial Z(\mathbf{x})}{\partial x^1}\right)' \mathbf{R}^{-1} \left(\frac{\partial Z(\mathbf{x})}{\partial x^1}\right) + \left(\frac{\partial Z(\mathbf{x})}{\partial x^2}\right)' \mathbf{R}^{-1} \left(\frac{\partial Z(\mathbf{x})}{\partial x^2}\right),$$

où $\mathbf{R}$ est la matrice de covariance entre les p variables. Pour $\mathbf{R}^{-1} = (r_{kl})_{k,l=1,\ldots,p}$, la fonction systémique a pour expression :

$$\breve{W} = \left[\sum_{k=1}^{p}\sum_{l=1}^{p} r_{kl}(\breve{W}_{1,k}\breve{W}_{1,l} + \breve{W}_{2,k}\breve{W}_{2,l})\right]^{\frac{1}{2}},$$

où $\breve{W}_{i,j},\ i = 1, 2$ est la $j^{ème}$ coordonnée du vecteur $\partial Z(\mathbf{x}) \ / \ \partial x^i$.

Les pixels sélectionnés comme dans la dernière étape de la description du wombling font apparaître une ou plusieurs barrières. Il est possible de déterminer l'influence individuelle des variables dans les barrières grâce aux mesures suivantes :

– *Importance des barrières* :

$$c = \sum_{i \in \mathcal{B}}^{n_{\mathcal{B}}} \breve{W}_i^2,$$

où $\mathcal{B}$ constitue l'ensemble des barrières observées, $n_{\mathcal{B}}$ représente le nombre de pixels dans les barrières et $\breve{W}_i^2$ est le carré de la fonction systémique au $i^{ème}$ pixel.

– *Composante de la barrière dûe à la $k^{ème}$ variable* :

$$c_k = \sum_{i \in \mathcal{B}}^{n_{\mathcal{B}}} \left\{ \left(\frac{\partial Z_k(x)}{\partial x^1} \right)_i \sum_{l=1}^{p} \left(\frac{\partial Z_l(x)}{\partial x^1} \right)_i r_{kl} + \left(\frac{\partial Z_k(x)}{\partial x^2} \right)_i \sum_{l=1}^{p} \left(\frac{\partial Z_l(x)}{\partial x^2} \right)_i r_{kl} \right\}$$

– *Poids de la $k^{ème}$ variable sur la barrière* : c_k / c.

La notion d'importance des barrières a fait émerger l'idée d'étudier les différentes barrières, en particulier leur structure et leur taille [Gleyze *et al.* (2001)].

L'introduction de l'inverse de la matrice de covariance dans la définition de la fonction systémique et l'importance des barrières offrent la perspective que seules les variables les plus influentes pourraient être prises en compte dans la détermination des ZCAs.

Gleyze, Bacro et Allard, 2001

La procédure de Womble détectant toujours des ZCAs, Gleyze *et al.* (2001) ont essayé de répondre au problème de savoir si les zones détectées pouvaient être expliquées par des fluctuations purement aléatoires ou si elles étaient réellement significatives. Pour se faire, ils ont développé le test d'hypothèses :

H_0 : la distribution des ZCAs est aléatoire,

contre H_1 : les ZCAs sont spatialement structurées.

Ils considèrent dans un premier temps une variable Z sans dépendance spatiale. Dans ce cas, un grand nombre de petites ZCAs est détecté et ces ZCAs ne font pas apparaître de distribution spatiale particulière. Inversement, lorsqu'il y a un petit nombre de grandes ZCAs, ils supposent l'existence d'un schéma particulier pour les fortes valeurs du gradient. Afin d'étudier la significativité de "grand nombre de petites ZCAs", ils proposent le test d'hypothèses H_0 contre H_1 cité ci-dessus. Lorsque le schéma n'est pas statistiquement différent de celui obtenu à partir d'une variable sans corrélation spatiale, les ZCAs sont considérées comme non significatives (*i.e.* ne présentent pas de structure spatiale). Leur procédure de test est basée sur la permutation des données par la méthode de Monte-Carlo.

Gleyze *et al.* (2001) définissent une statistique de test S. Pour S_{obs} la statistique associée au test unilatéral, calculée sur les données observées de Z, H_0 est rejetée au niveau α

lorsque $S_{obs} \geq F^{*\,-1}(1-\alpha)$, où F^* est la distribution empirique de S (la distribution F de S étant inconnue). En principe le calcul de F^* entraîne la réalisation des $M = N!$ permutations des valeurs de Z, et pour chaque permutation (indexée par m) l'évaluation de S_m. A partir de l'ensemble de valeurs $(\, S_m \,)_{m=1,...,M}$, F^* sous H_0 est définie par :

$$F^*(s) = \frac{1}{M} \text{Card } \{S_m \leq s,\ m = 1, ..., M\}\,.$$

D'un point de vue pratique, le nombre de permutations augmentant avec la taille de la grille, ils utilisent une approche Monte-Carlo qui consiste à tirer m permutations uniformes et indépendantes parmi les $N!$ possibles.

Lorsque Z est considéré comme stationnaire avec un variogramme donné, il faut prendre en compte cette structure dans le test. Le problème est qu'il est difficile de proposer des définitions précises de H_0 et de la significativité d'une ZCA contenant la dépendance spatiale. Aussi Gleyze *et al.* (2001) ont décidé de suivre la même approche que dans le cas précédent, où H_0 postule que la distribution des ZCAs est purement aléatoire. La procédure statistique reste la même.

Gleyze *et al.* (2001) proposent comme statistique de test la taille moyenne des ZCAs. D'autres statistiques pourraient être utilisées comme la taille de la plus grande ZCA ou la structure des différentes ZCAs.

La méthode proposée par Gleyze *et al.* (2001) présente l'originalité de vouloir tester la significativité des ZCAs détectées. Cependant leur méthode fait face à certaines limites et aux échecs suivants. Elle ne permet pas de séparer la structure des ZCAs (lorsqu'elle existe) de la distribution des variables. En effet, la méthode propose quatre situations possibles, schématisées dans le tableau 1.1. Dans leur cas il est difficile de placer H_1 et

	Dépendance spatiale	
ZCAs	absence	présence
sans	H_0	
avec		

TAB. 1.1: *Configurations de test possibles proposées par Gleyze et al. (2001).*

la procédure de permutations considère que H_1 est dans les trois cases blanches. De plus, Gleyze *et al.* (2001) ne sont pas en mesure d'expliquer l'origine des Zones de Changement Abrupt. Il ne leur est pas possible de dire si les ZCAs détectées caractérisent une fonction aléatoire stationnaire d'ordre deux, *i.e.* de moyenne constante sur le domaine, ou une transition entre deux zones dans un contexte non stationnaire.

1.2 Méthodes statistiques

Hall, Peng et Rau, 2001

Hall *et al.* (2001) proposent une méthode non paramétrique basée sur les approximations locales de la vraisemblance qu'une discontinuité existe en un point donné du plan, comme une fonction de ce point. La discontinuité est retrouvée par maximisation de la vraisemblance locale.

La méthodologie adoptée pour la détection de failles dans des surfaces de réponse est la suivante. Soient (X_i, Y_i), $1 \leq i \leq n$, des données observées. Les X_i représentent des vecteurs de dimension 2 représentant les variables exploratoires et les Y_i correspondent à des réponses univariées. La surface de réponse, d'équation $y = g(x)$ où $g(x) = \mathbb{E}[Y_1 \mid X_1 = x]$, est supposée lisse sauf le long d'une faille F (en général supposée contigüe et ne s'intersectant pas). Hall *et al.* (2001) définissent le test d'hypothèse nulle "g est localement constante en un point x du plan" versus l'alternative "x appartient à la faille F, avec F localement linéaire en x et ayant sa normale dans la direction du vecteur unité θ et g est localement constante de part et d'autre de F". Les erreurs $Y_i - g(X_i)$ sont supposées indépendantes, identiquement normalement distribuées et de variance connue fixée. Le rapport de la (log)-vraisemblance locale pondérée par noyau est :

$$w(x,\theta) = \frac{\sum_{i\in I^+} K_i(x) \sum_{i\in I^-} K_i(x)}{\sum_i K_i(x)} \left[\frac{\sum_{i\in I^+} Y_i K_i(x)}{\sum_{i\in I^+} K_i(x)} - \frac{\sum_{i\in I^-} Y_i K_i(x)}{\sum_{i\in I^-} K_i(x)} \right]^2,$$

avec $K_i(x) = K((X_i - x)/h)$ et $I^{\pm} = \{i : \pm\theta(X_i - x) \geq 0\}$. F est estimée comme étant une ligne de crête de la fonction non négative $w(x,\theta)$. Pour se faire, un point de départ, $\hat{x}^{(0)}$, est estimé par :

$$\hat{x}^{(0)} = \arg\max_x \left\{ \sup_\theta w(x,\theta) \right\}.$$

$\hat{\theta}$ est exclusivement déterminé par $\hat{\theta} = \arg\sup_\theta \{w(x,\theta)\}$, excepté pour son signe. Soit $\hat{\theta}_\perp$ le vecteur unité perpendiculaire à $\hat{\theta}$. Afin de tracer la faille, Hall *et al.* (2001) utilisent l'algorithme suivant. Pour $\hat{x}^{(j)}$ le point le plus récent estimé de F, ils posent $\hat{\theta}^{(j)} = \hat{\theta}_\perp(\hat{x}^{(j)})$ et définissent $\hat{x}^{(j+1)} = \hat{x}^{(j)} + \delta\hat{\theta}^{(j)}$ avec δ est très petit. Cette procédure est réalisée en partant de part et d'autre de $\hat{x}^{(0)}$.

Chaudhuri et Marron, 1999

Chaudhuri et Marron (1999) ont proposé une méthode, SiZer (SIgnificant ZERo crossings of derivatives), permettant de tester la significativité de structures (telles que des

sommets ou des vallées) détectées dans des données unidimensionnelles. SiZer est issu des questions soulevées par l'estimation de densité par noyaux, méthode qui met en avant l'existence de structures mais qui ne précise pas la significativité de chacune d'elles. SiZer détermine les lieux où la dérivée de l'estimation de la densité s'annule, évalue la significativité de chacun des zéro obtenus et propose une visualisation des régions de significativité.

SiZer est donc basé sur l'estimation d'une densité, f, par la méthode des noyaux :

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i),$$

où h est la largeur de la bande, $K_h(\cdot)$ est un noyau gaussien et $X_1, \ldots, X_n$ est l'échantillon de données. Leur approche consiste à visualiser la significativité des structures dans une famille $\left\{\hat{f}_h(x) : h \in [h_{min}, h_{max}]\right\}$, à partir des intervalles de confiance des dérivées $\hat{f}'_h(x)$. Le comportement en x et h est représenté par un code couleur : le noir (resp. blanc) correspond à "$\hat{f}'_h(x)$ est significativement positive (resp. négative)", et le gris à "$\hat{f}'_h(x)$ n'est pas significativement différente de 0". Les intervalles de confiance des dérivées $\hat{f}'_h(x)$ sont :

$$\hat{f}'_h(x) \pm q_\alpha \widehat{SD}\left(\hat{f}'_h(x)\right),$$

où q_α est un quantile approprié et $\widehat{SD}\left(\hat{f}'_h(x)\right)$ est l'estimation de l'écart-type de $\hat{f}'_h(x)$. Le quantile q_α est déterminé de la façon suivante. Notons η le taux d'erreurs de type I prédéfini et $m(h)$ le nombre de blocs indépendants de taille moyenne donnée par l'échantillon de taille n :

$$m(h) = \frac{n}{avg_x ESS(x,h)},$$

avec $ESS(x,h)$ (*Effective Sample Size*) le nombre de points de données dans la fenêtre centrée en x :

$$ESS(x,h) = \frac{\sum_{i=1}^{n} K_h(x - X_i)}{K_h(0)},$$

et $avg_x ESS(x,h)$ sa moyenne.
Alors,

$$\begin{aligned}
\eta &= \mathbb{P}\left[\cup_{i=1}^{m(h)} |X_i| > q_\alpha\right] = 1 - \mathbb{P}\left[\cap_{i=1}^{m(h)} |X_i| < q_\alpha\right] \\
&= 1 - \prod_{i=1}^{m(h)} \left(1 - \mathbb{P}\left[|X_i| > q_\alpha\right]\right) \\
&= 1 - \left(1 - \mathbb{P}\left[|X_1| > q_\alpha\right]\right)^{m(h)}, \text{ par indépendance des tests} \\
&= 1 - \left(1 - \left(1 - \mathbb{P}\left[X_1 < q_\alpha\right] + \mathbb{P}\left[X_1 < -q_\alpha\right]\right)\right)^{m(h)} \\
&= 1 - \left(2\Phi(q_\alpha) - 1\right)^{m(h)},
\end{aligned}$$

où Φ désigne la fonction de répartition d'une variable gaussienne standard. Ainsi,

$$q_\alpha = \Phi^{-1}\left(\frac{1+(1-\eta)^{1/m(h)}}{2}\right), \quad i.e. \quad \alpha = \frac{1-(1-\eta)^{1/m(h)}}{2}. \tag{1.3}$$

Godtliebsen, Marron, Pizer, 2002

L'approche proposée dans Godtliebsen *et al.* (2001, 2002), S^3 (Significance in Scale Space), est une généralisation de SiZer au cas bi-dimensionnel. L'étendue de la visualisation des régions de significativité au cas bi-dimensionnel n'est pas facile. La visualisation est donnée par des films où le temps correspond à la largeur de la bande et les images font apparaître des signaux indiquant la significativité des gradients locaux et des courbures locales.

Le modèle statistique est :

$$Y_{i,j} = s(i,j) + \epsilon_{i,j},$$

où $i = 1, \ldots, n$ et $j = 1, \ldots, m$ sont les indices des pixels de la grille, s représente le signal sous-jacent et ϵ est un bruit supposé indépendant. Le signal estimé est de la forme :

$$\hat{s}_h(i,j) = \sum_{i'=1}^{n}\sum_{j'=1}^{m} Y_{i',j'} K_h(i-i', j-j'), \tag{1.4}$$

où K_h est un noyau gaussien. L'équation (1.4) s'écrit en utilisant un autre système de notations : $\hat{s}_h = K_h * Y$. Une composante importante de S^3 est le nombre de points dans chaque fenêtre : $ESS = (K_h * \mathbf{1})/K_h(0,0)$, où $\mathbf{1}$ est la $n \times m$-matrice constituée de 1. Selon le développement effectué par Chaudhuri et Marron (1999), ESS peut être utilisé pour faire l'inférence de S^3 simultanément sur plusieurs sites et ils définissent :

$$\overline{ESS} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} ESS(i,j)}{nm}.$$

Puisqu'il y a nm points de mesure de $ESS(i,j)$, le processus de lissage peut être vu comme la moyenne dans des groupes de taille $\overline{ESS}$. Le nombre de moyennes indépendantes est approximativement :

$$l = \frac{nm}{\overline{ESS}}.$$

Gradient significatif

En un pixel $[i,j]$, la norme du gradient du signal sous-jacent s, $G(s)$, est estimée par :

$$\widehat{G}_h(s) = \left[\hat{s}_{h,1}^2 + \hat{s}_{h,2}^2\right]^{1/2},$$

où $\hat{s}_{h,1} = K_{h,1} * Y$ et $\hat{s}_{h,2} = K_{h,2} * Y$ représentent l'estimation des dérivées partielles. Un pixel est marqué comme significatif lorsque $\widehat{G}_h(s)$ est plus grand que le bruit, dans le sens que l'hypothèse de la forme $G(s) = 0$ est rejetée. La distribution de ce test est basée sur la distribution gaussienne bivariée :

$$\begin{pmatrix} \hat{s}_{h,1} \\ \hat{s}_{h,2} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_2^2 \end{pmatrix} \right).$$

Ainsi, l'hypothèse nulle est rejetée pour les pixels tels que $\hat{s}_{h,1}^2/\sigma_1^2 + \hat{s}_{h,2}^2/\sigma_2^2 > q_{\chi^2(2)}(\alpha)$, où $q_{\chi^2(2)}(\alpha)$ est le quantile approprié d'une $\chi^2(2)$. Le niveau α est choisi de manière à effectuer une inférence simultanée sur les pixels. Il est basé sur le "nombre de blocs indépendants" l. Notons R_k l'évènement "la k-ième hypothèse est rejetée" et $\overline{R}_k$ son complémentaire. Pour atteindre le niveau η simultanément sur l tests indépendants, il faut que :

$$\begin{aligned} \eta &= \mathbb{P}\left[\cup_{k=1}^{l} R_k\right] = 1 - \mathbb{P}\left[\cap_{k=1}^{l} \overline{R}_k\right] \\ &= 1 - \left(\mathbb{P}\left[\overline{R}_1\right]\right)^l, \text{ par indépendance des tests} \\ &= 1 - (1-\alpha)^l, \end{aligned}$$

Il en ressort que :

$$\alpha = 1 - (1-\eta)^{1/l}. \tag{1.5}$$

Courbure significative

En un pixel $[i, j]$ les propriétés de la courbure locale du signal s sont décrites par la matrice Hessienne :

$$H(s) = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}, \tag{1.6}$$

où s_{11} est la dérivée partielle seconde dans la direction verticale (indexée par i), s_{22} est la dérivée partielle seconde dans la direction horizontale (indexée par j) et s_{12} est la dérivée partielle seconde dans les deux directions. Ces dérivées secondes sont estimées par :

$$\hat{s}_{h,11} = K_{h,11} * Y, \quad \hat{s}_{h,12} = K_{h,12} * Y \text{ et } \hat{s}_{h,22} = K_{h,22} * Y. \tag{1.7}$$

Les valeurs propres de la matrice Hessienne résument le type et l'ampleur des courbures locales. Elles sont solution de :

$$\lambda_{\pm} = \left[(s_{11} + s_{22}) \pm \sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2} \right] /2. \tag{1.8}$$

Lorsque la courbure est significative, les pixels correspondants sont assignés comme significatifs. Cela se quantifie par le paramètre $T = \max\{|\lambda_-|, |\lambda_+|\}$ et est formulé par l'hypothèse : $H_0 : T = 0$. Une statistique de test est $\hat{T} = \max\left\{|\hat{\lambda}_-|, |\hat{\lambda}_+|\right\}$, où les valeurs propres estimées sont obtenues en remplaçant les dérivées secondes de (1.8) par leur version estimée (1.7). La distribution de cette statistique est basée sur la distribution tri-gaussienne :

$$\begin{pmatrix} \hat{s}_{h,11} \\ \hat{s}_{h,12} \\ \hat{s}_{h,22} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tilde{\sigma}^2_{11} & \tilde{\sigma}^2_{12} & \tilde{\sigma}^2_{13} \\ \tilde{\sigma}^2_{12} & \tilde{\sigma}^2_{22} & \tilde{\sigma}^2_{23} \\ \tilde{\sigma}^2_{13} & \tilde{\sigma}^2_{23} & \tilde{\sigma}^2_{33} \end{pmatrix}\right).$$

L'hypothèse nulle est rejetée au niveau α à un pixel donné si $\hat{T} > q_{\hat{T}}$, où $q_{\hat{T}}$ représente le quantile au niveau α de la distribution de $\hat{T}$. Comme dans ce qui précède, le niveau η fixé est atteint sur les l tests indépendants lorsque $\alpha = 1 - (1 - \eta)^{1/l}$.

Visualisation

La représentation des gradients significatifs et des courbures significatives se combine sur une seule image. A chaque pixel, lorsque le gradient est significatif une flèche est tracée en utilisant la direction du vecteur $(\hat{s}_{h,1}, \hat{s}_{h,2})'$. Si la courbure n'est pas significative, la flèche est colorée en vert. Si la courbure est significative, le code couleur utilisé est celui de la table 1.2.

Couleur	Type de structure	Caractérisation
jaune	trou	$\hat{\lambda}_+, \hat{\lambda}_- > q_{\hat{T}}$
orange	longue vallée	$\hat{\lambda}_+ > q_{\hat{T}}$, $\lvert\hat{\lambda}_-\rvert < q_{\hat{T}}$
rouge	point de selle	$\hat{\lambda}_+ > q_{\hat{T}}$, $\hat{\lambda}_- < -q_{\hat{T}}$
violet	longue crête	$\lvert\hat{\lambda}_+\rvert < q_{\hat{T}}$, $\hat{\lambda}_- < -q_{\hat{T}}$
bleu	sommet	$\hat{\lambda}_+, \hat{\lambda}_- < -q_{\hat{T}}$

TAB. 1.2: *Configurations des courbures en fonction des valeurs propres.*

Lorsque le gradient n'est pas significatif, mais que la courbure l'est, alors des points colorés sont tracés, toujours selon le code couleur de la table 1.2.

1.3 Problèmes résultants

D'une approche exploratoire des Zones de Changement Abrupt via la sélection des 5% des valeurs les plus élevées du gradient des variables considérées [Womble (1951)],

l'évolution du wombling a abouti à une analyse statistique dont les tests portent sur la significativité des ZCAs détectées. En effet, [Gleyze *et al.* (2001)] ont cherché à tester si les ZCAs détectées sont spatialement structurées.

Dans les méthodes décrites dans la section 1.1, les structures spatiales des variables n'ont pas été prises en compte dans la définition des ZCAs. Une des conséquences est que l'échantillonnage doit être régulier et suffisamment dense pour que le wombling soit pertinent. En effet, si l'interpolation est faite à partir de données éparses le wombling mettra en valeur la géométrie du schéma d'échantillonnage d'origine et ignorera les zones sous-représentées. En supposant que ce problème soit résolu, la construction du test de significativité proposée par Gleyze *et al.* (2001) reste à améliorer. Actuellement, l'hypothèse nulle est que la variable est sans dépendance spatiale, ce qui est une hypothèse très restrictive et qui ne reflète pas les phénomènes biologiques qui nous intéressent. Il est donc indispensable de pouvoir considérer l'existence de structures spatiales de la variable dans ce test, structures spatiales dont les variations ne doivent pas être prises en compte dans la définition des ZCAs.

Les méthodes proposées dans la section 1.2 présentent également des inconvénients non négligeables. La méthode non paramétrique proposée par Hall *et al.* (2001) nécessite un point de départ dans le voisinage de la discontinuité, ce qui suppose implicitement l'existence d'une courbe de discontinuité sur le domaine et par conséquent ne fournit pas de significativité sur les courbes détectées. De plus cette méthode n'intègre pas l'aspect spatial puisque les données sont supposées indépendantes et nécessite un grand nombre de données. L'étendue de SiZer au cas bi-dimensionnel [Godtliebsen *et al.* (2001, 2002)] nécessite de lourds calculs et offre une visualisation difficile à interpréter.

La méthode que nous proposons dans la deuxième partie répond à tous ces problèmes. Elle prend en compte les structures spatiales de la variable, n'impose pas de considérer un échantillon régulier et de forte densité et propose une visualisation facile à interpréter. La méthode ne permet de traiter qu'une variable à la fois. Cette variable est modélisée par un champ aléatoire que nous supposons gaussien. La méthode se décompose en deux étapes. Dans la première étape des ZCAs potentielles sont estimées. Elles ne décrivent plus un taux fixe (5%) des plus grandes valeurs du gradient, mais un taux, α, dépendant du schéma d'échantillonnage et de la fonction de covariance de la variable. Dans la deuxième étape, nous testons la significativité de chaque ZCA potentielle, à partir de sa taille. Nous testons si chacune d'elles est issue d'un champ aléatoire stationnaire ou présentant de fortes variations.

Les chapitres suivants ont pour objectif de rappeler quelques résultats théoriques sur lesquels repose la méthode de détection de ZCAs. Nous énonçons en particulier la définition de la stationnarité et les principales propriétés de continuité et de différentiabilité des champs aléatoires. Travaillant à partir d'une variable spatialisée, nous détaillons la méthode géostatistique d'interpolation spatiale utilisée pour prédire la variable sur l'ensemble du domaine d'étude.

Chapitre 2

Champs aléatoires

Afin de détecter des Zones de Changement Abrupt, nous supposons que le champ aléatoire considéré est un champ gaussien stationnaire, $Z(\cdot)$. Les tests mis en œuvre pour définir et tester les ZCAs sont liés au comportement d'un champ de χ^2 non stationnaire défini à partir du gradient et de la fonction de covariance de $Z(\cdot)$. La distribution de χ^2 provient du fait que nous considérons les carrés de champs gaussiens. La non stationnarité est issue de la géométrie des points d'échantillonnage, variable autour du point d'interpolation considéré. Nous allons dans ce chapitre définir la stationnarité et énoncer les propriétés de continuité et de différentiabilité de champs aléatoires stationnaires. Nous spécifierons également ce que sont des champs gaussiens et des champs de χ^2.

Commençons par définir un champ aléatoire. Soient $\mathcal{D}$ un sous ensemble de $\mathbb{R}^d$ et $(\Omega, \mathcal{B}, P)$ un espace de probabilité. Un champ aléatoire est une fonction de deux variables $Z(\mathbf{x}, \omega)$ telle que pour tout $\mathbf{x} \in \mathcal{D}$, $Z(\mathbf{x}, \cdot)$ est une variable aléatoire sur $(\Omega, \mathcal{B}, P)$. Chaque fonction $Z(\cdot, \omega)$ est une réalisation de la variable aléatoire. Pour simplifier les notations, dans la suite le champ aléatoire sera noté $Z(\cdot)$.

Définition 1 *Un champ aléatoire réel, $Z(\cdot)$, sur $\mathbb{R}^d$ est un ensemble constitué de variables aléatoires $Z(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, ayant des distributions de dimensions finies $F_{\mathbf{x}_1,\ldots,\mathbf{x}_n}$ de la forme :*

$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(B) = \mathbb{P}\left[(Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)) \in B\right],\ B \subseteq \mathbb{R}^n,\ \mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{R}^d,$$

et satisfaisant les conditions de Kolmogorov suivantes :

(i) Condition de symétrie :

Pour B de la forme $B = B_1 \times \cdots \times B_n$ ($B_i \subseteq \mathbb{R}$), considérons les permutations $i_1, \ldots, i_n$ des indices $1, \ldots, n$. Puisque "$(Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)) \in (B_1 \times \cdots \times B_n)$" et "$(Z(\mathbf{x}_{i_1}), \ldots, Z(\mathbf{x}_{i_n})) \in (B_{i_1} \times \cdots \times B_{i_n})$" sont les mêmes évènements :

$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(B_1 \times \cdots \times B_n) = F_{\mathbf{x}_{i_1},\ldots,\mathbf{x}_{i_n}}(B_{i_1} \times \cdots \times B_{i_n}).$$

(ii) Condition de consistence :

$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}}\left(B_1 \times \cdots \times B_{n-1}\right) = F_{\mathbf{x}_1,\ldots,\mathbf{x}_{n-1},\mathbf{x}_n}\left(B_1 \times \cdots \times B_{n-1} \times \mathbb{R}\right).$$

En géostatistique les variables aléatoires ne sont pas indépendantes. Afin de surmonter le problème de n'avoir qu'une unique réalisation d'un phénomène et l'absence d'indépendance, il est nécessaire d'avoir une abondance de données et d'établir certaines hypothèses. Les grandes hypothèses de la géostatistique sont la stationnarité et l'ergodicité.

2.1 Stationnarité

Un cas particulier de champ aléatoire est lorsque les distributions de dimensions finies sont invariantes par rapport à une translation arbitraire de vecteur $\mathbf{h}$ des points de données. Physiquement cela signifie que le phénomène est homogène dans l'espace.

Définition 2 *Un champ aléatoire $Z(\cdot)$ est dit strictement stationnaire si pour un vecteur $\mathbf{h}$ quelconque de $\mathbb{R}^d$,*

$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(z_1,\ldots,z_n) = F_{\mathbf{x}_1+\mathbf{h},\ldots,\mathbf{x}_n+\mathbf{h}}(z_1,\ldots,z_n),\ \forall \mathbf{h},\ \forall n,\ \forall \mathbf{x}_1,\ldots,\mathbf{x}_n \in \mathbb{R}^d,$$

avec $z_i = Z(\mathbf{x}_i),\ i = 1,\ldots,n.$

Lorsque le champ aléatoire est strictement stationnaire, ses moments (s'ils existent) sont invariants par translation. Considérons seulement les deux premiers moments : espérance $m(\mathbf{x})$ et covariance. Rappelons que la fonction de covariance d'un champ aléatoire $Z(\cdot)$ est la fonction $C_Z(\mathbf{x},\mathbf{y}) : \mathbb{R}^{2d} \to \mathbb{R}$ définie par :

$$C_Z(\mathbf{x},\mathbf{y}) = \mathrm{Cov}\left(Z(\mathbf{x}), Z(\mathbf{y})\right) = \mathbb{E}\left[\left(Z(\mathbf{x}) - m(\mathbf{x})\right)\left(Z(\mathbf{y}) - m(\mathbf{y})\right)\right].$$

Pour des points $\mathbf{x}$ et $\mathbf{x}+\mathbf{h}$ de $\mathbb{R}^d$, la stationnarité stricte entraîne les deux propriétés suivantes :

$$\begin{cases} \mathbb{E}\left[Z(\mathbf{x})\right] = m, \\ \mathbb{E}\left[\left(Z(\mathbf{x}) - m\right)\left(Z(\mathbf{x}+\mathbf{h}) - m\right)'\right] = C_Z(\mathbf{h}), \end{cases} \tag{2.1}$$

i.e. la moyenne est constante et la fonction de covariance dépend seulement de $\mathbf{h}$.

Définition 3 *Un champ aléatoire satisfaisant les conditions (2.1) est dit stationnaire d'ordre 2 (ou faiblement stationnaire).*

Un champ aléatoire stationnaire d'ordre 2 est dit isotrope si sa fonction de covariance ne dépend que de la longueur $\|\mathbf{h}\|$ du vecteur $\mathbf{h}$ et pas de ses directions. La covariance est une fonction paire et d'après l'inégalité de Schwarz elle est bornée par sa valeur à l'origine, *i.e.* par la variance du champ aléatoire :

$$C_Z(\mathbf{h}) = C_Z(-\mathbf{h}) \ ; \ |C_Z(\mathbf{h})| \leq C_Z(\mathbf{0}).$$

Plus précisément la covariance est une fonction définie positive. Comme le champ aléatoire a une variance finie, il fluctue autour de sa moyenne.

L'*hypothèse intrinsèque* de stationnarité des accroissements $Y_{\mathbf{h}}(\mathbf{x}) = Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x}) \ \forall \mathbf{h}$ ignore la propriété de stationnarité de la fonction. Elle suppose que les accroissements sont de moyenne nulle et que leur variance ne dépend que du vecteur entre les observations et non de leurs positions. Un champ aléatoire intrinsèque est caractérisé par les relations suivantes :

$$\begin{cases} \mathbb{E}\left[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})\right] = \mathbf{0}, \\ \mathrm{Var}\left(Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})\right) = 2\gamma(\mathbf{h}). \end{cases}$$

La fonction $\gamma(\mathbf{h})$ est appelée variogramme. L'estimateur classique du variogramme proposée par Matheron (1962) est :

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \left(Z(\mathbf{x}_i) - Z(\mathbf{x}_j)\right)^2, \tag{2.2}$$

où la somme est prise sur $N(\mathbf{h}) = \{(i,j) : \mathbf{x}_i - \mathbf{x}_j \simeq \mathbf{h}\}$ et $|N(\mathbf{h})|$ est le nombre d'éléments distincts de $N(\mathbf{h})$. Le variogramme montre comment les dissimilarités entre $Z(\mathbf{x}+\mathbf{h})$ et $Z(\mathbf{x})$ évoluent avec $\mathbf{h}$. Le variogramme est une fonction paire, non négative et vaut zéro à l'origine :

$$\gamma(\mathbf{h}) = \gamma(-\mathbf{h}) \ ; \ \gamma(\mathbf{h}) \geq \mathbf{0} \ ; \ \gamma(\mathbf{0}) = \mathbf{0}.$$

Un champ aléatoire stationnaire d'ordre 2 est un champ aléatoire intrinsèque et par conséquent possède un variogramme. Dans ce cas variogramme et covariance sont liés par la relation :

$$\gamma(\mathbf{h}) = C_Z(\mathbf{0}) - C_Z(\mathbf{h}). \tag{2.3}$$

Le variogramme d'un champ stationnaire d'ordre 2 est borné par $C_Z(\mathbf{0})$. La quantité $C_Z(\mathbf{0})$ est appelée *palier* du variogramme. L'équation (2.3) montre que si la covariance est connue, le variogramme l'est aussi. Inversement, si le variogramme d'un champ aléatoire intrinsèque est borné par une valeur finie, alors $\gamma(\mathbf{h})$ est de la forme (2.3) pour une

covariance stationnaire et le champ aléatoire intrinsèque ne diffère d'un champ stationnaire d'ordre 2 que par une constante aléatoire. Le variogramme a par conséquent un degré de généralité supérieur à celui de la covariance car la variance du champ aléatoire $Z(\cdot)$ n'existe pas nécessairement.

2.2 Ergodicité

L'ergodicité permet de travailler avec une réalisation unique d'un phénomène. Pour un champ aléatoire stationnaire $Z(\cdot)$ de moyenne m, de variance $\sigma^2 = C_Z(\mathbf{0})$ et de fonction de corrélation $\rho(\mathbf{h}) = C_Z(\mathbf{h})/C_Z(\mathbf{0})$, nous ne connaissons qu'une réalisation $\{z(\mathbf{x}) : \mathbf{x} \in \mathcal{D}\}$ de $Z(\cdot)$ sur le domaine $\mathcal{D}$ et les paramètres m, σ^2 et ρ sont inconnus. Un estimateur sans biais de m est :

$$Z(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} Z(\mathbf{x}) \, \mathrm{d}\mathbf{x},$$

où $|\mathcal{D}|$ représente le volume du domaine $\mathcal{D}$. La variance de cet estimateur est :

$$\mathrm{Var}\left(Z(\mathcal{D})\right) = \frac{\sigma^2}{|\mathcal{D}|^2} \int_{\mathcal{D}} \int_{\mathcal{D}} \rho(\mathbf{x} - \mathbf{y}) \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y}.$$

Un champ aléatoire $Z(\cdot)$ est ergodique si $\mathrm{Var}\left(Z(\mathcal{D})\right)$ tend vers zéro quand $\mathcal{D}$ tend vers l'infini :

$$\lim_{\mathcal{D} \to \infty} \mathbb{E}\left[\left(Z(\mathcal{D}) - m\right)^2\right] = 0, \tag{2.4}$$

ce qui signifie que $Z(\mathcal{D})$ converge vers m en moyenne quadratique. L'hypothèse d'ergodicité rend possible l'estimation de m à partir d'une seule réalisation du phénomène. En effet, regardons l'intégrale de l'espérance (2.4) :

$$\lim_{\mathcal{D} \to \infty} \int_{\Omega} \left[Z(\mathcal{D}, \omega) - m\right]^2 \, \mathrm{d}P(\omega) = \mathbf{0},$$

où ω désigne les réalisations du champ aléatoire. Si le champ aléatoire satisfait les conditions classiques de régularité, il est possible de permuter limite et intégrale :

$$\int_{\Omega} \lim_{\mathcal{D} \to \infty} \left[Z(\mathcal{D}, \omega) - m\right]^2 \, \mathrm{d}P(\omega) = \mathbf{0}. \tag{2.5}$$

(2.5) n'a lieu que si :

$$\lim_{\mathcal{D} \to \infty} \left[Z(\mathcal{D}, \omega) - m\right]^2 = 0 \text{ presque sûrement},$$

c'est-à-dire si :

$$\lim_{\mathcal{D} \to \infty} Z(\mathcal{D}, \omega) = m.$$

2.3 Continuité et différentiabilité

Il existe différentes sortes de convergences et donc de continuités ou de différentiabilités de champs aléatoires.
Un champ aléatoire $Z(\cdot)$ est *presque sûrement (p.s.) continu* en $\mathbf{x}^*$, si pour toute suite $(\mathbf{x}_n)_n$ telle que $\|\mathbf{x}_n - \mathbf{x}^*\| \to 0$ quand $n \to \infty$, nous avons $Z(\mathbf{x}_n) \xrightarrow{p.s.} Z(\mathbf{x}^*)$, *i.e.*

$$\mathbb{P}\left[\, \| Z(\mathbf{x}_n) - Z(\mathbf{x}^*) \|_{\underset{n\to\infty}{\longrightarrow}} 0 \right] = 1.$$

Dans la suite, nous allons utiliser l'information contenue dans la fonction de covariance pour étudier le comportement des champs aléatoires. Le comportement de la fonction covariance à l'origine est en particulier très informatif sur les propriétés de continuité de $Z(\cdot)$.
Comme la fonction de covariance est un moment d'ordre 2, nous préfèrerons utiliser la continuité et la différentiabilité en moyenne quadratique. Pour $(\mathbf{x}_n)_n$ une suite de points et $\mathbf{x}^*$ fixé dans $\mathbb{R}^d$ tel que $\|\mathbf{x}_n - \mathbf{x}^*\| \to 0$ quand $n \to \infty$, si $\mathbb{E}\left[\, \|Z(\mathbf{x}_n) - Z(\mathbf{x})\|^2 \,\right] \underset{n\to\infty}{\longrightarrow} 0$, alors $Z(\cdot)$ est *continu en moyenne quadratique.*

Théorème 1 (Adler, 1981) *Un champ aléatoire $Z(\cdot)$ est continu en moyenne quadratique en $\mathbf{x}^*$, si et seulement si sa fonction de covariance $(\mathbf{x}, \mathbf{y}) \mapsto C_Z(\mathbf{x}, \mathbf{y})$ est continue pour $\mathbf{x} = \mathbf{y} = \mathbf{x}^*$.*
Si $C_Z(\cdot, \cdot)$ est continue sur la diagonale, alors elle est continue partout.

Théorème 2 (Adler, 1981) *Si $\partial^2 C_Z(\mathbf{x}, \mathbf{y})/\partial x^i \partial y^i$ existe et est finie au point $(\mathbf{x}, \mathbf{x}) \in \mathbb{R}^{2d}$, alors :*

$$\frac{\partial Z(\mathbf{x})}{\partial x^i} = \lim_{\epsilon \to 0} \frac{Z(\mathbf{x} + \epsilon \delta_i) - Z(\mathbf{x})}{\epsilon} \tag{2.6}$$

existe et est appelée dérivée en moyenne quadratique de $Z(\cdot)$ en $\mathbf{x}$.
Si cette dérivée existe $\forall \mathbf{x} \in \mathbb{R}^d$, alors $Z(\cdot)$ possède une dérivée en moyenne quadratique. La fonction de covariance de $\partial Z(\mathbf{x})/\partial x^i$ est $\partial^2 C_Z(\mathbf{x}, \mathbf{y})/\partial x^i \partial y^i$.

Par conséquent pour un champ stationnaire d'ordre 2, si $C_Z(.)$ est dérivable deux fois à l'origine, alors Z est dérivable en moyenne quadratique en tout point.
Lorsque le champ aléatoire est stationnaire d'ordre 2, les conditions assurant la continuité en moyenne quadratique et l'existence de dérivées en moyenne quadratique deviennent simples. D'après le théorème 1, $Z(\cdot)$ est continu en moyenne quadratique en $\mathbf{x}$ si sa fonction de covariance est continue en $(\mathbf{x}, \mathbf{x})$. A cause de la stationnarité d'ordre 2, c'est

équivalent à demander la continuité de la fonction de covariance en $\mathbf{x} - \mathbf{x} = \mathbf{0}$. Ainsi, si $C_Z(\cdot)$ est continue à l'origine alors le champ $Z(\cdot)$ est continu en moyenne quadratique :

$$\lim_{\mathbf{h}\to\mathbf{0}} C_Z(\mathbf{h}) = C_Z(\mathbf{0}) \Rightarrow \lim_{\mathbf{h}\to\mathbf{0}} \mathbb{E}\left[(Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x}))^2\right] = 0, \ \forall \mathbf{x} \in \mathbb{R}^d.$$

Comme la convergence en moyenne quadratique n'implique pas la convergence presque sûre, la continuité en moyenne quadratique d'un champ aléatoire n'implique pas la continuité de ses réalisations. Notons également que si $C_Z(.)$ n'est pas continue à l'origine, alors $Z(\cdot)$ ne sera pas continu partout en moyenne quadratique.

Des propriétés similaires sont associées au variogramme. La continuité à l'origine de $\gamma(\cdot)$ est équivalente à la continuité de $Z(\cdot)$ en moyenne quadratique :

$$\mathbb{E}\left[(Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x}))^2\right] \to 0 \text{ si et seulement si } \gamma(\mathbf{h}) \to 0 \text{ quand } \|\mathbf{h}\| \to 0.$$

Si $\gamma(\mathbf{h})$ ne tend pas vers 0 quand $\mathbf{h}$ est proche de l'origine, alors $Z(\cdot)$ n'est pas continu en moyenne quadratique et est irrégulier. La discontinuité de $\gamma(\cdot)$ à l'origine est appelée *effet de pépite*. Si $\gamma(\mathbf{h})$ est une constante positive (sauf à l'origine où il est nul), alors $Z(\mathbf{x}_i)$ et $Z(\mathbf{x}_j)$ sont non corrélés pour tout $\mathbf{x}_i \neq \mathbf{x}_j$ sans se soucier de leur proximité. Dans ce cas $Z(\cdot)$ est souvent appelé *bruit blanc*.

Ainsi, pour un champ aléatoire stationnaire d'ordre 2, le comportement de sa fonction de covariance dans un voisinage de l'origine est un facteur très informatif sur les propriétés locales du champ (continuité, différentiabilité) en moyenne quadratique.

Extremum d'une fonction

Considérons maintenant la fonction déterministe : $F : \mathbb{R}^d \to \mathbb{R}$, deux fois différentiable. Un point $\mathbf{x}_0$ tel que toutes les dérivées partielles sont nulles est un *point critique* de F : $\frac{\partial F}{\partial x^1}(\mathbf{x}_0) = \cdots = \frac{\partial F}{\partial x^d}(\mathbf{x}_0) = 0$.

Théorème 3 *Soient F définie sur $\mathbb{R}^2$, deux fois différentiable et $\mathbf{x}_0$ un point critique de F. Pour,*

$$r = \frac{\partial^2 F}{\partial (x^1)^2}(\mathbf{x}_0), \quad s = \frac{\partial^2 F}{\partial x^1 \partial x^2}(\mathbf{x}_0), \quad t = \frac{\partial^2 F}{\partial (x^2)^2}(\mathbf{x}_0),$$

- *si $s^2 - rt < 0$, alors $\mathbf{x}_0$ est un extremum : minimum pour $r > 0$ (figure 2.1a) et maximum pour $r < 0$ (figure 2.1b),*
- *si $s^2 - rt > 0$, alors $\mathbf{x}_0$ est un col (figure 2.1c),*
- *si $s^2 - rt = 0$, il faut chercher le signe de $F(\mathbf{x}) - F(\mathbf{x}_0)$ au voisinage de $\mathbf{x}_0$.*

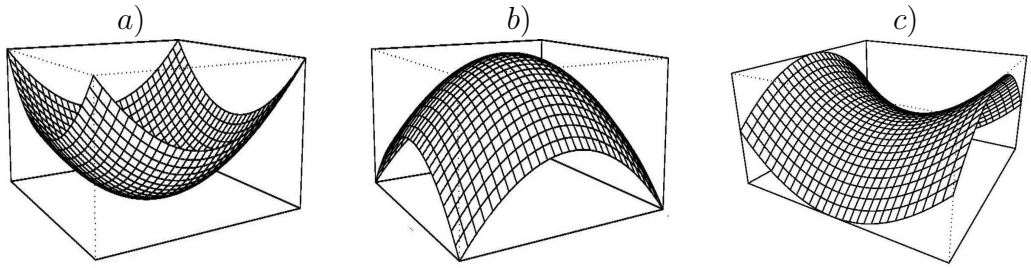


FIG. 2.1: *Illustration d'un minimum a), d'un maximum b) et d'un point col c).*

2.4 Champ aléatoire conditionnel

Afin d'étudier dans la section 4.3 et dans le chapitre 6 la structure d'un maximum local d'un champ aléatoire, il est nécessaire d'introduire la notion de champ conditionnel. Pour un champ aléatoire stationnaire, $Z(\cdot)$, nous noterons $Z_z(\cdot)$ le champ conditionnel :

$$Z_z(\mathbf{x}) = Z(\mathbf{x}) \mid A,\ \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d, \tag{2.7}$$

où A dénote l'événement "$Z(\cdot)$ possède un maximum local de hauteur z en $\mathbf{x} = \mathbf{0}$". Pour étudier les propriétés du champ $Z(\cdot)$ sachant qu'il possède à l'origine un maximum de hauteur z, nous approcherons cet événement conditionnel par une série d'évènements :

$$\begin{aligned} A(h, h') \quad = \quad & Z(\cdot) \text{ a un maximum local de hauteur dans } (z, z+h) \\ & \text{en des points } \mathbf{y} \text{ de la sphère } \|\mathbf{x}\| \leq h'. \end{aligned}$$

Pour B un sous-ensemble quelconque de la tribu $\mathcal{B}$, $Z_z(\cdot)$ a pour distribution :

$$\mathbb{P}\left[Z_z(\mathbf{x}) \in B\right] = \lim_{h \to 0} \lim_{h' \to 0} \mathbb{P}\left[Z_z(\mathbf{x}) \in B \ \mid \ A(h, h')\right].$$

L'événement A est donc obtenu en faisant tendre h et h' vers 0, voir par exemple Adler (1981) pour plus de précisions.
Ce conditionnement est aussi appelé conditionnement au sens ergodique ou par fenêtre horizontale (*Horizontal Window conditioning* ou *Slepian model process*). Nous noterons $\|$ le conditionnement au sens ergodique et $\mid$ le condionnement au sens usuel.

Le conditionnement au sens usuel sera souvent utilisé par la suite, aussi nous allons en rappeler quelques propriétés importantes, en particulier portant sur l'espérance conditionnelle. Nous nous limiterons à l'aspect intuitif de ces notions et nous occulterons les notations très concises liées à la théorie de la mesure dont elles sont issues.

Soient $\mathbf{X} = (X_1, \ldots, X_n)'$ et $\mathbf{Y} = (Y_1, \ldots, X_d)'$ deux champs aléatoires multivariables sur $\mathbb{R}^n$ et $\mathbb{R}^d$ respectivement, de densité $f_{\mathbf{X}}$ et $f_{\mathbf{Y}}$. Sous l'hypothèse d'existence d'une densité jointe $f_{\mathbf{X},\mathbf{Y}}$, l'espérance conditionnelle de $\mathbf{Y}$ sachant $\mathbf{X}$ est :

$$\mathbb{E}[\mathbf{Y} \mid \mathbf{X}] = [\mathbb{E}[Y_1 \mid \mathbf{X}], \ldots, \mathbb{E}[Y_d \mid \mathbf{X}]] \text{ où } \mathbb{E}[Y_i \mid \mathbf{X} = \mathbf{x}] = \int_{\mathbb{R}} y f_{Y_i|\mathbf{X}=\mathbf{x}}(y)\, \mathrm{d}y,$$

avec $f_{Y_i|\mathbf{X}=\mathbf{x}}(y) = f_{Y_i,\mathbf{X}}(y, \mathbf{x})/f_{\mathbf{X}}(\mathbf{x})$ et $\mathbf{x} \notin N$ où N est l'évènement de probabilité nulle : $N = \left\{\mathbf{x} \in \mathbb{R}^d : f_{\mathbf{X}}(\mathbf{x}) = 0\right\}$.
Pour Y un champ aléatoire monovariable, l'espérance conditionnelle $\mathbb{E}[Y \mid \mathbf{X}]$ est une variable aléatoire réelle définie partout sauf sur l'évènement de probabilité nulle N.
Les principales propriétés de l'espérance conditionnelle sont les suivantes :

(i) $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid \mathbf{X}]]$,

(ii) $\mathbb{E}[\{Y - \mathbb{E}[Y \mid \mathbf{X}]\} \mid h(\mathbf{X})] = 0, \ \forall\ h(\mathbf{X})$ de carré intégrable,

(iii) $\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid \mathbf{X})] + \mathrm{Var}(\mathbb{E}[Y \mid \mathbf{X}])$.

2.5 Champs gaussiens

Une variable aléatoire $Z(\mathbf{x})$ est dite gaussienne si son espérance $m = \mathbb{E}[Z(\mathbf{x})]$ et sa variance $\sigma^2 = \mathbb{E}\left[(Z(\mathbf{x}) - m)^2\right]$ telle que $0 < \sigma^2 < \infty$ sont finies et si sa loi marginale est :

$$f_Z(z) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{(z-m)^2}{\sigma^2}\right\}.$$

Nous résumerons en écrivant : $Z \sim \mathcal{N}(m, \sigma^2)$. Une variable gaussienne est dite centrée lorsque $m = 0$ et réduite lorsque $\sigma^2 = 1$.

Un vecteur aléatoire $\mathbf{Z}$ sur $\mathbb{R}^N$ est multigaussien (ou suit une loi multinormale) si pour tout N-uplet de nombres réels $(\alpha_1, ..., \alpha_N)$, $\mathcal{Z} = \sum_{i=1}^N \alpha_i Z_i$ est une variable gaussienne. Dans ce cas, la densité du vecteur $\mathbf{Z}$ est donnée par :

$$f_{\mathbf{Z}}(\mathbf{z}) = (2\pi)^{-N/2}\det(\mathbf{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{z}-\mathbf{m})'\mathbf{\Sigma}^{-1}(\mathbf{z}-\mathbf{m})\right\} \tag{2.8}$$

où $\mathbf{m}$ est le N-vecteur d'éléments $m_i = \mathbb{E}[Z_i]$ et $\mathbf{\Sigma}$ est la $N \times N$-matrice de covariance, définie non-négative, d'éléments $\sigma_{ij}^2 = \mathbb{E}[(Z_i - m_i)(Z_j - m_j)]$.

Propriétés :

(i) Tout sous-vecteur d'un vecteur gaussien est un vecteur gaussien. En particulier les composantes d'un vecteur gaussien sont des variables aléatoires gaussiennes.
Pour $\mathbf{\Sigma}^{1/2}$ une racine carrée de la matrice $\mathbf{\Sigma}$:

(ii) $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{Id}) \Rightarrow \mathbf{m} + \mathbf{\Sigma}^{1/2}\mathbf{Z} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$, où $\mathbf{Id}$ est la matrice identité.

(iii) Si $\mathbf{Z} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ avec $\mathbf{\Sigma}$ symétrique, définie positive, alors $\mathbf{\Sigma}^{-1/2}(\mathbf{Z} - \mathbf{m}) \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$.

(iv) Si $\mathbf{Z}$ est un vecteur gaussien, une condition nécessaire et suffisante pour que ses composantes soient des variables aléatoires indépendantes est que $\mathbf{\Sigma}$ soit diagonale.

Si $\mathbf{Z}$ est partitionné en $(\mathbf{Z}_1, \mathbf{Z_2})' = ((Z_1, \ldots, Z_n), (Z_{n+1}, \ldots, Z_N))'$, $\mathbf{m}$ en $(\mathbf{m}_1, \mathbf{m}_2)' = ((m_1, \ldots, m_n), (m_{n+1}, \ldots, m_N))'$ et $\mathbf{\Sigma}$ en :

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$$

où $\mathbf{\Sigma}_{11}$ est une matrice de dimension $n \times n$, alors chaque $\mathbf{Z}_i$ a pour moyenne $\mathbf{m}_i$ et pour matrice de covariance $\mathbf{\Sigma}_{ii}$. De plus, la distribution conditionnelle de $\mathbf{Z}_1$ sachant $\mathbf{Z}_2$ est aussi gaussienne, de moyenne :

$$\mathbf{m}_{1|2} = \mathbf{m}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\mathbf{Z}_2 - \mathbf{m}_2\right), \tag{2.9}$$

et de matrice de covariance :

$$\mathbf{\Sigma}_{1|2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}. \tag{2.10}$$

Il est maintenant possible de définir un champ aléatoire gaussien. Il s'agit d'un champ aléatoire dont toutes les distributions de dimensions finies sont multigaussiennes. A partir de cette définition et des commentaires précédents, il est clair que toute distribution de dimension finie d'une variable gaussienne est complètement déterminée si l'on connaît les deux fonctions suivantes :

- sa moyenne : $\mathbf{m}(\mathbf{x}) = \mathbb{E}\left[Z(\mathbf{x})\right]$,
- sa covariance : $\mathbf{C}_Z(\mathbf{x}, \mathbf{y}) = \mathbb{E}\left[\left(Z(\mathbf{x}) - \mathbf{m}(\mathbf{x})\right)\left(Z(\mathbf{y}) - \mathbf{m}(\mathbf{y})\right)\right]$.

D'après la forme de la densité multinormale (2.8), nous voyons que si un champ gaussien est de moyenne constante et de fonction de covariance dépendante seulement de $\mathbf{x} - \mathbf{y}$, alors le champ est stationnaire d'ordre 2.

Proposition 1 *Si $Z(\cdot)$ est un champ aléatoire gaussien possédant une dérivée en moyenne quadratique, alors $\partial Z(\cdot)/\partial x^i$ est aussi un champ aléatoire gaussien.*

Cela vient directement du théorème 2 et du fait qu'un vecteur aléatoire est gaussien si et seulement si toute combinaison linéaire de ses composantes est une variable aléatoire gaussienne.

Le lemme suivant donne une représentation des dérivées premières et secondes d'un champ gaussien en terme de variables aléatoires indépendantes.

Lemme 1 (Worlsey, 1994) *Soit $Z(\mathbf{x})$ un champ gaussien stationnaire d'ordre 2.*

(i) $\partial Z(\mathbf{x})/\partial \mathbf{x} \sim \mathcal{N}_d(0, \mathbf{\Lambda})$ indépendamment de $Z(\mathbf{x})$ et de $\partial^2 Z(\mathbf{x})/\partial \mathbf{x} \partial \mathbf{x}'$,

où $\mathbf{\Lambda} = Var(\partial Z(\mathbf{x})/\partial \mathbf{x})$ est la $d \times d$-matrice de covariance des dérivées partielles de $Z(\mathbf{x})$, d'éléments λ_{kl}, avec $\lambda_{kl} = Cov\left(\partial Z(\mathbf{x})/\partial x^k, \partial Z(\mathbf{x})/\partial x^l\right)$, $k, l = 1, \ldots, d$.

(ii) Condionnellement à $Z(\mathbf{x})$,

$$\frac{\partial^2 Z(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \Big| Z(\mathbf{x}) \sim \mathcal{N}_{d \times d}\left(-Z(\mathbf{x})\mathbf{\Lambda}, \mathbf{M}(\mathbf{\Lambda})\right),$$

où les éléments de $\mathbf{M}(\mathbf{\Lambda})$ sont tels que :

$$Cov\left(\frac{\partial^2 Z(\mathbf{x})}{\partial x^i \partial x^j}, \frac{\partial^2 Z(\mathbf{x})}{\partial x^k \partial x^l} \Big| Z(\mathbf{x})\right) = \varepsilon(i,j,k,l) - \lambda_{ij}\lambda_{kl},$$

avec $\varepsilon(i,j,k,l)$ symétrique dans ses arguments.

Si $Z(\mathbf{x})$ est de plus de variance unité, la distribution de $Z(\mathbf{x})$ conditionnelle à "$Z(\mathbf{0}) = z$, $\partial Z(\mathbf{0})/\partial \mathbf{x}$, $\partial^2 Z(\mathbf{0})/\partial \mathbf{x} \partial \mathbf{x}'$" a été déterminée dans Adler (1981). Il a essentiellement été montré le lemme suivant.

Lemme 2 (Adler, 1981) *Soit $Z(\mathbf{x})$ un champ gaussien stationnaire d'ordre 2. Conditionnellement à $Z(\mathbf{0})$, $\partial Z(\mathbf{0})/\partial \mathbf{x}$, $\partial^2 Z(\mathbf{0})/\partial \mathbf{x} \partial \mathbf{x}'$, $Z(\mathbf{x})$ a la même distribution que :*

$$Z(\mathbf{0}) + \mathbf{x}' \frac{\partial Z(\mathbf{0})}{\partial \mathbf{x}} + \frac{1}{2}\mathbf{x}' \frac{\partial^2 Z(\mathbf{0})}{\partial \mathbf{x} \partial \mathbf{x}'} \mathbf{x} + Mo\left(\|\mathbf{x}\|^2\right) + \epsilon(\mathbf{x}),$$

où $M = \max\left(\|Z(\mathbf{0})\|, \|\partial Z(\mathbf{0})/\partial \mathbf{x}\|, \|\partial^2 Z(\mathbf{0})/\partial \mathbf{x} \partial \mathbf{x}'\|\right)$ et $\epsilon(\mathbf{x})$ est de loi normale centrée et de covariance $o\left(\|\mathbf{x}\|^4\right)$.

Ce résultat permet de conclure que le champ gaussien $Z(\mathbf{x})$ conditionnel à $Z(\mathbf{0})$, $\partial Z(\mathbf{0})/\partial \mathbf{x}$, $\partial^2 Z(\mathbf{0})/\partial \mathbf{x} \partial \mathbf{x}'$ peut être représenté par $Z(\mathbf{0}) + \mathbf{x}\frac{\partial Z(\mathbf{0})}{\partial \mathbf{x}} + \frac{1}{2}\mathbf{x}'\frac{\partial^2 Z(\mathbf{0})}{\partial \mathbf{x} \partial \mathbf{x}'}\mathbf{x} + Mo\left(\|\mathbf{x}\|^2\right) + \epsilon(\mathbf{x})$ où $\epsilon(\mathbf{x}) = o\left(\|\mathbf{x}\|^2\right)$ presque sûrement.

2.6 Champs de χ^2

Pour construire un champ de χ^2, Adler (1981) considère p champs gaussiens, $Z_1(\mathbf{x}), \ldots, Z_p(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, indépendants et stationnaires. Supposons que chaque $Z_i(\cdot)$ soit de moyenne nulle et que tous les $Z_i(\cdot)$ aient la même fonction de covariance, $C_Z(\mathbf{h})$, avec $C_Z(\mathbf{0}) = \sigma^2$. A partir de ces champs, nous définissons le champ $Y(\mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{R}^d$:

$$Y(\mathbf{x}) = [Z_1(\mathbf{x})]^2 + \cdots + [Z_p(\mathbf{x})]^2.$$

Ce champ est appelé champ de χ^2 à p degrés de liberté. Comme les $Z_i(\cdot),\ i = 1, ..., p$ sont strictement stationnaires, il en va de même pour le champ $Y(\cdot)$. Sa densité marginale est de la forme :

$$f_Y(y) = \frac{1}{\sigma^p 2^{p/2}\Gamma\left(p/2\right)} y^{(p-2)/2} \exp\left(-\frac{y}{2\sigma^2}\right),\ y \geq \mathbf{0}, \tag{2.11}$$

où $\Gamma(x)$ est la fonction gamma définie par : $\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)\,\mathrm{d}t$.
Etudions maintenant la structure de la fonction de covariance d'un champ de χ^2. Comme conséquence de (2.11), si $Y(\cdot)$ est un champ de χ^2 de paramètre p, alors pour tout $\mathbf{x} \in \mathbb{R}^d$:

$$\mathbb{E}\left[Y(\mathbf{x})\right] = p \text{ et } \mathrm{Var}(Y(\mathbf{x})) = 2p.$$

Proposition 2 (Adler, 1981) *Pour $C_Z(\mathbf{x}, \mathbf{y})$ la fonction de covariance commune des $Z_i(\cdot)$, la fonction de covariance, $C_Y(\mathbf{x}, \mathbf{y})$, du champ $\chi^2(p)$, $Y(\cdot)$, défini précédemment est :*

$$C_Y(\mathbf{x}, \mathbf{y}) = 2p\left[C_Z(\mathbf{x}, \mathbf{y})\right]^2.$$

La distribution de $Y(\cdot)$, ainsi que toutes ses propriétés statistiques, sont entièrement déterminées lorsque p et $C_Y(\cdot)$ (ou de façon équivalente $C_Z(\cdot)$) sont connus.
Si les champs gaussiens $Z_i(\cdot),\ i = 1, \ldots, p$, sont deux fois différentiable, *i.e.* si la fonction de covariance $C_Z(.)$ est quatre fois différentiable, alors le champ de χ^2, $Y(\cdot)$, est deux fois différentiable. Ses dérivées partielles premières et secondes sont les suivantes :

$$\frac{\partial Y(\mathbf{x})}{\partial x^i} = 2\sum_{k=1}^{p} Z_k(\mathbf{x})\frac{\partial Z_k(\mathbf{x})}{\partial x^i},\ \text{pour } i = 1, \ldots, d. \tag{2.12}$$

$$\frac{\partial^2 Y(\mathbf{x})}{\partial x^i \partial x^j} = 2\sum_{k=1}^{p} \frac{\partial Z_k(\mathbf{x})}{\partial x^i}\frac{\partial Z_k(\mathbf{x})}{\partial x^j} + 2\sum_{k=1}^{p} Z_k(\mathbf{x})\frac{\partial^2 Z_k(\mathbf{x})}{\partial x^i \partial x^j},\ \text{pour } i, j = 1, \ldots, d. \tag{2.13}$$

D'après les équations (2.12) et (2.13), $\partial Y(\mathbf{x})/\partial x^i$ et $\partial^2 Y(\mathbf{x})/\partial x^i \partial x^j$ sont continues si $\partial Z_k(\mathbf{x})/\partial x^i$ et $\partial^2 Z_k(\mathbf{x})/\partial x^i \partial x^j$ le sont, ce qui peut être déterminé à partir de $C_Z(\cdot)$ (ou $C_Y(\cdot)$) d'après les paragraphes précédents.

La somme des carrés de p variables gaussiennes $Z_i \sim \mathcal{N}(\mu_i, 1), i = 1, \ldots, p$, a pour distribution une χ^2 décentrée avec p degrés de liberté et de paramètre de décentrage $\mu = \sum_{i=1}^{p} \mu_i^2$. Si $Y(\cdot)$ représente un tel champ, alors pour tout $\mathbf{x} \in \mathbb{R}^d$:

$$\mathbb{E}\left[Y(\mathbf{x})\right] = p + \mu \text{ et } \mathrm{Var}(Y(\mathbf{x})) = 2(p + 2\mu).$$

Conclusion

Nous avons dans ce chapitre introduit la notion de champ aléatoire et énoncé les principales propriétés de continuité et de différentiabilité de ces champs. Nous consacrons le chapitre suivant à la description du krigeage, procédé de prédiction en géostatistique du champ aléatoire en tout point du domaine d'étude.

Chapitre 3

Géostatistique

La signification originelle du préfixe “géo” du terme géostatistique réfère aux statistiques relatives aux sciences de la terre. Hart (1954) inventa ce terme dans un contexte géographique afin de désigner les techniques statistiques qui prennent en compte l'emplacement des sites dans les surfaces de distribution. Matheron (1963) utilisa ce terme dans un contexte géologique pour décrire les méthodes inférantes aux réserves de minerai à partir de données distribuées spatialement dans une galerie. Selon Cressie (1993), la géostatistique s'est aujourd'hui détachée de ce qui la raccrochait aux sciences de la terre et a désormais un enjeu universel dans la théorie statistique pour des processus définis dans un contexte spatial.

Une des questions principales posées dans le contexte spatial est celle de l'interpolation en tout point à partir de mesures en quelques points donnés. Pour un champ aléatoire $Z(\cdot)$, la géostatistique propose une méthode, le krigeage, permettant de prédire la valeur du champ en un point arbitraire en utilisant les valeurs prises par les points de données. Nous allons dans cette partie définir différentes formes de krigeage et les propriétés associées à ces prédicteurs, puis nous définirons les modèles de covariances usuels en géostatistique. Nous référons le lecteur à Chilès et Delfiner (1999) et Cressie (1993) pour de plus amples détails de ces notions.

3.1 Krigeage

De façon générale le krigeage est un prédicteur de $Z(\mathbf{x}_0)$ devant vérifier les quatre conditions suivantes :

– une *contrainte de linéarité* : le prédicteur est une combinaison linéaire des données

disponibles,
- une *contrainte d'autorisation* : l'erreur de prédiction doit être autorisée pour que les manipulations à l'ordre 2 soient licites (cette contrainte n'existe que pour le cas intrinsèque et correspond à une contrainte sur la somme des pondérateurs),
- une *contrainte de non biais* : le prédicteur doit être sans biais,
- et une *contrainte de variance minimale* : la variance de l'erreur de prédiction doit être minimale.

Le krigeage est donc le meilleur prédicteur linéaire sans biais ou BLUP (*Best Linear Unbiased Predictor*) au sens des moindres carrés.

Nous adopterons les notations suivantes :
- $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$: ensemble des points de données,
- $Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)$: les données,
- $\mathbf{x}_0$: point où la prédiction doit être faite,
- $Z^*(\mathbf{x}_0)$: prédicteur de krigeage,
- $\mathbf{C} = (C_Z(\mathbf{x}_i, \mathbf{x}_j))_{i,j}$: matrice de covariance entre les données,
- $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$: vecteur de poids où λ_i est un poids associé à $Z(\mathbf{x}_i)$ dépendant de l'emplacement de $\mathbf{x}_0$.

Krigeage simple

Nous considérons dans ce paragraphe que le champ $Z(\cdot)$ est stationnaire d'ordre 2 et que sa moyenne est connue. Connaître la moyenne rend la théorie beaucoup plus simple et fournit au prédicteur de krigeage de bonnes propriétés. Par exemple, dans le cas d'un champ gaussien, il coïncide avec l'espérance conditionnelle $\mathbb{E}\left[Z(\mathbf{x}_0) \mid Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)\right]$ qui est le meilleur prédicteur de $Z(\mathbf{x}_0)$ au sens des moindres carrés. Les propriétés établies dans le cas de la moyenne connue peuvent être considérées comme des propriétés limites.

La moyenne m étant supposée connue, nous travaillons sur la variable centrée. Le prédicteur de krigeage s'écrit :

$$Z^*(\mathbf{x}_0) - m = \sum_{i=1}^{n} \lambda_i (Z(\mathbf{x}_i) - m).$$

Le prédicteur est dit sans biais si l'erreur de prédiction, $Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)$, est d'espérance

nulle. Appliqué au krigeage simple cela donne :

$$\begin{aligned}\mathbb{E}\left[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)\right] &= m + \sum_{i=1}^{n} \lambda_i \left(\mathbb{E}\left[Z(\mathbf{x}_i)\right] - m\right) - \mathbb{E}\left[Z(\mathbf{x}_0)\right] \\ &= m + \sum_{i=1}^{n} \lambda_i (m - m) - m \\ &= 0.\end{aligned}$$

La prédiction est sans biais quels que soient les poids λ_i, donc il n'y a pas de contrainte de non biais.
La variance de l'erreur de prédiction est donnée par :

$$\begin{aligned}\sigma_E^2 &= \mathrm{Var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) = \mathbb{E}\left[\left(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)\right)^2\right] \\ &= \mathbb{E}\left[(Z^*(\mathbf{x}_0))^2 + (Z(\mathbf{x}_0))^2 - 2Z^*(\mathbf{x}_0)Z(\mathbf{x}_0)\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) + C_Z(\mathbf{x}_0 - \mathbf{x}_0) - 2\sum_{i=1}^{n} \lambda_i C_Z(\mathbf{x}_i - \mathbf{x}_0).\end{aligned}$$

La variance estimée est minimale quand toutes les dérivées premières s'annulent :

$$\frac{\partial \sigma_E^2}{\partial \lambda_i} = 2\sum_{j=1}^{n} \lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) - 2C_Z(\mathbf{x}_i - \mathbf{x}_0) = 0,\ \forall i = 1, \ldots, n,$$

et la matrice des dérivées secondes est définie positive, ce qui est le cas d'après la propriété de positivité de la fonction de covariance (section 2.1). Les λ_i sont donc solutions des équations de krigeage simple :

$$\sum_{j=1}^{n} \lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) = C_Z(\mathbf{x}_i - \mathbf{x}_0),\ \forall i = 1, \ldots, n. \tag{3.1}$$

La partie gauche de l'équation décrit les covariances entre les points. La partie droite décrit les covariances entre chaque point et le point où la prédiction est faite. Une seconde quantité d'intérêt est la variance du krigeage simple. Elle est obtenue en substituant la partie gauche du système de krigeage par sa partie droite dans le premier terme de l'expression de la variance σ_E^2 :

$$\begin{aligned}\sigma_{KS}^2 &= \sum_{i=1}^{n} \lambda_i C_Z(\mathbf{x}_i - \mathbf{x}_0) + C_Z(\mathbf{x}_0 - \mathbf{x}_0) - 2\sum_{i=1}^{n} \lambda_i C_Z(\mathbf{x}_i - \mathbf{x}_0) \\ &= C_Z(\mathbf{0}) - \sum_{i=1}^{n} \lambda_i C_Z(\mathbf{x}_i - \mathbf{x}_0).\end{aligned}$$

L'écart-type de krigeage, σ_{KS}, fournit une mesure de l'erreur associée au prédicteur de krigeage. Notons que la variance de krigeage ne dépend pas de la valeur des données mais seulement de leurs emplacements et du modèle de covariance.

En notant $C(\mathbf{x}_0)$ le vecteur de covariance entre $\mathbf{x}_0$ et les points de données et $\mathbf{C}$ la matrice de covariance entre les données, la représentation matricielle du prédicteur de krigeage simple est :

$$Z^*(\mathbf{x}_0) = m + C(\mathbf{x}_0)'\mathbf{C}^{-1}\left(\mathbf{Z} - m\right). \tag{3.2}$$

Krigeage ordinaire

Le krigeage ordinaire permet de prédire une valeur en un point d'un domaine pour un champ, $Z(\cdot)$, stationnaire d'ordre 2 dont la covariance est connue et la moyenne est inconnue, en utilisant les données contenues dans le voisinage du point de prédiction et en le combinant linéairement avec des poids :

$$Z^*(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i).$$

La condition de non biais est $\sum_{j=1}^{n} \lambda_j = 1$, en effet :

$$\begin{aligned}
\mathbb{E}\left[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)\right] &= \mathbb{E}\left[\sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i) - Z(\mathbf{x}_0)\sum_{i=1}^{n} \lambda_i\right] \\
&= \sum_{i=1}^{n} \lambda_i \mathbb{E}\left[Z(\mathbf{x}_i) - Z(\mathbf{x}_0)\right] \\
&= 0.
\end{aligned}$$

La variance de l'erreur de prédiction est donnée par :

$$\sigma_E^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) + C_Z(\mathbf{x}_0 - \mathbf{x}_0) - 2\sum_{i=1}^{n} \lambda_i C_Z(\mathbf{x}_i - \mathbf{x}_0).$$

σ_E^2 doit être minimale sous la contrainte : $\sum_{i=1}^{n} \lambda_i = 1$. La méthode des pondérateurs de Lagrange fournit le système de krigeage issu de la minimisation de la variance de l'erreur :

$$\begin{cases} \sum_{j=1}^{n} \lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) - \mu = C_Z(\mathbf{x}_i - \mathbf{x}_0), \forall i = 1, \ldots, n \\ \sum_{j=1}^{n} \lambda_j = 1. \end{cases}$$

La variance du krigeage ordinaire est alors de la forme :

$$\sigma^2_{KO} = C_Z(\mathbf{x}_0 - \mathbf{x}_0) - \sum_{i=1}^{n} \lambda_i C_Z(\mathbf{x}_i - \mathbf{x}_0) + \mu.$$

Pour $\mathbf{1}$ le vecteur de dimension n d'éléments 1, en utilisant les notations de (3.2), le prédicteur du krigeage ordinaire a la forme matricielle suivante :

$$Z^*(\mathbf{x}_0) = C'(\mathbf{x}_0)\mathbf{C}^{-1}\mathbf{Z} + (1 - C'(\mathbf{x}_0)\mathbf{C}^{-1}\mathbf{1})\frac{\mathbf{1}'\mathbf{C}^{-1}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}}\mathbf{Z}. \tag{3.3}$$

Nous pouvons constater que, comparé au krigeage simple (équation (3.2)), le krigeage ordinaire présente un second terme correspondant à l'estimation de la moyenne, *i.e.* au krigeage de la moyenne :

$$m^* = \frac{\mathbf{1}'\mathbf{C}^{-1}\mathbf{Z}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}}. \tag{3.4}$$

Krigeage intrinsèque

Supposons maintenant que le champ aléatoire vérifie les conditions de la stationnarité intrinsèque et non plus celle de la stationnarité d'ordre 2. Cela signifie que l'incrément $Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})$ est d'espérance nulle et que $\mathrm{Var}\,(Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})) = 2\gamma(h)$. Ces hypothèses ne permettent pas de travailler directement sur $Z(\cdot)$ mais seulement sur des accroissements de $Z(\cdot)$, d'où la nécessité d'introduire la notion de combinaison linéaire autorisée. Ces combinaisons linéaires permettent de travailler dans le cadre du modèle intrinsèque : elles assurent l'existence de l'espérance et de la variance des accroissements. Pour $Z^*(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i)$, la condition de combinaison linéaire autorisée pour l'accroissement $Z(\mathbf{x}_0+\mathbf{h}) - Z(\mathbf{x}_0)$ est $\sum_{i=1}^n \lambda_i - 1 = 0$, où 1 correspond au poids de $Z(\mathbf{x}_0)$. Remarquons que la condition d'autorisation du krigeage intrinsèque est la condition de non biais du krigeage ordinaire. Par conséquent les équations du krigeage intrinsèque sont celles du krigeage ordinaire mais dans lesquelles intervient le variogramme et non la fonction de covariance :

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) - \mu = \gamma(\mathbf{x}_i - \mathbf{x}_0), \forall i = 1, \ldots, n \\ \sum_{j=1}^n \lambda_j = 1. \end{cases}$$

3.2 Propriétés du prédicteur de krigeage

Consistence avec les points de données

Les prédicteurs de krigeage est un interpolateur exact. Cela signifie que si $\mathbf{x}_0$ coïncide avec un des points de données, par exemple $\mathbf{x}_1$, alors $Z^*(\mathbf{x}_0) = Z(\mathbf{x}_1)$. L'ensemble de

poids $\{\lambda_1 = 1,\ \lambda_i = 0 \text{ si } i \neq 1\}$ satisfait l'équation (3.1) et représente l'unique solution. $Z^*(\mathbf{x}_0) = Z(\mathbf{x}_1)$ est le meilleur prédicteur de $Z(\mathbf{x}_1)$ au sens des moindres carrés car l'erreur de prédiction est nulle. La variance de krigeage est également nulle.
Comme le krigeage est un prédicteur linéaire, les prédictions sont moins dispersées que les données. En effet,

$$\begin{aligned}
\mathrm{Var}\left(Z(\mathbf{x}_0)\right) &= \mathrm{Var}\left(Z(\mathbf{x}_0) - Z^*(\mathbf{x}_0) + Z^*(\mathbf{x}_0)\right) \\
&= \mathrm{Var}\left(Z(\mathbf{x}_0) - Z^*(\mathbf{x}_0)\right) + \mathrm{Var}\left(Z^*(\mathbf{x}_0)\right) - 2\mathrm{Cov}\left(Z(\mathbf{x}_0) - Z^*(\mathbf{x}_0), Z^*(\mathbf{x}_0)\right) \\
&= \sigma^2_{KS} + \mathrm{Var}\left(Z^*(\mathbf{x}_0)\right).
\end{aligned}$$

D'où $\mathrm{Var}\left(Z^*(\mathbf{x}_0)\right) \leq \mathrm{Var}\left(Z(\mathbf{x}_0)\right)$. La variance de la prédiction $\mathrm{Var}\left(Z^*(\mathbf{x}_0)\right)$ diffère de $C_Z(\mathbf{0})$ par une quantité égale à la variance de krigeage (qui dépend de $\mathbf{x}_0$). Les prédictions tendent vers la moyenne entre les points de données, mais passent par les points de données.

Cas d'un champ gaussien

Si $Z(\cdot)$ est un champ aléatoire gaussien, alors le prédicteur de krigeage coïncide avec l'espérance conditionnelle $\mathbb{E}\left[Z(\mathbf{x}_0) \mid Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)\right]$. Cela vient directement de la linéarité de la fonction de régression pour une distribution gaussienne multivariée et d'une des propriétés de l'espérance conditionnelle (section 2.4).
En effet nous avons,

$$\mathbb{E}\left[\{\lambda_1 Z(\mathbf{x}_1) + \cdots + \lambda_n Z(\mathbf{x}_n) - Z(\mathbf{x}_0)\}\, Z(\mathbf{x}_i)\right] = 0,\ i = 1, \ldots, n$$

ce qui correspond aux équations de krigeage simple (3.1).

Prédiction des dérivées d'un champ aléatoire

D'après les relations (3.2) et (3.3) le champ aléatoire $Z^*(\mathbf{x}), \mathbf{x} \in \mathcal{D}$ hérite des propriétés de différentiabilité de la fonction $C(\mathbf{x})$, *i.e.* de $C_Z(\mathbf{x} - \mathbf{x}_i), \forall i = 1, \ldots, n$. Si la fonction de covariance est supposée indéfiniment différentiable, $Z^*(\cdot)$ sera indéfiniment différentiable pour tout $\mathbf{x}$, sauf peut-être en les points d'échantillonnage, *i.e.* en les points où $\mathbf{x} - \mathbf{x}_i = \mathbf{0}$. Cela implique que $Z^*(\cdot)$ est presque sûrement indéfiniment différentiable. Pour $\mathbf{Z}$ un vecteur gaussien de variance finie, il en découle l'indéfinie différentiabilité en moyenne quadratique. En les points d'échantillonnage, le champ $Z^*(\cdot)$ sera continu si $C_Z(\mathbf{h})$ est continue à l'origine. De façon plus générale, $Z^*(\cdot)$ sera k-fois différentiable si $C_Z(\mathbf{h})$ est

$2k$-fois différentiable à l'origine. Si $C_Z(\mathbf{h})$ possède un effet de pépite, *i.e.* si $C_Z(\mathbf{h})$ est discontinue en $\mathbf{h} = \mathbf{0}$, alors $Z^*(\cdot)$ sera discontinu aux points d'échantillonnage. L'ensemble des points d'échantillonnage étant un ensemble de mesure nulle dans $\mathcal{D}$, cela ne représente pas une limite sérieuse.

La condition mathématique de différentiabilité d'un champ aléatoire $Z(\cdot)$ est la double différentiabilité à l'origine de la fonction de covariance du champ (théorème 2). Les modèles de covariance les plus utilisés en pratique (modèles sphérique et exponentiel) ne sont pas différentiables à l'origine (section 3.3). Les dérivées du champ ne sont par conséquent pas définies mathématiquement (variance infinie) et ne peuvent en principe pas être prédites ou utilisées pour la prédiction. Nous allons pourtant supposer que ces dérivées peuvent être approchées. Supposons que le champ $Z(\cdot)$ soit définit comme sur la figure 3.1 (trait noir). Ce champ est non différentiable, en revanche sa prédiction (trait orange) est par construction plus lisse et différentiable, sauf éventuellement en les points de données (cercles rouges). Nous allons supposer que la dérivée de la prédiction en certains points (carrés verts) fournit une bonne approximation de la prédiction de la dérivée du champ.

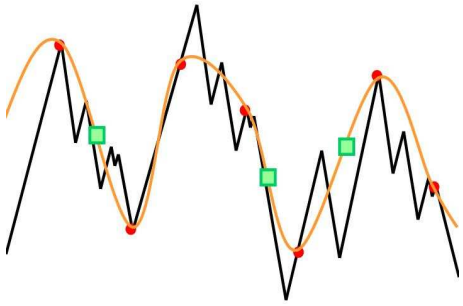


FIG. 3.1: *Représentation d'un champ aléatoire (en noir) et de sa prédiction (en orange) passant par les points d'échantillonnage (cercles rouges). Les carrés verts correspondent aux points en lesquels nous voulons prédire la dérivée du champ.*

Soit $\delta_i,\ 1 \leq i \leq d$ le vecteur de $\mathbb{R}^d$ nul en toutes ses composantes sauf en la $i^{ème}$ qui est égale à 1. Par linéarité du système de krigeage, le prédicteur optimal de la différence $Z(\mathbf{x} + \epsilon\delta_i) - Z(\mathbf{x})$ est $Z^*(\mathbf{x} + \epsilon\delta_i) - Z^*(\mathbf{x})$. En passant à la limite comme en (2.6), nous avons d'après Chilès et Delfiner (1999) :

$$\left(\frac{\partial Z}{\partial x^i}\right)^* = \frac{\partial Z^*}{\partial x^i}. \tag{3.5}$$

C'est-à-dire que le prédicteur de krigeage du champ $\partial Z(\cdot)/\partial x^i$ est égal à la dérivée du prédicteur de krigeage. L'équation (3.5) présume que les dérivées de $Z(\cdot)$ et de $Z^*(\cdot)$

existent, en réalité seul $Z^*(\cdot)$ doit être différentiable. Par exemple, si le variogramme de $Z(\cdot)$ est linéaire près de l'origine, $Z^*(\cdot)$ est différentiable sauf peut-être en les points d'échantillonnage. Par extension, $\partial Z^*/\partial \mathbf{x}^i$ est accepté comme prédicteur de la dérivée. Notons que dériver le prédicteur de krigeage n'entraîne pas une augmentation du bruit dans les dérivées comme ce serait le cas en dérivant Z, puisque Z^* est une version lissée de Z, sauf en les points d'échantillonnage.

3.3 Modèles de covariance

Nous verrons dans la section 5.1 que la méthode de détection de Zones de Changement Abrupt que nous proposons nécessite des hypothèses sur la régularité de la fonction de covariance du champ aléatoire considéré. Nous présentons dans ce paragraphe les modèles auxquels nous nous référons dans ce travail.

Modèle gaussien

$$C_Z(\mathbf{h}) = c_0 \exp\left(-\frac{\|\mathbf{h}\|^2}{b}\right), \text{ avec } b, c_0 > 0. \tag{3.6}$$

Le modèle gaussien est associé à un champ aléatoire stationnaire dont la fonction de covariance est infiniment différentiable à l'origine. Dans ce cas le champ aléatoire $Z(\cdot)$ est aussi infiniment différentiable. Ce modèle est donc extrêmement régulier.

Modèle exponentiel

$$C_Z(\mathbf{h}) = c_0 \exp\left(-\frac{\|\mathbf{h}\|}{b}\right), \text{ avec } b, c_0 > 0. \tag{3.7}$$

La fonction $C_Z(\mathbf{h})$ diminue exponentiellement lorsque la distance augmente. Le paramètre b détermine la rapidité avec laquelle la fonction diminue. Pour $\|\mathbf{h}\| = 3b$ la fonction a diminué de 95% de sa valeur d'origine. Cette distance est appelée *portée pratique* du modèle de covariance.
Le modèle exponentiel est continu mais pas différentiable à l'origine, ce qui entraîne que le champ n'est pas non plus différentiable.

Modèle à effet de pépite pur

$$C_Z(\mathbf{h}) = \begin{cases} c_0, & \text{si } \|\mathbf{h}\| = 0 \\ 0, & \text{si } \|\mathbf{h}\| > 0 \end{cases} \tag{3.8}$$

Cette fonction de covariance modélise une discontinuité à l'origine, où c_0 est une valeur positive. Ce modèle est également appelé *bruit blanc*.

Modèle sphérique

$$C_Z(\mathbf{h}) = \begin{cases} c_0 \left(1 - \frac{3}{2}\frac{\|\mathbf{h}\|}{b} + \frac{1}{2}\frac{\|\mathbf{h}\|^3}{b^3}\right), & \text{si } 0 \leq \|\mathbf{h}\| \leq b \\ 0, & \text{si } \|\mathbf{h}\| > b \end{cases} \tag{3.9}$$

Le paramètre b indique la portée de la covariance sphérique : les corrélations deviennent nulles lorsque la portée est atteinte. Le paramètre c_0 représente la valeur maximale de la covariance. Le modèle à effet de pépite pur est un cas particulier du modèle sphérique et correspond au cas où la portée est infiniment petite. Cependant il y a une différence importante entre les deux modèles : le modèle pépitique décrit un phénomène discontinu dont les valeurs changent brusquement d'un site à l'autre, tandis que le modèle sphérique représente un phénomène continu (mais non différentiable).

La figure 3.2 illustre les modèles de covariance gaussien, exponentiel et sphérique, pour une portée $b = 0.25$ et une variance $c_0 = 1$. Cette figure montre les différents comportements à l'origine des modèles de covariance usuels en géostatistique.

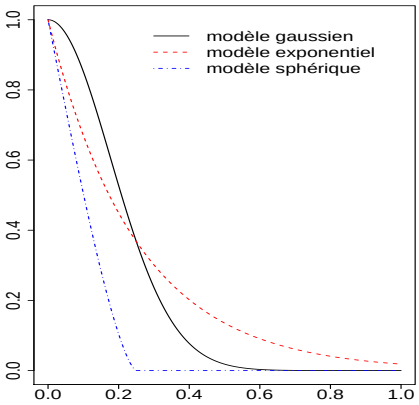


FIG. 3.2: *Modèles de covariance gaussien, exponentiel et sphérique, pour une portée fixée à* $b = 0.25$.

Les propriétés des interpolateurs spatiaux dépendent du comportement local du champ aléatoire. En pratique ce comportement local n'est pas connu et doit être estimé à partir des données utilisées pour l'interpolation. Il est difficile de sélectionner un modèle de covariance qui reflète parfaitement les comportements locaux de la variable d'étude. Bon

nombre de modèles usuels de covariances, comme les modèles gaussiens, exponentiels ou sphériques, fournissent peu de flexibilité aux comportements locaux et cela suppose généralement qu'ils sont connus *a priori*. Un modèle plus général est le modèle de Matérn [voir par exemple Stein (1999)] :

$$C_Z(\mathbf{h}) = \frac{\pi^{1/2}\phi}{2^{\nu-1}\Gamma(\nu+1/2)\alpha^{2\nu}} (\alpha\|\mathbf{h}\|)^\nu \mathcal{K}_\nu(\alpha\|\mathbf{h}\|), \text{ avec } \nu > 0, \phi > 0 \text{ et } \alpha > 0. \quad (3.10)$$

$\mathcal{K}_\nu$ est la fonction de Bessel modifiée de deuxième espèce :

$$\mathcal{K}_\nu(x) = \frac{\pi}{2}\frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\pi\nu)}, \text{ avec } I_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(\nu+k+1)} \left(\frac{x}{2}\right)^{2k}.$$

ν est un paramètre critique : plus ν est élevé, plus le champ $Z(\cdot)$ est lisse. En particulier, $Z(\cdot)$ sera k-fois différentiable au sens des moindres carrés si et seulement si $\nu > k$. Le paramètre ϕ correspond au rapport entre la variance c_0 et la portée α^{-1} à la puissance $2\nu^{\text{ème}}$: $\phi = c_0\alpha^{2\nu}$. Ce modèle peut avoir tout type de comportement à l'origine puisqu'il se comporte comme $\|\mathbf{h}\|^{2\nu}$ si ν n'est pas un entier et comme $\|\mathbf{h}\|^{2\nu}\log(\|\mathbf{h}\|)$ si ν est un entier. Il dépend d'un paramètre permettant tout degré de différentiabilité du champ aléatoire. La figure 3.3 illustre différents comportements que peut prendre le modèle de Matérn. Sur la figure 3.3*a* la portée est fixe : $\alpha^{-1} = 0.25$ et le paramètre de lissage ν varie. Cette figure montre que plus ν est élevé, plus la covariance est régulière à l'origine. Sur la figure 3.3*b* la portée est variable, de sorte que les différents modèles représentés aient la même portée pratique et le paramètre de lissage ν varie.

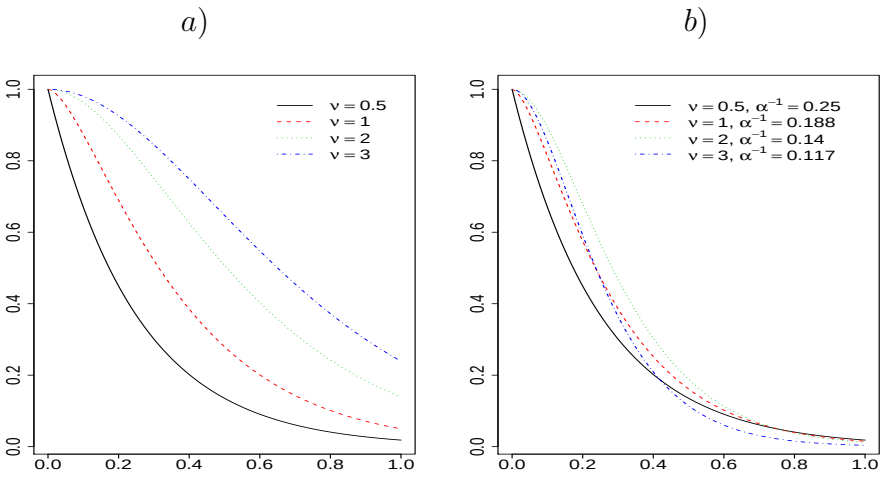


FIG. 3.3: *Exemples de modèles de Matérn en fonction des paramètres de portée, α^{-1}, et de lissage, ν.*

Le modèle de Matérn inclut les modèles gaussien et exponentiel : lorsque $\nu \to \infty$ le modèle de Matérn correspond au modèle gaussien et lorsque $\nu = 1/2$, au modèle exponentiel.

Nous verrons (section 5.1) que le modèle de Matérn répond aux conditions de régularité imposées par la méthode de détection de ZCAs.

Chapitre 4

Ensembles d'excursion de champs aléatoires

Dans la méthode de détection de Zones de Changement Abrupt, nous cherchons à déterminer le comportement d'un champ de χ^2 au dessus d'un seuil. Nous appelons ensemble d'excursion d'un champ aléatoire l'ensemble des points où le champ est supérieur au seuil. Pour $Z(\cdot)$ un champ aléatoire et $\mathcal{D}$ un sous-ensemble compact de $\mathbb{R}^d$, l'ensemble d'excursion du champ $Z(\cdot)$ au dessus du seuil z dans $\mathcal{D}$ est défini par :

$$\mathcal{A}_z = \mathcal{A}_z(Z, \mathcal{D}) = \{\mathbf{x} \in \mathcal{D} \ : \ Z(\mathbf{x}) \geq z\}\,. \tag{4.1}$$

La connexité d'ensembles aléatoires peut être décrite par différentes caractéristiques. Les plus usuelles sont des grandeurs scalaires : la caractéristique d'Euler et le nombre de composantes connexes. Dans ce chapitre nous définissons ces notions et les relions à l'étude des ensembles d'excursion de champs aléatoires.

4.1 Nombre de composantes connexes

La connaissance de la distribution du maximum d'un champ aléatoire permet, à l'aide d'un test statistique, de détecter des changements locaux dans des données. C'est en particulier le cas en analyse d'images médicales obtenues par émission de positrons ou par résonnance magnétique fonctionnelle (Worsley et Friston dans les années 1990). Un test statistique, basé sur les réactions à un stimulus, permet d'identifier les zones du cerveau où le signal s'accroît. L'image obtenue représente la statistique de test, modélisée par un champ aléatoire, $Z(\cdot)$. La distribution du maximum de ce champ aléatoire permet alors

de déterminer la valeur z du seuil du test, au niveau α fixé. Lorsque la statistique de test est supérieure à z, les régions du cerveau correspondantes sont déclarées significativement positives au stimulus. Les premiers travaux mis en œuvre pour donner une approximation de la distribution du maximum sont ceux de Rice dans les années 1940 [voir Delmas (2001) pour une synthèse de ces travaux]. Ils consistent à utiliser le nombre de *upcrossings* du niveau z du champ aléatoire $Z(\cdot)$ (supposé assez régulier) sur un intervalle $[0, X] \subset \mathbb{R}$:

$$U_z^Z\left([0,x']\right) = \sharp\left\{x \in [0,x'] : Z(x) = z, Z'(x) > 0\right\},$$

de la façon suivante :

$$\mathbb{P}\left[\max_{x\in[0,x']} Z(x) \geq z\right] \leq \mathbb{P}\left[Z(0) \geq z\right] + \mathbb{E}\left[U_z^Z\left([0,x']\right)\right].$$

Une approche plus récente [Adler (1981) et Worsley dans les années 1990] consiste à généraliser la notion de upcrossing en dimension un, par la caractéristique d'Euler de l'ensemble d'excursion $\mathcal{A}_z(Z,\mathcal{D}) = \{\mathbf{x} \in \mathcal{D} \ : \ Z(\mathbf{x}) \geq z\}$ en dimension supérieure. Afin de déterminer la distribution du maximum de $Z(\cdot)$ par :

$$\mathbb{P}\left[\max_{\mathbf{x}\in\mathcal{D}} Z(\mathbf{x}) \geq z\right] = \mathbb{P}\left[\text{Nombre de composantes connexes de } \mathcal{A}_z(Z,\mathcal{D}) \geq 1\right],$$

il est nécessaire d'avoir un compteur du nombre de composante connexes de $\mathcal{A}_z(Z,\mathcal{D})$. Or le nombre de composantes connexes est difficile à étudier pour des dimensions strictement supérieures à un, ce qui n'est pas le cas de la caractéristique d'Euler (proche du nombre de composantes connexes) pour laquelle il existe de nombreux outils [Adler (1981,2000)]. Avant de voir comment la caractéristique d'Euler est utilisée pour approcher la distribution du maximum d'un champ aléatoire, nous faisons un rappel sur la connexité.

Deux points d'un ensemble sont connectés s'il est possible d'aller de l'un à l'autre par un chemin continu dans cet ensemble. C'est la notion de connexité par arc.

Définition 4 *Une partie A de $\mathbb{R}^d$ est dite connexe par arc si pour tout $\mathbf{x}, \mathbf{y} \in A$, il existe une application continue f d'un intervalle $[u, v]$ de $\mathbb{R}$ dans A telle que $f(u) = \mathbf{x}$ et $f(v) = \mathbf{y}$.*

Définition 5 *La réunion de toutes les parties connexes par arc contenant $\mathbf{x} \in B$ est une partie connexe appelée composante connexe de $\mathbf{x}$ et notée $\mathcal{C}_{\mathbf{x}}(B)$.*

Nous pouvons également donner une définition axiomatique des composantes connexes.

Définition 6 *Soit B une partie quelconque de $\mathbb{R}^d$. Une composante connexe en $\mathbf{x}$ de B, notée $\mathcal{C}_{\mathbf{x}}(B)$, est une partie de $\mathbb{R}^d$ vérifiant les axiomes suivants :*

(i) Pour tout $\mathbf{x}$ de $\mathbb{R}^d$, $\mathbf{x} \notin B \Rightarrow \mathcal{C}_{\mathbf{x}}(B) = \emptyset$ (figure 4.1a).

(ii) Pour tout $\mathbf{x}$ de $\mathbb{R}^d$, si $B = \{\mathbf{x}\}$, $\mathcal{C}_{\mathbf{x}}(B) = \{\mathbf{x}\}$ (figure 4.1b).

(iii) Pour tout $\mathbf{x}$, $\mathbf{y}$ de $\mathbb{R}^d$, $\mathcal{C}_{\mathbf{x}}(B)$ et $\mathcal{C}_{\mathbf{y}}(B)$ sont égales (figure 4.1c) ou disjointes (figure 4.1d).

(iv) $\cup_{\mathbf{x}_i \in \mathbb{R}^d} \mathcal{C}_{\mathbf{x}_i}(B) = B$ (figure 4.1e).

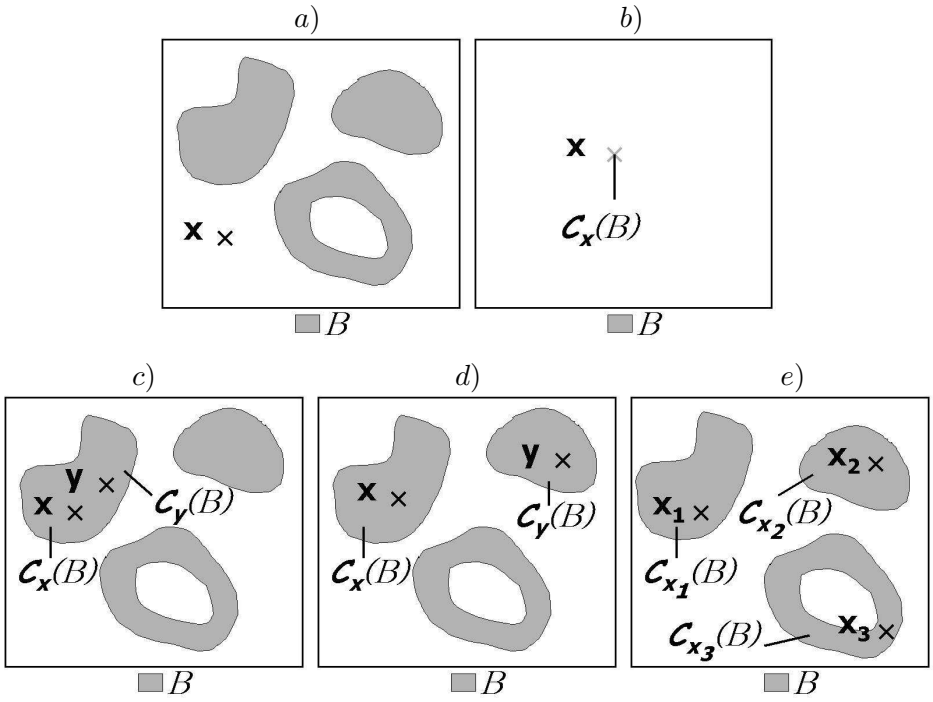


FIG. 4.1: *Illustration des propriétés de la définition axiomatique d'une composante connexe (définition 6).*

Dans le cas $d = 1$, la géométrie des composantes connexes est simple : il s'agit simplement de segments de droite. Dans ce cas la caractéristique d'Euler de l'ensemble d'excursion est égale au nombre de composantes connexes. Pour des dimensions plus grandes la topologie de ces composantes connexes est plus complexe. En dimension deux, la caractéristique d'Euler est égale au nombre de composantes connexes, diminué du nombre total de trous situés dans celles-ci. En dimension trois, il s'agit du nombre de composantes connexes diminué du nombre de circuits simples de chacune de celles-ci. Un circuit simple d'un ensemble connexe est une boucle fermée qu'il est possible de tracer sur sa surface sans le séparer en deux parties connexes.

La caractéristique d'Euler apparaît donc comme une alternative possible au nombre de composantes connexes. Comme un champ aléatoire suffisamment régulier présente un comportement paraboloïde au voisinage de ses hauts maxima locaux, nous pouvons penser que, pour z suffisamment grand, la caractéristique d'Euler des ensembles d'excursion de $Z(\cdot)$ au-dessus de z vaut un (les ensembles d'excursion étant alors des ellipsoïdes). La caractéristique d'Euler de l'ensemble d'excursion $\mathcal{A}_z(Z, \mathcal{D})$, notée $\chi(\mathcal{A}_z)$, correspond alors à son nombre de composantes connexes (à des termes de bord près). Par conséquent lorsque $z \to \infty$ et que le champ aléatoire est suffisamment régulier, nous pouvons approcher le nombre de composantes connexes par la caractéristique d'Euler et ainsi obtenir une approximation de la distribution du maximum de $Z(\cdot)$:

$$\begin{aligned} \mathbb{P}\left[\max_{\mathbf{x}\in\mathcal{D}} Z(\mathbf{x}) \geq z\right] &= \mathbb{P}\left[\text{Nombre de composantes connexes de } \mathcal{A}_z(Z,\mathcal{D}) \geq 1\right] \\ &\leq \mathbb{E}\left[\text{Nombre de composantes connexes de } \mathcal{A}_z(Z,\mathcal{D})\right] \\ &\approx \mathbb{E}\left[\chi(\mathcal{A}_z)\right]. \end{aligned} \tag{4.2}$$

Nous venons de voir que pour des ensembles d'excursion dont le seuil est élevé et pour des champs aléatoires suffisamment réguliers, il existe une relation étroite entre la caractéristique d'Euler et le nombre de maxima locaux. En effet, lorsque le seuil tend à être élevé, l'ensemble d'excursion du champ aléatoire tend à avoir une topologie simple. Par exemple en dimension deux les trous tendent à disparaître et il ne reste que des régions isolées ayant chacune un maximum local. Dans ce cas la caractéristique d'Euler compte juste le nombre de composantes connexes, ou compte le nombre de maxima locaux. Comme nous le verrons dans ce chapitre, la caractéristique d'Euler peut approcher le maximum global pour des seuils élevés. Effectivement, si le seuil est très élevé "avoir plus d'une composante connexe" devient un évènement rare. Par conséquent, si l'ensemble d'excursion n'est pas vide, il peut être constitué d'une seule composante connexe et donc avoir un maximum local qui sera alors le maximum global. La figure 4.2 illustre ce phénomène. Nous avons déterminé des ensembles d'excursion à partir du champ représenté sur l'image artificielle 4.2*a*. Les figures 4.2*b*, 4.2*c* et 4.2*d* représentent les ensembles d'excursion $\mathcal{A}_z$ pour différents niveaux : $z = 6$, 10.5 et 13.4 respectivement. Il s'agit de l'ensemble des pixels colorés en noir. Plus le seuil z est élevé, plus l'ensemble d'excursion devient petit et plus les trous tendent à disparaître (figure 4.2*c*). Lorsque z est très élevé (figure 4.2*d*), l'ensemble d'excursion ne contient plus qu'une composante connexe.

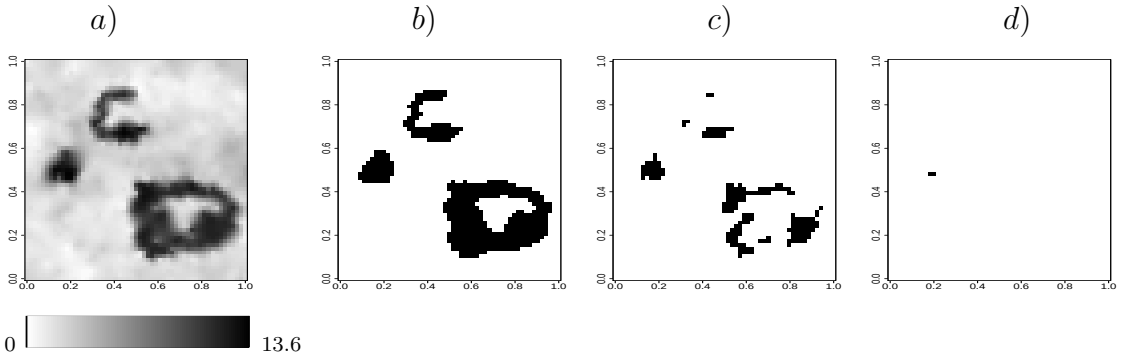


FIG. 4.2: *a) Image artificielle dans* $\mathbb{R}^2$*. Ensemble d'excursion* $\mathcal{A}_z$ *au dessus de : b)* $z = 6$*, c)* $z = 10.5$ *et d)* $z = 13.4$*.*

4.2 Taille d'une composante connexe

La caractéristique d'Euler compte le nombre de composantes connexes pour des seuils élevés. Une autre mesure de ces composantes connexes est leur taille. Lorsque les seuils sont de plus en plus élevés, les composantes connexes deviennent plus régulières et deviennent des ensembles convexes avec un maximum local. L'étude de la forme de ces ensembles convexes mène à avoir une approximation de leur taille. La forme des champs aléatoires gaussiens stationnaires est étudiée dans Adler (1981). Il a montré que les champs gaussiens ont une forme quadratique lorsque les seuils sont élevés, ce qui implique que les composantes connexes de l'ensemble d'excursion sont des ellipsoïdes. Dans le théorème suivant, ainsi que ceux de la section 4.3, la fonction de covariance du champ aléatoire $Z(\cdot)$ est supposée 4 fois différentiable à l'origine. Notons que le résultat suivant ne nécessite pas que z soit un maximum local, mais seulement qu'il soit élevé.

Théorème 4 (Cao, 1997) *Soit* $Z(\cdot)$ *un champ gaussien, stationnaire, de moyenne nulle et satisfaisant la condition :*

$$\max_{i,j} \mathbb{E}\left[\mid \partial^2_{ij} Z(\mathbf{x}) - \partial^2_{ij} Z(\mathbf{0}) \mid^2 \right] \leq c\|\mathbf{x}\|^2, \tag{4.3}$$

pour $c > 0$ *et pour tout* $\mathbf{x}$ *suffisament petit, avec* $\partial^2_{ij} Z$ *la composante* $[i, j]$ *de la matrice Hessienne de* Z*. Sachant que le champ* $Z(\cdot)$ *prend la valeur* z *quand* $\mathbf{x} = \mathbf{0}$*, et en supposant que les dérivées premières du champ en* $\mathbf{0}$ *soient nulles, le champ a alors, avec une probabilité tendant vers 1 quand* $z \to \infty$*, la représentation suivante sur la composante connexe contenant* $\mathbf{0}$ *:*

$$Z(\mathbf{x}) = z - \frac{1}{2} z \mathbf{x}' \mathbf{\Lambda} \mathbf{x} + o\left(\frac{1}{z}\right),$$

où $\mathbf{\Lambda}$ *est la matrice de covariance de* $\partial Z / \partial \mathbf{x}$ *en* $\mathbf{0}$*.*

4.3 Ensembles d'excursion de champs de χ^2

La caractéristique d'Euler de l'ensemble d'excursion d'un champ aléatoire a servi d'outil pour analyser les images obtenues par tomographie par émission de positrons et par résonnance magnétique fonctionnelle. Ces images sont modélisées comme des champs gaussiens. La caractéristique d'Euler est à la base de l'estimateur proposé pour le nombre de signaux de l'image (section 4.1). Dans ce cadre d'analyse statistique des changements fonctionnels et structuraux du cerveau humain, une statistique indiquant l'activité neuronale en chaque partie du cerveau donne une image en trois dimensions. Le maximum global de ces images a été proposé par Friston *et al.* (1991) et Worsley *et al.* (1992) pour tester les régions sujettes à des changements importants supposés locaux. Ces images ont été modélisées comme la réalisation d'un champ aléatoire stationnaire sous l'hypothèse nulle. La valeur critique associée au maximum global est approchée par l'espérance de la caractéristique d'Euler de l'ensemble d'excursion caractérisant les changements locaux. Cette espérance a été calculée pour un champ gaussien [Adler (1981)] et étendue aux champs χ^2 [Worlsey (1994)]. Certains auteurs ont préféré utiliser la taille de la plus grande composante connexe de l'ensemble d'excursion comme statistique de test et ont préalablement déterminé la distribution asymptotique de la taille d'une composante connexe de l'ensemble d'excursion. Appliquant ces travaux aux champs gaussiens, Friston *et al.* (1994) ont présenté une analyse approximative pour évaluer l'importance des activités par leur répartition spatiale. Leurs travaux ont été étendus aux champs de χ^2 par Cao (1999) dont nous allons énoncer les principaux résultats.

Nous nous intéressons donc aux ensembles d'excursion de champs de χ^2 comme définis dans la section 2.6. Nous noterons $Y(\cdot)$ le champ de χ^2 associé aux p champs gaussiens indépendants, identiquement distribués et stationnaires : $Z_1(\mathbf{x}), \ldots, Z_p(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, où chaque $Z_i(\cdot)$ est de moyenne nulle et avec $\mathbf{\Lambda} = \mathrm{Var}\,(\partial Z_i(\mathbf{x})/\partial \mathbf{x})$ la $d \times d$ matrice de covariance de $\partial Z_i(\mathbf{x})/\partial \mathbf{x}$. Plus particulièrement, nous allons étudier la courbure et la hauteur de tels champs près des maxima locaux, ainsi que la taille des composantes connexes d'ensembles d'excursion. Afin d'alléger les notations, nous noterons $\dot{Z}$ et $\ddot{\mathbf{Z}}$ les dérivées premières et secondes d'un champ $Z(\cdot)$.

Nous allons déterminer la distribution conditionnelle de la courbure $\ddot{\mathbf{Y}}$ en $\mathbf{0}$ sachant que $\mathbf{0}$ est un maximum de hauteur y. Ceci revient à conditionner par $Y(\mathbf{0}) = y$, $\dot{Y}(\mathbf{0}) = \mathbf{0}$, et $\ddot{\mathbf{Y}}(\mathbf{0}) < \mathbf{0}$. Des problèmes similaires ont été étudiés pour des champs gaussiens dans Adler (1981) et pour des champs de χ^2 dans Aronowich et Adler (1986, 1988) en dimension une et deux. L'outil approprié pour une telle étude est le conditionnement par fenêtre

horizontale (section 2.4).

Lemme 3 (Worsley, 1994) *Les dérivées premières et secondes d'un champ $Y = Y(\cdot) \sim \chi^2(p)$ peuvent s'écrire de la façon suivante, où $\stackrel{\mathcal{L}}{=}$ désigne l'égalité en loi :*
(i) $\dot{Y} \stackrel{\mathcal{L}}{=} 2Y^{1/2}\mathbf{z}$,
(ii) $\ddot{\mathbf{Y}} \stackrel{\mathcal{L}}{=} 2\left(\mathbf{P} + \mathbf{z}\mathbf{z}' - Y\mathbf{\Lambda} + Y^{1/2}\mathbf{H}\right)$,
où $\mathbf{z} \sim \mathcal{N}_d(0, \mathbf{\Lambda})$, $\mathbf{P} \sim Wishart_d(\mathbf{\Lambda}, p-1)$ *et* $\mathbf{H} \sim \mathcal{N}_{d\times d}(\mathbf{0}, \mathbf{M}(\mathbf{\Lambda}))$ *sont des vecteurs aléatoires tous indépendants.*

Ce lemme a permis de montrer le résultat suivant.

Théorème 5 (Cao, 1999) *Pour un champ de χ^2, $Y(\cdot)$, ayant un maximum local en $\mathbf{0}$ de hauteur y,*

$$y^{-1}\ddot{\mathbf{Y}}(\mathbf{0}) \xrightarrow{\mathcal{L}} -2\mathbf{\Lambda}, \text{ quand } y \to \infty. \tag{4.4}$$

Les résultats de cette section supposent que les champs gaussiens satisfont l'équation (4.3). Cette condition assure les conditions de régularité des champs de χ^2 requises pour l'utilisation du conditionnement par fenêtre horizontale.
Le résultat suivant permet d'obtenir la distribution conditionnelle de la hauteur du champ $Y(\cdot)$ en un maximum local.

Théorème 6 (Cao, 1999) *Pour un champ de χ^2, $Y(\cdot)$, ayant un maximum local en $\mathbf{0}$ de hauteur Y_0 supérieure à y,*

$$\lim_{y\to\infty} \mathbb{P}\left[Y_0 > y + y' \,\|\, Y_0 > y;\ \textit{maximum}\ \right] = \exp\left(-y'/2\right), \text{ pour } y' > 0.$$

Le lemme 2 permet de montrer le théorème suivant sur la structure du champ $Y(\cdot)$ près d'un maximum local.

Théorème 7 (Cao, 1999) *Si $\mathbf{0}$ est un maximum local d'un champ de χ^2, $Y(\cdot)$, de hauteur y et si $Y_y(\cdot)$ est le champ conditionnel défini par (2.7), alors $Y_y(\mathbf{x}/\sqrt{y}) - y$ converge uniformément vers le paraboloïde elliptique $-\mathbf{x}'\mathbf{\Lambda}\mathbf{x}$ quand $y \to \infty$.*

Il est intéressant de voir que ce champ non gaussien a le même comportement en un maximum local qu'un champ gaussien. Pour un seuil élevé un champ de χ^2 montre une forme quadratique, comme un champ gaussien.
Ces résultats rendent possible la détermination de la distribution asymptotique de la taille d'une composante connexe d'un ensemble d'excursion pour un seuil élevé. Supposons que

l'ensemble d'excursion au-dessus du seuil y contienne $\mathbf{x} = \mathbf{0}$ comme maximum local et définissons la composante connexe $\mathcal{C}_{\mathbf{0}}(\mathcal{A}_y)$ contenant $\mathbf{0}$ par :

$$\mathcal{C}_{\mathbf{0}}(\mathcal{A}_y) = \{\mathbf{x} : Y(\mathbf{x}) \geq y,\ \mathbf{x} \text{ est connecté avec } \mathbf{0}\}.$$

De plus, notons S_y la taille de $\mathcal{C}_{\mathbf{0}}(\mathcal{A}_y)$: $S_y = \lambda(\mathcal{C}_{\mathbf{0}}(\mathcal{A}_y))$, où λ désigne la mesure de Lebesgue dans $\mathbb{R}^d$. Alors nous avons le théorème suivant sur la distribution asymptotique de S_y.

Théorème 8 (Cao, 1999) *Pour $Y(\cdot)$ un champ de χ^2, quand $y \to \infty$,*

$$y^{d/2} S_y \longrightarrow a_d \det(\mathbf{\Lambda})^{-1/2} E^{d/2},$$

où E est une variable aléatoire exponentielle d'espérance 2 et $a_d = \pi^{d/2}/\Gamma(d/2+1)$ est la mesure de Lebesgue de la sphère unité dans $\mathbb{R}^d$.

Ce théorème permet donc d'obtenir la distribution asymptotique de la taille S_y d'une composante connexe : $y^{d/2} S_y \det(\mathbf{\Lambda})^{1/2}/a_d \longrightarrow E^{d/2}$ quand $y \to \infty$.

Considérant que les changements fonctionnels et structuraux du cerveau humain sont locaux, Friston *et al.* (1991) et Worsley *et al.* (1992) ont utilisé le maximum global, Y_{max}, pour tester les régions de changements en des emplacements inconnus. Cependant le test basé sur le maximum global ne prend pas en compte l'étendue spatiale de la région détectée. Une alternative est la taille de la plus grande composante connexe, S_{max}, de l'ensemble d'excursion [Friston *et al.* (1994)].
Cao (1999) a pu donner une approximation de la distribution asymptotique de S_{max} dans le cas d'un champ χ^2, en utilisant une heuristique poissonnienne [Aldous (1989)]. Pour un seuil y élevé, les centres des composantes connexes $\mathcal{C}_1, \ldots, \mathcal{C}_L$ de l'ensemble d'excursion $\mathcal{A}_y = \{\mathbf{x} : Y(\mathbf{x}) \geq y\}$ peuvent être vus comme un processus ponctuel multidimensionnel avec absence de mémoire, *i.e.* lorsque $y \to \infty$, le nombre L de composantes connexes suit une distribution de Poisson et les tailles $S_i,\ i = 1, \ldots, L$ sont indépendantes et identiquement distribuées. Ainsi,

$$\mathbb{P}[L = l] \approx \frac{\mathbb{E}[L]^l}{l!} \exp(-\mathbb{E}[L]),$$

et

$$\begin{aligned}\mathbb{P}\left[S_{max} \leq s \mid L \geq 1\right] &= \sum_{l=1}^{\infty} \mathbb{P}\left[S_i \leq s, 1 \leq i \leq l, L = l \mid L \geq 1\right] \\ &\approx \frac{1}{\mathbb{P}\left[L \geq 1\right]} \sum_{l=1}^{\infty} \frac{\mathbb{E}\left[L\right]^l}{l!} \exp\left(-\mathbb{E}\left[L\right]\right) \mathbb{P}\left[S \leq s\right]^l \\ &= \frac{\exp\left(-\mathbb{E}\left[L\right] \mathbb{P}\left[S \geq s\right]\right) - \exp\left(-\mathbb{E}\left[L\right]\right)}{1 - \exp\left(-\mathbb{E}\left[L\right]\right)}. \end{aligned} \tag{4.5}$$

L'équation (4.5) relie la distribution de S_{max} à celle de S, dont la limite a été évaluée dans le théorème 8. Dans le cas extrême où $\mathbb{E}\left[L\right]$ tend vers 0, $\mathbb{P}\left[S_{max} \leq x\right] \rightarrow \mathbb{P}\left[S \leq x\right]$ et alors S_{max} a la même distribution que S. Les espérances $\mathbb{E}\left[L\right]$ et $\mathbb{E}\left[\chi\left(\mathcal{A}_y\right)\right]$, où $\chi\left(\mathcal{A}_y\right)$ est la caractéristique d'Euler de l'ensemble d'excursion $\mathcal{A}_y$, sont asymptotiquement égales. $\mathbb{E}\left[\chi\left(\mathcal{A}_y\right)\right]$ a pu être évaluée dans Adler (1981) pour un champ gaussien et dans Worsley (1994) pour un champ de χ^2.

Conclusion

Dans la détection de zones de changement structurel et fonctionnel du cerveau humain, le calcul d'une statistique de test permet d'obtenir une image de l'activité cérébrale. Afin d'identifier les zones de changement significatives, la taille de la plus grande composante connexe de l'ensemble d'excursion est proposée comme statistique de test. Cette étude statistique des changements structuraux du cerveau est un exemple de problème de comparaisons multiples : comme dans la méthode que nous proposons, les tests locaux doivent être agrégés afin de tester la significativité des changements détectés. Nous discuterons des problèmes d'agrégation de tests statistiques dans la section 7.2 lorsque nous proposerons une méthode pour évaluer le niveau local de significativité à niveau global fixé.

Deuxième partie

Détection de Zones de Changement Abrupt

Introduction

Nous proposons une méthode permettant de détecter les zones où une variable spatialisée et faiblement échantillonnée dans le plan varie brusquement. Notre modèle général est que sous l'hypothèse nulle le champ aléatoire, $Z(\cdot)$, modélisant la variable observée est d'espérance constante sur un domaine $\mathcal{D}$. L'alternative est une discontinuité de $\mathbb{E}[Z(\cdot)]$ le long d'un ensemble de courbes Γ. Aucune hypothèse n'est faite sur la forme des discontinuités. Comme nous considérons le cas d'un échantillonnage irrégulier et de faible densité de points, nous ne pouvons pas parfaitement estimer ces courbes mais seulement les zones où de fortes variations ont lieu. Nous parlons alors de l'estimation des Zones de Changement Abrupt (ZCAs).

Nous avons choisi une approche différente de celles présentées dans le chapitre 1, basée sur les propriétés de l'estimation du gradient lorsque la norme du gradient est élevée par rapport à l'estimation de sa variance. Pour un champ aléatoire gaussien $Z(\cdot)$, nous supposons que sous l'hypothèse nulle $Z(\cdot)$ est stationnaire d'ordre 2, *i.e.* sa fonction de covariance $C_Z(\mathbf{x}, \mathbf{y})$ ne dépend que de $\mathbf{x} - \mathbf{y}$. Les outils géostatistiques permettent d'obtenir une interpolation optimale du gradient. La variable étant supposée spatialisée et faiblement échantillonnée, le gradient prédit est gaussien et non stationnaire. Nous nous intéressons aux maxima locaux d'un champ gaussien.

Différents travaux portent sur la distribution du maximum d'un champ aléatoire. C'est en particulier le cas de Adler (1981, 2000) et Worsley dans les années 1990. Dans ces travaux l'espérance de la caractéristique d'Euler de l'ensemble d'excursion $\mathcal{A}_y = \mathcal{A}_y(Y, \mathcal{D}) = \{\mathbf{x} \in \mathcal{D} : Y(\mathbf{x}) \geq y\}$ permet de donner une approximation de la distribution du maximum du champ aléatoire $Y(\cdot)$ (équation (4.2)). En effet, le nombre de composantes connexes est difficile à déterminer. La caractéristique d'Euler, pour laquelle il existe de nombreux résultats [Adler (1981, 2000)], est proche du nombre de composantes connexes (section 4.1). Lorsque y est suffisamment grand, chaque maximum local au dessus de y engendre un ensemble d'excursion ellipsoïde dont la caractéristique d'Euler vaut un, ainsi la caractéristique vaut approximativement le nombre de composantes connexes sur le domaine.

Le problème dans l'approximation de la distribution du maximum du champ aléatoire $Y(\cdot)$ par l'espérance de la caractéristique d'Euler est qu'il n'existe pas de lien direct entre les deux objets, et que par conséquent rien ne permet de conclure sur l'erreur d'approximation.

Delmas (2001) et Azaïs et Delmas (2002) ont pu démontrer des résultats plus généraux que ceux de Adler et Worsley. Ils ont établi le développement asyomptotique de l'espérance du nombre de maxima locaux pour un champ gaussien dans le cas général, *i.e.* en n'imposant pas la stationnarité du champ aléatoire.

Dans notre approche, nous ne nous intéressons pas précisément au nombre de maxima locaux d'un champ aléatoire, mais à la surface des composantes connexes d'un ensemble d'excursion. En effet, notre méthode se schématise de la façon suivante. Nous testons localement si $H_0(\mathbf{x})$: "$\mathbb{E}\left[Z(\cdot)\right]$ est constante dans le voisinage de $\mathbf{x}$", contre $H_1(\mathbf{x})$: "$\mathbb{E}\left[Z(\cdot)\right]$ possède une discontinuité en $\mathbf{x}$". La statistique associée à ce test est une forme quadratique de la prédiction du gradient normé par sa variance et a pour distribution une χ^2 à deux degrés de liberté, $\chi^2(2)$. Les ZCAs potentielles sont définies comme l'ensemble des points où l'hypothèse nulle est rejétée, *i.e.* l'ensemble d'excursion de la statistique de test local au dessus du $(1-\alpha)$-quantile de la $\chi^2(2)$. Afin d'agréger les tests locaux pour déterminer si $\mathbf{Z}$ est issu d'un champ aléatoire stationnaire (hypothèse nulle) ou d'un champ présentant des discontinuités (alternative), nous travaillons sur les parties connexes des ZCAs potentielles détectées. Le test utilise la distribution asymptotique de la taille de chaque composante connexe sous l'hypothèse nulle. Un niveau de significativité global, $1-\eta$, est donc nécessaire. η correspond au faux taux de découverte global ou gFDR (*global False Discovery Rate*) et est par exemple fixé à 5%. La valeur de α permettant d'atteindre le gFDR est fonction de η, mais dépend aussi d'autres paramètres tels que la discrétisation de la grille utilisée en pratique, la portée du modèle de covariance, la densité d'échantillonnage, ... La théorie reliant α à η est celle des ensembles d'excursion de champs de χ^2 développée dans Adler (1981, 2000), Aronowich et Adler (1988), Worsley (1994) et Cao (1999). Les champs de χ^2 standards sont définis comme la somme du carré de champs aléatoires gaussiens stationnaires, indépendants et identiquement distribués (section 2.6). Cependant dans notre cas les champs gaussiens ne sont ni indépendants, ni stationnaires. Aussi, nous avons développé des résultats théoriques pour de tels champs de χ^2 non stationnaires.

Cette partie s'organise de la façon suivante. Dans le chapitre 5 nous décrivons le test

local de détection de ZCAs et illustrons ces résultats à travers un exemple sur simulations. Dans le chapitre 6 nous développons le test de significativité et tous les résultats théoriques liés à son élaboration. L'exemple précédent est utilisé pour illustrer les résultats de ce test. Enfin, nous proposons des solutions aux problèmes liés à l'implémentation de la méthode (chapitre 7).

Chapitre 5

Détection locale des ZCAs

5.1 Notations

Soit $\mathbf{Z} = (Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n))'$ un échantillon de valeurs réelles d'un champ aléatoire gaussien $Z(\cdot)$ en les points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ d'un domaine $\mathcal{D}$ de $\mathbb{R}^2$. Tout point $\mathbf{x}$ de $\mathcal{D}$ a pour coordonnées $(x^1, x^2)'$. Notons $C_Z(\mathbf{x}, \mathbf{y}) = \mathrm{Cov}\,(Z(\mathbf{x}), Z(\mathbf{y}))$ la fonction de covariance de ce champ et faisons les hypothèses suivantes :

$\mathcal{H}_1$: la fonction de covariance de $Z(\cdot)$ est stationnaire :

$$C_Z(\mathbf{x}, \mathbf{y}) = C_Z(\mathbf{x} - \mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathcal{D},$$

$\mathcal{H}_2$: $C_Z(\mathbf{h})$ est indéfiniment différentiable pour tout $\mathbf{h}$ tel que $\|\mathbf{h}\| > 0$.

L'hypothèse $\mathcal{H}_1$ est usuelle en géostatistique. Elle est nécessaire lors de l'estimation de la fonction de covariance lorsqu'il n'y a pas de répétition des données et que la densité d'échantillonnage n'est pas très élevée. Nous verrons dans la section 8.2 que cette hypothèse peut être affaiblie et que considérer que $Z(\cdot)$ est un champ aléatoire vérifiant les conditions de stationnarité intrinsèque suffit. Cependant pour plus de clarté nous considèrerons l'hypothèse plus restrictive $\mathcal{H}_1$. L'hypothèse $\mathcal{H}_2$ est moins usuelle car elle porte sur la régularité de la fonction de covariance dans un voisinage de $\mathbf{0}$. Elle est vérifiée par un grand nombre de fonctions de covariance. C'est par exemple le cas des fonctions de covariance de la classe de Matérn (section 3.3), qui inclut le modèle de covariance exponentiel. Remarquons qu'aucune hypothèse n'est exigée en $\mathbf{h} = \mathbf{0}$. En particulier la continuité et la différentiabilité à l'origine ne sont pas requises. L'hypothèse $\mathcal{H}_2$ est allégée dans la section 8.2, seule la différentiabilité à l'ordre 3 de $C_Z(\mathbf{h})$, pour tout $\mathbf{h}$ tel que $\|\mathbf{h}\| > 0$,

est exigée. Nous verrons que la fonction de covariance sphérique peut aussi être utilisée.

Nous cherchons à tester si $\mathbf{Z}$ est issu d'un champ aléatoire stationnaire, ce qui se traduit par l'absence de ZCAs. L'hypothèse nulle d'absence de ZCAs et l'hypothèse alternative associées à ce test sont définies par :

H_0 : $\mathbb{E}[Z(\mathbf{x})] = m$, pour tout $\mathbf{x} \in \mathcal{D}$,

H_1 : il existe des courbes $\Gamma \subset \mathcal{D}$ sur lesquelles $\mathbb{E}[Z(\cdot)]$ est discontinue.

Nous ne faisons aucune hypothèse sur Γ. Il peut s'agir d'une courbe régulière dans le plan, d'un ensemble de courbes s'intersectant ou non. De plus, sans perte de généralité, nous considèrerons que $m = 0$ sous l'hypothèse nulle. Le cas général où $m \neq 0$ est inconnue sera discuté section 8.1.

Comme sous l'hypothèse nulle $Z(\cdot)$ est un champ aléatoire gaussien centré, le prédicteur optimal en $\mathbf{x}$, noté $Z^*(\mathbf{x})$, est le krigeage simple (section 3.1),

$$Z^*(\mathbf{x}) = C(\mathbf{x})'\mathbf{C}^{-1}\mathbf{Z}, \tag{5.1}$$

où $C(\mathbf{x})$, de dimension $n \times 1$, est le vecteur de covariance entre $\mathbf{x}$ et les points de données :

$$C(\mathbf{x}) = (C_Z(\mathbf{x} - \mathbf{x}_1), \ldots, C_Z(\mathbf{x} - \mathbf{x}_n))',$$

$\mathbf{C}$, de dimension $n \times n$, est la matrice de covariance entre les points de données :

$$\mathbf{C} = \mathbb{E}[\mathbf{Z}\mathbf{Z}'],$$

d'éléments $\mathbf{C}_{[ij]} = C_Z(\mathbf{x}_i - \mathbf{x}_j)$, $i, j = 1, \ldots, n$ et $\mathbf{Z}$, de dimension $n \times 1$, représente une réalisation donnée :

$$\mathbf{Z} = (Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n))'.$$

La fonction de covariance de cet estimateur est :

$$\mathrm{Cov}(Z^*(\mathbf{x}), Z^*(\mathbf{y})) = C(\mathbf{x})'\mathbf{C}^{-1}\mathbb{E}[\mathbf{Z}\mathbf{Z}']\,\mathbf{C}^{-1}C(\mathbf{y}) = C(\mathbf{x})'\mathbf{C}^{-1}C(\mathbf{y}),$$

ce qui montre que le champ $Z^*(\cdot)$ est non stationnaire.

Désormais nous considèrerons que x n'est pas un point d'échantillonnage et que la fonction de covariance est supposée connue.

5.2 Test local de changement abrupt

Dans un premier temps nous détectons localement la présence d'une discontinuité au point $\mathbf{x} \in \mathcal{D}$. Il y a un changement abrupt en un point $\mathbf{x}$ si le gradient du champ $Z(\cdot)$ en $\mathbf{x}$ est "élevé". Par conséquent, afin de définir le terme "élevé" nous avons besoin d'une part d'un prédicteur du gradient de $Z(\cdot)$ et d'autre part d'un test statistique. Les sauts de discontinuité de l'espérance du champ correspondent à un gradient local élevé.
Les hypothèses locales se modélisent par :

$H_0(\mathbf{x})$: $\mathbb{E}[Z(\mathbf{y})] = 0$, pour tout $\mathbf{y}$ dans un voisinage de $\mathbf{x}$,

$H_1(\mathbf{x})$: $\mathbf{x} \in \Gamma$, où Γ est l'ensemble des courbes de discontinuité.

Dans la suite nous noterons ∂_i la dérivée partielle le long de la coordonnée x^i, $i = 1, 2$: $\partial_i = \partial/\partial x^i$. D'après la section 3.2, le gradient $W(\mathbf{x})$ de $Z^*(\mathbf{x})$ existe et nous avons :

$$W(\mathbf{x}) = (W_1(\mathbf{x}), W_2(\mathbf{x}))' = (\partial_1 Z^*(\mathbf{x}), \partial_2 Z^*(\mathbf{x}))', \tag{5.2}$$

avec

$$\partial_i Z^*(\mathbf{x}) = \partial_i C'(\mathbf{x}) \mathbf{C}^{-1} \mathbf{Z},$$

où $\partial_i C'(\mathbf{x}) = \partial_i \left(C_Z(\mathbf{x} - \mathbf{x}_1), \ldots, C_Z(\mathbf{x} - \mathbf{x}_n)\right) = (\partial_i C_Z(\mathbf{x} - \mathbf{x}_1), \ldots, \partial_i C_Z(\mathbf{x} - \mathbf{x}_n))$. Si $C_Z(\mathbf{h})$ est deux fois différentiable en $\mathbf{h} = \mathbf{0}$, alors, par linéarité, $W(\mathbf{x})$ est le meilleur prédicteur linéaire sans biais du gradient de $Z(\mathbf{x})$. Si $C_Z(\mathbf{h})$ n'est pas deux fois différentiable en $\mathbf{h} = \mathbf{0}$, ce qui est par exemple le cas pour une fonction de covariance exponentielle, alors $Z(\mathbf{x})$ n'a pas de dérivée mais $Z^*(\mathbf{x})$ est différentiable, sauf aux points d'échantillonnage. Ainsi d'après la relation (3.5), $W(\mathbf{x})$ est un prédicteur local du gradient.
Notons $\mathbf{\Sigma}(\mathbf{x})$ la matrice de covariance de $W(\mathbf{x})$:

$$\mathbf{\Sigma}(\mathbf{x}) = \mathbb{E}[W(\mathbf{x})W'(\mathbf{x})] = \partial C'(\mathbf{x}) \mathbf{C}^{-1} \partial C(\mathbf{x}),$$

avec $\partial C'(\mathbf{x}) = (\partial_1 C'(\mathbf{x}), \partial_2 C'(\mathbf{x}))'$ de dimension $2 \times n$. Comme $Z(\cdot)$ est un champ gaussien, nous avons $W(\mathbf{x}) \sim \mathcal{N}_2(0, \mathbf{\Sigma}(\mathbf{x}))$.

Nous supposerons dans toute la suite que $\mathbf{\Sigma}(\mathbf{x})$ est strictement positive et inversible quelquesoit $\mathbf{x} \in \mathcal{D}$.
Nous excluons donc la possibilité que deux points d'échantillonnage puissent avoir les mêmes coordonnées.

En un point $\mathbf{x}$, le champ $T(\cdot)$,

$$T(\mathbf{x}) = W'(\mathbf{x}) \mathbf{\Sigma}(\mathbf{x})^{-1} W(\mathbf{x})$$

a pour distribution une χ^2 à deux degrés de liberté sous l'hypothèse locale nulle. Ce champ compare localement le gradient estimé à sa variance. L'hypothèse locale nulle est rejetée si la statistique $T(\mathbf{x})$ est supérieure à t_α, quantile d'ordre $1-\alpha$ de la distribution $\chi^2(2)$.

Nous pouvons désormais définir les ZCAs potentielles. Elles correspondent à l'ensemble d'excursion de la statistique de test local au dessus du niveau t_α :

$$\mathcal{A}_{t_\alpha} = \{\mathbf{x} \,:\, T(\mathbf{x}) \geq t_\alpha\}\,.$$

5.3 Illustration

En pratique, cette procédure est appliquée aux nœuds d'une grille superposée sur le domaine $\mathcal{D}$, ce qui permet de représenter et de visualiser les ZCAs. Voici une illustration sur simulation.

Dans un carré unitaire, nous avons simulé un échantillon $\mathbf{Z}$ de 100 points dispersés aléatoirement, issus d'un champ aléatoire gaussien $\mathcal{N}(0,1)$ de fonction de covariance exponentielle $C_Z(\mathbf{h}) = \exp(-\|\mathbf{h}\|/b)$, de portée $b = 0.1$. Nous avons introduit une discontinuité le long de la ligne $x^1 = 0.4$ en ajoutant une constante a aux points d'échantillonnage situés dans la zone $x^1 < 0.4$. Le cas $a = 0$ correspond à H_0 (absence de discontinuité). La figure 5.1 représente une réalisation pour $a = 0$ (figure 5.1a) et pour $a = 2.5$ (figure 5.1b), avec en pointillés la localisation de la discontinuité. Les cercles correspondent aux points

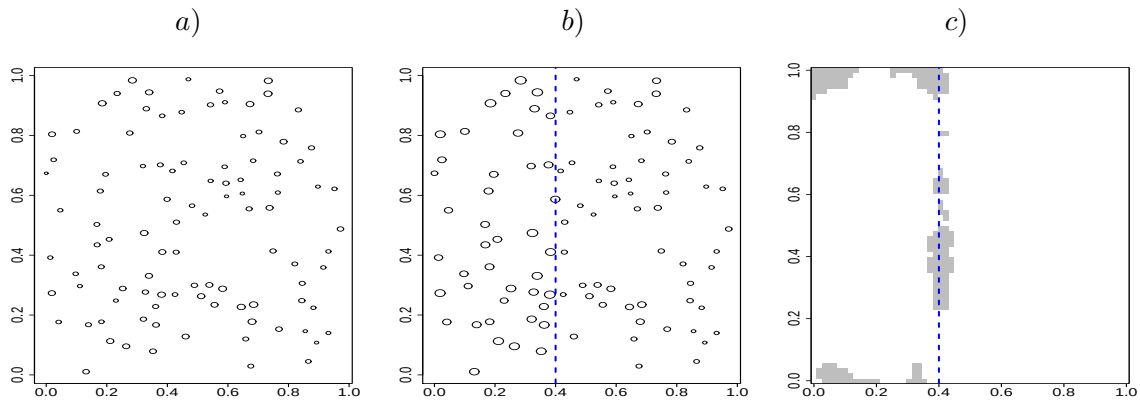


FIG. 5.1: *Echantillon de 100 points répartis aléatoirement, issus d'un processus : a) gaussien stationnaire, centré-réduit, de fonction de covariance exponentielle de portée $b = 0.1$, b) avec une discontinuité le long des pointillés ($a = 2.5$). c) ZCAs potentielles (en gris) obtenues sous l'alternative.*

d'échantillonnage et leur taille est proportionnelle à leur valeur. La figure 5.1c donne les résultats des tests locaux au niveau de confiance $1-\alpha = 0.9994$ pour $a = 2.5$. La grille

d'interpolation utilisée est de dimension 60×60. En chaque pixel de cette grille nous testons si la statistique de T est supérieure à $t_{0.0006} = 14.927$. Lorsque c'est le cas, le pixel correspondant est marqué, sur la figure 5.1*c* ils sont colorés en gris. Aucun pixel n'est marqué sous H_0. En revanche, en présence d'une discontinuité, une bande verticale est visible autour de la ligne de discontinuité. Le champ étant supposé d'espérance nulle à l'extérieur du domaine, un effet de bord apparaît dans les deux coins gauches du domaine.

Les ZCAs sont définies comme l'ensemble des lieux où $H_0(\mathbf{x})$ est rejetée. Dans le cas d'un champ stationnaire, nous devons avoir une fréquence de rejet de l'ordre de α sous H_0, et nous nous attendons à ce que les pixels de la grille où l'hypothèse nulle est rejetée soient distribués aléatoirement. En revanche, en présence d'une discontinuité de l'espérance du champ aléatoire, les pixels correspondant au rejet de H_0 devraient être plus nombreux et situés le long ou à proximité de la discontinuité.

Notre prochaine étape consiste à proposer un test global afin de rejeter H_0 versus H_1. Cela correspond à tester simultanément les tests locaux $H_0(\mathbf{x}_i)$ versus $H_1(\mathbf{x}_i)$, où i est un indice désignant les points appartenant aux ZCAs potentielles détectées. Comme ces tests sont très dépendants, nous ne pouvons pas utiliser la méthode proposée dans Shen *et al.* (2002) (section 7.2). Leur procédure, appelée le *Enhanced False Discovery Rate* (EFDR), prend en compte l'information contenue dans toutes les valeurs critiques et suppose l'indépendance entre les tests. Nous allons suivre une approche similaire à celle usuellement utilisée en fMRI (*fonctional Magnetic Resonance Imaging*) et proposée dans Worsley (2001) et Cao (1999), dont les principaux résultats ont été donnés dans la section 4.3. Cette approche va nous permettre d'obtenir la distribution asymptotique de la taille des composantes connexes des ZCAs afin d'en tester la significativité.

Chapitre 6

Détection globale des ZCAs

Dans cette partie nous allons établir des résultats théoriques sur le champ $T(\cdot)$ sous l'hypothèse nulle, en particulier sur sa courbure en un maximum local. Les ZCAs potentielles correspondent à un ensemble d'excursion, $\mathcal{A}_{t_\alpha}$, du champ de $\chi^2(2)$, $T(\cdot)$, au dessus du niveau t_α :

$$\mathcal{A}_{t_\alpha} = \{\mathbf{x} : T(\mathbf{x}) \geq t_\alpha\}.$$

Nous montrerons que pour des seuils élevés t_α, *i.e.* lorsque α tend vers zéro, la taille d'une composante connexe $\mathcal{C}_{t_\alpha}$ de l'ensemble d'excursion est reliée à la courbure du champ $T(\cdot)$ au maximum local de $\mathcal{C}_{t_\alpha}$. Ce résultat nous permettra de tester la significativité de chaque composante connexe. Ainsi, pour chacune des composantes connexes, nous allons tester si elle est issue :

H_0 : d'un champ aléatoire stationnaire,

H_1 : d'un champ aléatoire présentant de fortes variations.

6.1 Courbure du champ $T(\cdot)$

Effectuons le changement de variables $U(\mathbf{x}) = \mathbf{\Sigma}^{-1/2}(\mathbf{x})W(\mathbf{x})$ de sorte que $T(\mathbf{x})$ se décompose en la somme du carré de deux champs gaussiens $\mathcal{N}(0,1)$. $\mathbf{\Sigma}^{-1/2}(\mathbf{x})$ représente l'inverse de la racine carrée de la matrice $\mathbf{\Sigma}(\mathbf{x})$. La racine carrée d'une matrice n'est pas unique. Comme $\mathbf{\Sigma}(\mathbf{x})$ est définie positive, nous avons choisi la décomposition de Cholesky dont les éléments diagonaux sont positifs (annexe A).

Proposition 3 *Le champ $T(\mathbf{x})$ se décompose en $T(\mathbf{x}) = U_1^2(\mathbf{x}) + U_2^2(\mathbf{x})$, avec :*

$$U_1(\mathbf{x}) = \frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})} \text{ et } U_2(\mathbf{x}) = \frac{1}{\sqrt{1-\rho^2(\mathbf{x})}} \left\{ \frac{W_2(\mathbf{x})}{\sigma_2(\mathbf{x})} - \rho(\mathbf{x}) \frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})} \right\}, \tag{6.1}$$

où $\sigma_i(\mathbf{x}) = \sqrt{\mathbf{\Sigma}_{[ii]}(\mathbf{x})}$ *et* $\rho(\mathbf{x}) = \mathbf{\Sigma}_{[12]}(\mathbf{x})/\sigma_1(\mathbf{x})\sigma_2(\mathbf{x})$.

La démonstration de cette proposition est donnée dans l'annexe A.

Propriétés de $U(\cdot)$

Soit $\rho_{ij}(\mathbf{x}, \mathbf{y})$ la fonction de corrélation entre $W_i(\mathbf{x})$ et $W_j(\mathbf{y})$:

$$\rho_{ij}(\mathbf{x}, \mathbf{y}) = \mathbb{E}\left[\frac{W_i(\mathbf{x})}{\sigma_i(\mathbf{x})}\frac{W_j(\mathbf{y})}{\sigma_j(\mathbf{y})}\right] = \frac{\partial_i C'(\mathbf{x})\mathbf{C}^{-1}\partial_j C(\mathbf{y})}{\sigma_i(\mathbf{x})\sigma_j(\mathbf{y})}. \tag{6.2}$$

Alors, $\rho_{12}(\mathbf{x}, \mathbf{x}) = \rho(\mathbf{x})$ et $\rho_{11}(\mathbf{x}, \mathbf{x}) = \rho_{22}(\mathbf{x}, \mathbf{x}) = 1$.

Proposition 4 *Pour* $U_1(\cdot)$ *et* $U_2(\cdot)$ *définis dans la proposition 3,*

(i) $U_1(\mathbf{x})$ *et* $U_2(\mathbf{x})$ *sont indépendants,*

(ii)
$$\begin{aligned}
\mathbb{E}\left[U_1(\mathbf{x})U_1(\mathbf{y})\right] &= \rho_{11}(\mathbf{x}, \mathbf{y}),\\
\mathbb{E}\left[U_1(\mathbf{x})U_2(\mathbf{y})\right] &= \left\{\rho_{12}(\mathbf{x}, \mathbf{y}) - \rho(\mathbf{y})\rho_{11}(\mathbf{x}, \mathbf{y})\right\}/\sqrt{1-\rho^2(\mathbf{y})},\\
\mathbb{E}\left[U_2(\mathbf{x})U_2(\mathbf{y})\right] &= \{\rho_{22}(\mathbf{x}, \mathbf{y}) - \rho(\mathbf{x})\rho_{12}(\mathbf{x}, \mathbf{y}) - \rho(\mathbf{y})\rho_{21}(\mathbf{x}, \mathbf{y})\\
&\quad +\rho(\mathbf{x})\rho(\mathbf{y})\rho_{11}(\mathbf{x}, \mathbf{y})\}/\sqrt{(1-\rho^2(\mathbf{x}))(1-\rho^2(\mathbf{y}))}.
\end{aligned}$$

Ces propriétés sont démontrées dans l'annexe B.
La proposition 4 indique que les champs gaussiens $U_1(\cdot)$ et $U_2(\cdot)$ ne sont pas stationnaires, leurs fonctions de covariance dépendent de $\mathbf{x}$ et $\mathbf{y}$. De plus ces champs ne sont pas identiquement distribués, ni indépendants :

$$\mathbb{E}\left[U_1(\mathbf{x})U_1(\mathbf{y})\right] \neq \mathbb{E}\left[U_2(\mathbf{x})U_2(\mathbf{y})\right], \quad \mathbb{E}\left[U_1(\mathbf{x})U_2(\mathbf{y})\right] \neq 0.$$

Des résultats sur la géométrie des champs χ^2 définis à la section 2.6 sont exposés dans Adler (1981) et Cao (1999). D'après la proposition 4 nous ne pouvons pas utiliser ces résultats. En effet, $T(\mathbf{x})$ ne se décompose pas en la somme du carré de champs aléatoires stationnaires, indépendants et identiquement distribués. Des propriétés sur les champs $U_i(\cdot), i = 1, 2$ sont énoncées ci-dessous en vue de les utiliser lors de la démonstration de résultats à venir sur la géométrie d'un champ de χ^2 non stationnaire.
Notons $\partial^{(k)}U_1(\mathbf{x})$ et $\partial^{(l)}U_2(\mathbf{x})$ les dérivées k-ième de $U_1(\mathbf{x})$ et l-ième de $U_2(\mathbf{x})$, avec

$$\partial^{(k)} = \frac{\partial^{(k)}}{\partial(x^1)^{k_1}\partial(x^2)^{k_2}}, \; k = k_1 + k_2.$$

L'hypothèse $\mathcal{H}_2$, concernant l'indéfinie différentiabilité de $C_Z(\mathbf{h})$ pour tout $\|\mathbf{h}\| > 0$, permet de montrer les deux résultats suivants.

Lemme 4 *Pour $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ les champs définis dans la proposition 3, si la fonction de covariance de $Z(\mathbf{x})$ est au moins 2-fois différentiable, alors pour $i \neq j$:*

(i) $\frac{\partial}{\partial x^k}\mathbb{E}\left[U_i(\mathbf{x}) \mid U_j(\mathbf{x})\right] = \mathbb{E}\left[\frac{\partial U_i(\mathbf{x})}{\partial x^k} \mid U_j(\mathbf{x})\right]$, $k \in \{1,2\}$,

(ii) $\frac{\partial}{\partial x^k}\mathbb{E}\left[U_i(\mathbf{x}) \mid \frac{\partial U_j(\mathbf{x})}{\partial x^l}\right] = \mathbb{E}\left[\frac{\partial U_i(\mathbf{x})}{\partial x^k} \mid \frac{\partial U_j(\mathbf{x})}{\partial x^l}\right]$, $k,l \in \{1,2\}$.

La démonstration des ces deux propriétés est donnée dans l'annexe B.
Le lemme 4 a permis de montrer la proposition suivante sur l'indépendance des dérivées k-ième de $U_1(\mathbf{x})$ et l-ième de $U_2(\mathbf{x})$.

Proposition 5 *Pour $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ les champs définis dans la proposition 3, si la fonction de covariance de $Z(\mathbf{x})$ est au moins $2m$-fois différentiable, alors pour $k,l \leq m$:*

$$\mathbb{E}\left[\partial^{(k)}U_1(\mathbf{x})\partial^{(l)}U_2'(\mathbf{x})\right] = \mathbf{0}.$$

Ce résultat est démontré par récurrence dans l'annexe B.

Introduisons maintenant quelques notations. Pour un champ aléatoire, $T(\cdot)$, supposé deux fois différentiable, $\dot{T}(\mathbf{x})$ est le vecteur $(\partial_1 T(\mathbf{x}), \partial_2 T(\mathbf{x}))'$ et $\ddot{\mathbf{T}}(\mathbf{x})$ la 2×2 matrice d'éléments $\partial^2_{kl}T(\mathbf{x})$, $1 \leq k,l \leq 2$. Définissons également la matrice de covariance du gradient du champ $U_i(\cdot), i = 1,2$:

$$\mathbf{\Lambda}_i(\mathbf{x}) = \mathbb{E}\left[\dot{U}_i(\mathbf{x})\dot{U}_i'(\mathbf{x})\right] = \mathrm{Var}\left(\dot{U}_i(\mathbf{x})\right).$$

Les éléments des matrices $\mathbf{\Lambda}_i(\mathbf{x})$, $i = 1,2$ sont donnés dans le résultat suivant dans lequel nous enlevons la dépendance en $\mathbf{x}$ pour alléger les notations et posons : $D_1 = D_1(\mathbf{x}) = \partial_1 C(\mathbf{x})$ et $D_2 = D_2(\mathbf{x}) = \partial_2 C(\mathbf{x})$.

Résultat 1 *Pour k,l = 1,2, les éléments $\mathbf{\Lambda}_{i\ [kl]}$ de la matrice $\mathbf{\Lambda}_i$, $i = 1,2$ sont :*

$$\begin{aligned}
\mathbf{\Lambda}_{1\ [kl]} &= \{\partial_k D_1'\mathbf{C}^{-1}\partial_l D_1 - \partial_k\sigma_1\partial_l\sigma_1\}/\sigma_1^2 \\
\mathbf{\Lambda}_{2\ [kl]} &= \{\partial_k D_2'\mathbf{C}^{-1}\partial_l D_2/\sigma_2^2 - \partial_k\sigma_2\partial_l\sigma_2/\sigma_2^2 + \partial_k\rho\partial_l\rho - \partial_k\rho\partial_l D_2'\mathbf{C}^{-1}D_1/\sigma_1\sigma_2 \\
&\quad -\partial_l\rho\partial_k D_2'\mathbf{C}^{-1}D_1/\sigma_1\sigma_2 + \rho[(\partial_k\sigma_2/\sigma_2)(\partial_l D_1'\mathbf{C}^{-1}D_2/\sigma_1\sigma_2 + \partial_l\rho) \\
&\quad +(\partial_l\sigma_2/\sigma_2)(\partial_k D_1'\mathbf{C}^{-1}D_2/\sigma_1\sigma_2 + \partial_k\rho) + (\partial_l\sigma_1/\sigma_1)(\partial_k D_2'\mathbf{C}^{-1}D_1/\sigma_1\sigma_2) \\
&\quad +(\partial_k\sigma_1/\sigma_1)(\partial_l D_2'\mathbf{C}^{-1}D_1/\sigma_1\sigma_2) - \partial_k D_1'\mathbf{C}^{-1}\partial_l D_2/\sigma_1\sigma_2 - \partial_k D_2'\mathbf{C}^{-1}\partial_l D_1/\sigma_1\sigma_2] \\
&\quad +\rho^2[\partial_k D_1'\mathbf{C}^{-1}\partial_l D_1/\sigma_1^1 - \partial_k\sigma_1\partial_l\sigma_1/\sigma_1^1 - \partial_k\sigma_1\partial_l\sigma_2/\sigma_1\sigma_2 \\
&\quad -\partial_k\sigma_2\partial_l\sigma_1/\sigma_1\sigma_2]\}/(1-\rho^2) - \rho^2(\partial_k\rho\partial_l\rho)/(1-\rho^2)^2.
\end{aligned}$$

Preuve : En ce qui concerne $\mathbf{\Lambda}_{1\ [kl]}$ la démonstration est rapide.

$$\begin{aligned}\mathbf{\Lambda}_{1\ [kl]} &= \mathbb{E}[\partial_k U_1 \partial_l U_1] \\ &= \mathbb{E}[\partial_k (W_1/\sigma_1)\partial_l (W_1/\sigma_1)] \\ &= \{\partial_k D_1' \mathbf{C}^{-1} \partial_l D_1 - \partial_k \sigma_1 \partial_l \sigma_1\}/\sigma_1^2.\end{aligned}$$

En revanche pour $\mathbf{\Lambda}_{2\ [kl]}$ le développement de $\mathbb{E}[\partial_k U_2 \partial_l U_2]$ est beaucoup plus long (annexe B). Le résultat est obtenu en partant du même principe que pour $\mathbf{\Lambda}_{1\ [kl]}$ et en utilisant,

$$\partial_k \rho = (\partial_k D_1' \mathbf{C}^{-1} D_2 + \partial_k D_2' \mathbf{C}^{-1} D_1)/\sigma_1\sigma_2 - \rho\left(\partial_k \sigma_1/\sigma_1 + \partial_k \sigma_2/\sigma_2\right).$$

□

Proposition 6 *Conditionnellement à $U_i(\cdot)$, $\dot{U}_i(\cdot)$ et $\ddot{\mathbf{U}}_i(\cdot)$ en $\mathbf{0}$, dans un voisinage de $\mathbf{0}$ les champs $U_i(\cdot)$ définis dans la proposition 3 ont la même distribution que :*

$$U_i(\mathbf{0}) + \mathbf{x}'\dot{U}_i(\mathbf{0}) + \frac{1}{2}\mathbf{x}'\ddot{\mathbf{U}}_i(\mathbf{0})\mathbf{x} + \epsilon_i(\mathbf{x}), \tag{6.3}$$

où $\epsilon_i(\mathbf{x})$ est un champ gaussien de moyenne $\mathbb{E}\left[\epsilon_i(\mathbf{x})\right] = o\left(\|\mathbf{x}\|^2\right)$ et de fonction de covariance $\mathbb{E}\left[\epsilon_i(\mathbf{x})\epsilon_i(\mathbf{y})\right] = o\left(\|\mathbf{x}\|^2\|\mathbf{y}\|^2\right)$.

Preuve : Les champs $U_i(\mathbf{x}), i = 1, 2$ peuvent s'écrire $U_i(\mathbf{x}) = A_i'(\mathbf{x})\mathbf{C}^{-1}\mathbf{Z}$ où $A_i(\mathbf{x}), i = 1, 2$ est un vecteur de dimension n défini par :

$$A_1(\mathbf{x}) = \frac{\partial_1 C'(\mathbf{x})}{\sigma_1(\mathbf{x})} \text{ et } A_2(\mathbf{x}) = \frac{1}{\sqrt{1-\rho(\mathbf{x})}}\left\{\frac{\partial_2 C'(\mathbf{x})}{\sigma_2(\mathbf{x})} - \rho(\mathbf{x})\frac{\partial_1 C'(\mathbf{x})}{\sigma_1(\mathbf{x})}\right\}.$$

Nous avons :

$$U_i(\mathbf{x}) = \left(A_{i,1}(\mathbf{x}), \ldots, A_{i,n}(\mathbf{x})\right)\mathbf{C}^{-1}\mathbf{Z} = \sum_{k=1}^{n} A_{i,k}(\mathbf{x})\left[\mathbf{C}^{-1}\mathbf{Z}\right]_k,$$

où $A_{i,k}$ et $[\mathbf{C}^{-1}\mathbf{Z}]_k$ sont les k-ièmes éléments des vecteurs A_i et $\mathbf{C}^{-1}\mathbf{Z}$. Comme $A_i(\mathbf{x}), i = 1, 2$ est indéfiniment différentiable presque sûrement et au sens des moindres carrés, le développement de Taylor à l'ordre 2 de $A_{i,k}(\mathbf{x})$ autour de $\mathbf{0}$,

$$A_{i,k}(\mathbf{x}) = A_{i,k}(\mathbf{0}) + \mathbf{x}'\dot{A}_{i,k}(\mathbf{0}) + \frac{1}{2}\mathbf{x}'\ddot{A}_{i,k}(\mathbf{0})\mathbf{x} + O\left(\|\mathbf{x}\|^3\right)$$

implique :

$$\begin{aligned} U_i(\mathbf{x}) &= U_i(\mathbf{0}) + \mathbf{x}'\dot{U}_i(\mathbf{0}) + \frac{1}{2}\mathbf{x}'\ddot{U}_i(\mathbf{0})\mathbf{x} + \sum_{k=1}^{n} R_{i,k}(\mathbf{x}) \left[\mathbf{C}^{-1}\mathbf{Z}\right]_k \\ &= U_i(\mathbf{0}) + \mathbf{x}'\dot{U}_i(\mathbf{0}) + \frac{1}{2}\mathbf{x}'\ddot{\mathbf{U}}_i(\mathbf{0})\mathbf{x} + R_i'(\mathbf{x})\mathbf{C}^{-1}\mathbf{Z}, \end{aligned}$$

avec $R_i(\mathbf{x}) = (R_{i,1}(\mathbf{x}), \ldots, R_{i,n}(\mathbf{x}))'$, $R_{i,k}(\mathbf{x}) \leq M_{i,k}\|\mathbf{x}\|^3$ et les $M_{i,k}$ sont des constantes. Conditionnellement à $U_i(\cdot)$, $\dot{U}_i(\cdot)$ et $\ddot{\mathbf{U}}_i(\cdot)$ en $\mathbf{0}$,

$$U_i(\mathbf{0}) + \mathbf{x}'\dot{U}_i(\mathbf{0}) + \frac{1}{2}\mathbf{x}'\ddot{U}_i(\mathbf{0})\mathbf{x} + \epsilon_i(\mathbf{x}),$$

avec $\epsilon_i(\mathbf{x}) = R_i'(\mathbf{x})\mathbf{C}^{-1}\mathbf{Z}_{|U_i(\mathbf{0}),\dot{U}_i(\mathbf{0}),\ddot{\mathbf{U}}_i(\mathbf{0})}$.
Si nous notons $\mathbf{V}_i = (U_i(\mathbf{0}), \partial_1 U_i(\mathbf{0}), \partial_2 U_i(\mathbf{0}), \partial^2_{11} U_i(\mathbf{0}), \partial^2_{12} U_i(\mathbf{0}), \partial^2_{22} U_i(\mathbf{0}))'$ le vecteur conditionnel, $K_{\mathbf{V}_i\mathbf{V}_i} = \mathbb{E}[\mathbf{V}_i\mathbf{V}_i']$ sa matrice de covariance et $K_{\mathbf{Z}\mathbf{V}_i} = \mathbb{E}[\mathbf{Z}\mathbf{V}_i']$ la matrice de covariance entre $\mathbf{Z}$ et $\mathbf{V}_i$, alors,

$$\mathbb{E}[\epsilon_i(\mathbf{x})] = R_i'(\mathbf{x})\mathbf{C}^{-1}\mathbb{E}\left[\mathbf{Z} \mid U_i(\mathbf{0}), \dot{U}_i(\mathbf{0}), \ddot{\mathbf{U}}_i(\mathbf{0})\right] = R_i'(\mathbf{x})\mathbf{C}^{-1}K_{\mathbf{Z}\mathbf{V}_i}K^{-1}_{\mathbf{V}_i\mathbf{V}_i}\mathbf{V}_i,$$

et

$$\begin{aligned} \mathbb{E}[\epsilon_i(\mathbf{x})\epsilon_i(\mathbf{y})] &= R_i'(\mathbf{x})\mathbf{C}^{-1}\mathbb{E}[\mathbf{Z}\mathbf{Z}' \mid \mathbf{V}_i]\,\mathbf{C}^{-1}R_i(\mathbf{y}) \\ &= R_i'(\mathbf{x})\mathbf{C}^{-1}\left(\mathbf{C} - K_{\mathbf{Z}\mathbf{V}_i}K^{-1}_{\mathbf{V}_i\mathbf{V}_i}K_{\mathbf{V}_i\mathbf{Z}}\right)\mathbf{C}^{-1}R_i(\mathbf{y}). \end{aligned}$$

Comme $R_i(\mathbf{x})$ (resp. $R_i(\mathbf{y})$) est $o(\|\mathbf{x}\|^2)$ (resp. $o(\|\mathbf{y}\|^2)$), le résultat est immédiat. □

La proposition 6 tient pour des hypothèses plus faibles que l'indéfinie différentiabilité de l'hypothèse $\mathcal{H}_2$. En effet, il est seulement nécessaire que $R_i(\mathbf{x})$, $K_{\mathbf{V}_i\mathbf{V}_i}$ et $K_{\mathbf{Z}\mathbf{V}_i}$ existent ce qui requiert la différentiabilité à l'ordre 3 de $C_Z(\mathbf{h})$ pour $\|\mathbf{h}\| > \mathbf{0}$.
Les distributions conditionnelles des dérivées secondes des champs $U_i(\mathbf{x})$ sachant $U(\mathbf{x})$ et $\dot{U}(\mathbf{x})$ sont déterminées dans le lemme suivant dont la démonstration est donnée dans l'annexe B. Ces propriétés permettront d'établir la distribution de la courbure du champ $T(\cdot)$, en particulier en un maximum local.

Lemme 5 *Pour $U_i(\cdot), i = 1, 2$ défini comme dans la proposition 3,*
(i) $U_i(\mathbf{x})$ et $\dot{U}_i(\mathbf{x})$ sont indépendants,
(ii) $Cov\left(\ddot{\mathbf{U}}_i(\mathbf{x}), \dot{U}_i(\mathbf{x})\dot{U}_i'(\mathbf{x}) \mid U(\mathbf{x})\right) = \mathbf{0}$,
(iii) $\ddot{\mathbf{U}}_i(\mathbf{x})_{|U(\mathbf{x})} \sim \mathcal{N}_{2\times 2}\left(-U_i(\mathbf{x})\mathbf{\Lambda}_i(\mathbf{x}), \mathbf{M}_i(\mathbf{\Lambda}_i(\mathbf{x}))\right)$,
où $\mathbf{M}_i(\mathbf{\Lambda}_i(\mathbf{x}))_{[kl]} = Cov(\partial^2_{kl}U_i(\mathbf{x}), \partial^2_{k'l'}U_i(\mathbf{x}) \mid U_i(\mathbf{x}))$ est la matrice de covariance conditionnelle des dérivées d'ordre deux de $U_i(\mathbf{x})$, elle ne dépend pas de $U_i(\mathbf{x})$.

(iv) $\partial^2_{kl}U_i(\mathbf{x})_{|U(\mathbf{x}),\dot{U}(\mathbf{x})} \sim \mathcal{N}\left(E_{\partial^2 U_i}, V_{\partial^2 U_i}\right)$,
avec $E_{\partial^2 U_i} = -U_i\mathbf{\Lambda}_i(\mathbf{x})_{[kl]} + \mathbb{E}\left[\partial^2_{kl}U_i(\mathbf{x})\dot{U}_i(\mathbf{x})\right]\mathbf{\Lambda}_i^{-1}\dot{U}_i(\mathbf{x})$
et $V_{\partial^2 U_i} = \mathbf{M}_i(\mathbf{\Lambda}_i(\mathbf{x}))_{[kl]} + \mathbb{E}\left[\partial^2_{kl}U_i(\mathbf{x})\dot{U}_i(\mathbf{x})\right]\mathbf{\Lambda}_i^{-1}(\mathbf{x})\mathbb{E}'\left[\partial^2_{kl}U_i(\mathbf{x})\dot{U}_i(\mathbf{x})\right]$.

Remarquons qu'il n'y a pas indépendance entre $\dot{U}_i(\mathbf{x})$ et $\ddot{\mathbf{U}}_i(\mathbf{x})$.

Propriétés de $T(\cdot)$

D'après les équations (2.12) et (2.13) et sous l'hypothèse $\mathcal{H}_2$, les dérivées premières et secondes de $T(\mathbf{x})$ s'écrivent :

$$\partial_k T(\mathbf{x}) = 2\sum_{i=1}^{2} U_i(\mathbf{x})\partial_k U_i(\mathbf{x}),\ k = 1, 2 \tag{6.4}$$

$$\partial^2_{kl}T(\mathbf{x}) = 2\sum_{i=1}^{2}\left(\partial_k U_i(\mathbf{x})\partial_l U_i(\mathbf{x}) + U_i(\mathbf{x})\partial^2_{kl}U_i(\mathbf{x})\right),\ k, l = 1, 2. \tag{6.5}$$

Proposition 7 *Pour le champ* $T(\mathbf{x}) = U_1^2(\mathbf{x}) + U_2^2(\mathbf{x})$, *avec* $U_1(\mathbf{x})$ *et* $U_2(\mathbf{x})$ *définis dans la proposition 3, nous avons :*
(i) $\dot{T}(\mathbf{x})_{|U(\mathbf{x})} \sim \mathcal{N}_2\left(\mathbf{0}, \sum_{i=1}^{2} 4U_i^2(\mathbf{x})\mathbf{\Lambda}_i(\mathbf{x})\right)$,
(ii) $\mathbb{E}\left[\partial^2_{kl}T(\mathbf{x}) \mid U(\mathbf{x})\right] = 2\sum_{i=1}^{2}\left(1 - U_i^2(\mathbf{x})\right)\mathbf{\Lambda}_{i[kl]}(\mathbf{x})$
et $Var(\partial^2_{kl}T(\mathbf{x}) \mid U(\mathbf{x})) = 4\sum_{i=1}^{2} U_i^2(\mathbf{x})\mathbf{M}_i(\mathbf{\Lambda}_i(\mathbf{x}))_{[kl]}$.

Cette proposition est démontrée dans l'annexe C.
Le théorème suivant établit la distribution asymptotique de la courbure du champ $T(\mathbf{x})$ lorsque $1/T(\mathbf{x}) = o_p(1)$. Nous verrons que ce résultat est encore vrai conditionnellement à l'évènement "$T(\cdot)$ possède un maximum en $\mathbf{0}$".

Théorème 9 *Pour* $U_1(\mathbf{x})$, $U_2(\mathbf{x})$ *et* $T(\mathbf{x})$ *les champs aléatoires définis précédemment,*

$$T^{-1}(\mathbf{x})\ddot{\mathbf{T}}(\mathbf{x})_{|U(\mathbf{x})} + 2\left\{v(\mathbf{x})\mathbf{\Lambda}_1(\mathbf{x}) + (1 - v(\mathbf{x}))\mathbf{\Lambda}_2(\mathbf{x})\right\} = o_p(1),$$

quand $T(\mathbf{x}) = U_1^2(\mathbf{x}) + U_2^2(\mathbf{x}) \to \infty$, *avec* $v(\mathbf{x}) = U_1^2(\mathbf{x})/T(\mathbf{x})$.

Preuve : La démonstration se fait grâce à la version multivariée de l'inégalité de Chebychev, à savoir : soit Y un vecteur aléatoire de moyenne m_Y et de matrice de covariance $\mathbf{C}_Y$. Alors, $P(||Y - m_Y|| \geq a) \leq \mathrm{tr}(\mathbf{C}_Y)/a^2,\ \forall a > 0$, où $\mathrm{tr}(\mathbf{C}_Y)$ désigne la trace de la matrice $(\mathbf{C}_Y)$.
D'après la proposition 7, $\mathrm{tr}(\mathrm{Var}(\ddot{\mathbf{T}}(\mathbf{x}) \mid U(\mathbf{x}))) = O(T(\mathbf{x}))$ quand $T(\mathbf{x}) \to \infty$. En appliquant l'inégalité de Chebychev à $\ddot{\mathbf{T}}(\mathbf{x})_{|U(\mathbf{x})}$ le résultat est immédiat. □

6.2 Champ conditionnel

Intéressons nous maintenant à la structure du champ $T(\mathbf{x})$ autour d'un maximum local. Pour cela nous allons utiliser la notion de champ conditionnel définie dans la section 2.4. D'après la relation (2.7), le champ $T(\mathbf{x})$ conditionnel à l'évènement :

$$A_0(T_0) = \text{“}T(\mathbf{x}) \text{ possède un maximum local en zéro de hauteur } T_0\text{”},$$

sera noté $T_{T_0}(\mathbf{x})$.
Afin de pouvoir étendre le théorème 9 sur la distribution asymptotique de la courbure du champ $T(\mathbf{x})$ au cas du champ conditionnel $T_{T_0}(\mathbf{x})$, il est nécessaire de connaître la distribution de $\ddot{\mathbf{T}}(\mathbf{x})$ conditionnellement à $U(\mathbf{x})$ et $\dot{U}(\mathbf{x})$.

Proposition 8 *Pour $T(\mathbf{x}) = U_1^2(\mathbf{x}) + U_2^2(\mathbf{x})$ défini dans la proposition 3,*

$$\partial^2_{kl} T(\mathbf{x})_{|U(\mathbf{x}),\dot{U}(\mathbf{x})} \sim \mathcal{N}\left(E_{\ddot{\mathbf{T}}_{kl}}, V_{\ddot{\mathbf{T}}_{kl}}\right),$$

avec :
$E_{\ddot{\mathbf{T}}_{kl}} = 2\sum_{i=1}^{2}\left(\partial_k U_i(\mathbf{x})\partial_l U_i(\mathbf{x}) - U_i^2(\mathbf{x})\mathbf{\Lambda}_i(\mathbf{x})_{[kl]} + U_i(\mathbf{x})\mathbb{E}\left[\partial^2_{kl}U_i'(\mathbf{x})\dot{U}_i(\mathbf{x})\right]\mathbf{\Lambda}_i^{-1}(\mathbf{x})\dot{U}_i(\mathbf{x})\right)$
et $V_{\ddot{\mathbf{T}}_{kl}} = 4\sum_{i=1}^{2} U_i^2(\mathbf{x})\left(\mathbf{M}_i(\mathbf{\Lambda}_i(\mathbf{x}))_{[kl]} + \mathbb{E}\left[\partial^2_{kl}U_i(\mathbf{x})\dot{U}_i(\mathbf{x})\right]\mathbf{\Lambda}_i^{-1}(\mathbf{x})\mathbb{E}'\left[\partial^2_{kl}U_i(\mathbf{x})\dot{U}_i(\mathbf{x})\right]\right).$

Cette proposition est démontrée dans l'annexe C.

Théorème 10 *Si $\mathbf{0}$ est un maximum local de hauteur T_0, alors lorsque $T_0 \to \infty$,*

$$T_0^{-1}\ddot{\mathbf{T}}_{T_0}(\mathbf{0}) + 2\left(v_0\mathbf{\Lambda}_1(\mathbf{0}) + (1-v_0)\mathbf{\Lambda}_2(\mathbf{0})\right) = o_p(1), \tag{6.6}$$

où $v_0 = U_1^2(\mathbf{0})/T_0$.

Preuve : Comme pour le théorème 9 nous utilisons la version multivariée de l'inégalité de Chebychev. D'après la proposition 8, si $\mathbf{0}$ est un maximum local,

$$\begin{aligned}\mathbb{E}\left[\partial^2_{kl}T_{T_0}(\mathbf{0})\right] &= 2\sum_{i=1}^{2}\left(-U_i^2(\mathbf{0})\mathbf{\Lambda}_i(\mathbf{0})_{[kl]} + U_i(\mathbf{0})\mathbb{E}\left[\partial^2_{kl}U_i(\mathbf{0})\dot{U}_i(\mathbf{0})\right]\mathbf{\Lambda}_i^{-1}(\mathbf{0})\dot{U}_i'(\mathbf{0})\right)\\ &= 2\sum_{i=1}^{2} -U_i^2(\mathbf{0})\mathbf{\Lambda}_i(\mathbf{0})_{[kl]} + O\left(\sqrt{T_0}\right)\end{aligned}$$

et $\mathrm{Var}\left(\partial^2_{kl}T_{T_0}(\mathbf{0})\right) = O\left(T_0\right)$.

Alors,

$$\mathbb{P}\left[\left\|\ddot{\mathbf{T}}_{T_0}(\mathbf{0}) + 2\sum_{i=1}^{2} U_i^2(\mathbf{0})\mathbf{\Lambda}_i(\mathbf{0}) + O\left(\sqrt{T_0}\right)\right\| \geq T_0\right] \leq \mathrm{tr}\left(O(T_0)\right)/T_0^2$$

d'où le résultat. □

Ce résultat est à relier avec le théorème 5 (section 4.3) établi lorsque les champs gaussiens sont identiquement distribués. En effet, si les champs $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ satisfaisaient cette condition, $\mathbf{\Lambda}_1(\mathbf{x})$ serait égal à $\mathbf{\Lambda}_2(\mathbf{x})$ et nous aurions la relation (4.4).

La proposition 6 permet d'établir le résultat suivant sur la distribution asymptotique du champ conditionnel $T_{T_0}(\mathbf{x})$.

Proposition 9 *Conditionnellement à $A_0(T_0)$, pour $\mathbf{x}$ dans un voisinage de $\mathbf{0}$:*

$$T_{T_0}\left(\mathbf{x}/\sqrt{T_0}\right) - T_0 + \mathbf{x}'\left(v_0\mathbf{\Lambda}_1(\mathbf{0}) + (1-v_0)\mathbf{\Lambda}_2(\mathbf{0})\right)\mathbf{x} = o_p(1)$$

quand $T_0 \to \infty$, avec $v_0 = U_1^2(\mathbf{0})/T_0$.

Preuve : L'équation (6.3) permet d'écrire :

$$\begin{aligned} T_{T_0}\left(\mathbf{x}/\sqrt{T_0}\right) &= U_1^2\left(\mathbf{x}/\sqrt{T_0}\right) + U_2^2\left(\mathbf{x}/\sqrt{T_0}\right) \\ &= U_1^2(\mathbf{0}) + U_2^2(\mathbf{0}) + \frac{2}{\sqrt{T_0}}\mathbf{x}'\left(U_1(\mathbf{0})\dot{U}_1(\mathbf{0}) + U_2(\mathbf{0})\dot{U}_2(\mathbf{0})\right) \\ &\quad + \frac{1}{T_0}\mathbf{x}'\left(\dot{U}_1(\mathbf{0})\dot{U}_1'(\mathbf{0}) + \dot{U}_2(\mathbf{0})\dot{U}_2'(\mathbf{0})\right)\mathbf{x} \\ &\quad + \frac{1}{T_0}\mathbf{x}'\left(U_1(\mathbf{0})\ddot{\mathbf{U}}_1(\mathbf{0}) + U_2(\mathbf{0})\ddot{\mathbf{U}}_2(\mathbf{0})\right)\mathbf{x} + \frac{1}{T_0}o_p\left(\|\mathbf{x}\|^2\right). \end{aligned}$$

D'où,

$$T_{T_0}\left(\mathbf{x}/\sqrt{T_0}\right) = T_0 + \frac{1}{\sqrt{T_0}}\mathbf{x}'\dot{T}(\mathbf{0}) + \frac{1}{2T_0}\mathbf{x}'\ddot{\mathbf{T}}(\mathbf{0})\mathbf{x} + \frac{1}{T_0}o_p\left(\|\mathbf{x}\|^2\right).$$

Conditionnellement à $A_0(T_0)$, nous avons $\dot{T}_{T_0}(\mathbf{0}) = 0$ et d'après l'équation (6.6) nous obtenons le résultat annoncé. □

Ce résultat est une extension du théorème 7 pour des champs de χ^2 standards $Y(\mathbf{x})$. Cao (1999) a pu montrer que $Y_{Y_0}\left(\mathbf{x}/\sqrt{Y_0}\right) - Y_0$ converge uniformément vers $-\mathbf{x}'\mathbf{\Lambda}\mathbf{x}$ quand $Y_0 \to \infty$, où $\mathbf{\Lambda}$ représente la matrice des moments spectraux d'ordre 2 des champs gaussiens stationnaires définissant le champ de χ^2 et Y_0 est un maximum local. En notant $\mathbf{\Lambda} = v\mathbf{\Lambda}_1 + (1-v)\mathbf{\Lambda}_2$, la proposition 9 donne un résultat similaire pour le champ χ^2 non standard $T(\mathbf{x})$.

6.3 Distribution asymptotique de la taille d'une ZCA

Il est désormais possible d'établir la loi asymptotique de la taille d'une composante connexe de l'ensemble d'excursion $\mathcal{A}_{t_\alpha}$ de $T(\mathbf{x})$ au-dessus du seuil t_α. D'après la proposition 9, conditionnellement à "$T(\cdot)$ possède un maximum local T_0 en $\mathbf{0}$", $T(\mathbf{x})$ est un paraboloïde elliptique dans le voisinage de $\mathbf{0}$. Notons $\mathcal{C}_0$ la composante connexe de $\mathcal{A}_{t_\alpha}$ contenant 0 et S_0 sa surface. S_0 représente la section horizontale du paraboloïde elliptique à la distance $T_0 - t_\alpha$ à partir du maximum, ce qui est illustré figure 6.1.

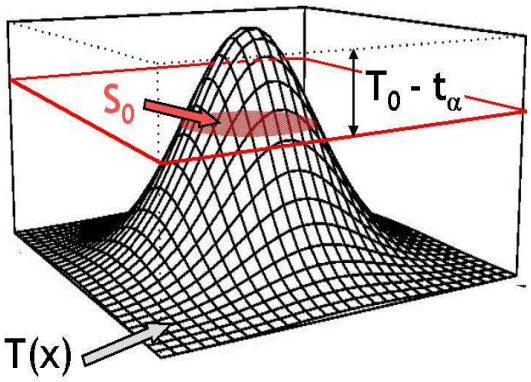


FIG. 6.1: *Représentation de la section horizontale, S_0, du paraboloïde elliptique, $T(\mathbf{x})$ dans un voisinage du maximum local $\mathbf{0}$, à la distance $T_0 - t_\alpha$ à partir du maximum.*

Proposition 10 *Conditionnellement à l'évènement "$T(\mathbf{x})$ possède un maximum en $\mathbf{0}$ de hauteur T_0 supérieur à t_α",*

$$S_0 = \frac{T_0 - t_\alpha}{T_0} \pi \det\left(\mathbf{\Lambda}(\mathbf{0})\right)^{-1/2} + o_p(1/T_0), \tag{6.7}$$

quand $t_\alpha \to \infty$, où $\mathbf{\Lambda}(\mathbf{0}) = v_0 \mathbf{\Lambda}_1(\mathbf{0}) + (1 - v_0)\mathbf{\Lambda}_2(\mathbf{0})$ et $v_0 = U_1^2(\mathbf{0})/T_0$.

Preuve : D'après la proposition 9, conditionnellement à "$T(\mathbf{0})$ est un maximum de hauteur T_0", $T(\mathbf{x})$ est représenté de la façon suivante sur la composante connexe $\mathcal{C}_0$ quand $t_\alpha \to \infty$:

$$T(\mathbf{x}) = T_0(1 - \mathbf{x}'\mathbf{\Lambda}(\mathbf{0})\mathbf{x}) + o_p(1).$$

La composante connexe $\mathcal{C}_0$ est alors : $\mathcal{C}_0 = \{\mathbf{x} : T_0(1 - \mathbf{x}'\mathbf{\Lambda}(\mathbf{0})\mathbf{x}) + o_p(1) > t_\alpha\}$, et sa mesure de Lebesgue est :

$$S_0 = \int_{T_0(1-\mathbf{x}'\mathbf{\Lambda}(\mathbf{0})\mathbf{x})+o_p(1)>t_\alpha} \mathrm{d}\mathbf{x} = \int_{\frac{T_0-t_\alpha}{T_0}+o_p(1/T_0)>\mathbf{x}'\mathbf{\Lambda}(\mathbf{0})\mathbf{x}} \mathrm{d}\mathbf{x}.$$

On effectue le changement de variable : $\mathbf{x} = \mathbf{\Lambda}(\mathbf{0})^{1/2}\mathbf{y}$, alors $\mathrm{d}\mathbf{x} = \det(\mathbf{\Lambda}(\mathbf{0}))^{-1/2}\,\mathrm{d}\mathbf{y}$ et

$$\begin{aligned} S_0 &= \det(\mathbf{\Lambda}(\mathbf{0}))^{-1/2} \int_{\frac{T_0-t_\alpha}{T_0}+o_p(1/T_0)>\mathbf{y}'\mathbf{y}} \mathrm{d}\mathbf{y} \\ &= \det(\mathbf{\Lambda}(\mathbf{0}))^{-1/2}\,\pi\frac{T_0-t_\alpha}{T_0} + o_p(1/T_0). \end{aligned}$$

□

Nous avons donc obtenu une équivalence asymptotique entre S_0 et $(T_0 - t_\alpha)/T_0$. Pour un champ de χ^2 standard, $Y(\mathbf{x})$, Aronowich et Adler (1988) ont montré que la variable aléatoire conditionnelle $Y_m - y_\alpha$, sachant que Y_m est un maximum local, est asymptotiquement distribuée comme une variable aléatoire exponentielle d'espérance 2, notée $E(2)$. Ce résultat a permis à Cao (1999) d'établir le théorème 8 sur la distribution symptotique de la taille d'une ZCA.

Le théorème suivant, qui généralise le théorème 8 au champ de χ^2 non stationnaire, est un résultat clé. Il fournit la distribution asymptotique de la taille S_{t_α} pour un seuil élevé t_α, ce qui va nous permettre d'utiliser S_{t_α} comme statistique de test pour le test global.

Théorème 11 *Conditionnellement à "$T(\mathbf{0})$ est un maximum en $\mathbf{0}$ de hauteur $T_0 > t_\alpha$",*

$$t_\alpha S_{t_\alpha} \xrightarrow{\mathcal{L}} \pi \det(\mathbf{\Lambda}(\mathbf{0}))^{-1/2}\,E(2),\ \textit{quand } t_\alpha \to \infty, \tag{6.8}$$

où $E(2)$ est une variable aléatoire exponentielle d'espérance 2 indépendante de T.

Preuve : Montrons d'abord que conditionnellement à $T_0 > t_\alpha$:

$$T_0 - t_\alpha = E(2) + o_p(1), \tag{6.9}$$

quand $t_\alpha \to \infty$.

Dans cette démonstration nous noterons $t = t_\alpha$. Nous allons utiliser un argument similaire à celui utilisé dans Aldous (1989) : les maxima locaux de $T(\mathbf{x})$ dans $\mathcal{D}$ peuvent être considérés comme un processus de points. Comme $T(\mathbf{x})$ est indéfiniment différentiable, ce processus de points est localement fini. Notons $\lambda(z)\,\mathrm{d}z$ le taux de maxima locaux de hauteur dans $[z, z + \mathrm{d}z[$, *i.e.* l'espérance du nombre de tels maxima locaux par unité de surface. Alors, comme la distribution d'un champ de χ^2 à deux degrés de liberté est la même que celle d'un champ exponentiel d'espérance deux,

$$\begin{aligned} \mathbb{P}\left[T(\mathbf{x}) > t\right] &= e^{-t/2} \\ &= \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\int_t^{+\infty} \lambda(z)\mathbf{1}_{\mathcal{C}_{\mathbf{y},z,t}}(\mathbf{x})\,\mathrm{d}z\,\mathrm{d}\mathbf{y} \\ &= \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\int_t^{+\infty} \lambda(z)\mathbf{1}_{\check{\mathcal{C}}_{\mathbf{0},z,t}\oplus(\mathbf{x}-\mathbf{y})}(\mathbf{y})\,\mathrm{d}z\,\mathrm{d}\mathbf{y} \end{aligned}$$

où $|\mathcal{D}|$ représente la surface du domaine $\mathcal{D}$, $\mathcal{C}_{\mathbf{y},z,t}$ dénote une composante connexe de l'ensemble d'excursion au dessus de t, ayant un maximum de hauteur z en $\mathbf{y}$, $\check{\mathcal{C}}_{\mathbf{0},z,t}$ est le symétrique de $\mathcal{C}_{\mathbf{0},z,t}$ par rapport à $\mathbf{0}$ et $\oplus(\mathbf{x}-\mathbf{y})$ est la translation de vecteur $\mathbf{x}-\mathbf{y}$. En ignorant les effets de bord,

$$e^{-t/2} = \frac{1}{|\mathcal{D}|}\int_t^{+\infty} \lambda(z) S(\mathbf{0}, z, t)\,\mathrm{d}z,$$

où $S(\mathbf{0}, z, t)$ est la surface de la composante connexe $\mathcal{C}_{\mathbf{0},z,t}$.
D'après la proposition 10, pour t élevé et conditionnellement à une hauteur z,

$$S(\mathbf{0}, z, t) = (z-t)\pi \det(\mathbf{\Lambda}(\mathbf{0}))^{-1/2}/z + o_p(1/z).$$

Ainsi, en négligeant $o_p(1/z)$ et après ré-arrangement,

$$K\int_t^{+\infty} \lambda(z)\frac{z-t}{z}\,\mathrm{d}z = e^{-t/2}, \tag{6.10}$$

où $K = \pi \det(\mathbf{\Lambda}(\mathbf{0}))^{-1/2}/|\mathcal{D}|$. Il est facile de montrer que :

$$\lambda(z) = \frac{1}{4K} z e^{-z/2}$$

est solution de l'équation (6.10). Cela nous mène à :

$$\mathbb{P}\left[T_0 > t+u \mid T_0 > t\right] = \frac{\int_{t+u}^{+\infty} \lambda(z)\,\mathrm{d}z}{\int_t^{+\infty} \lambda(z)\,\mathrm{d}z} = e^{-u/2}(1+O(1/t)).$$

La preuve du théorème est maintenant directe. D'après l'équation (6.9) quand $t_\alpha \to \infty$, $T_0 = t_\alpha + O_p(1)$. En insérant cette relation dans (6.7), nous avons :

$$S_{t_\alpha} = E(2)\pi \det\left(\mathbf{\Lambda}(\mathbf{0})\right)^{-1/2} / t_\alpha + O_p(1).$$

□

6.4 Test de significativité

Pour un niveau de confiance $1-\alpha$ nous avons vu dans la section 5.2 que les Zones de Changement Abrupt potentielles sont définies par l'ensemble des points $\mathbf{x}$ où la statistique $T(\mathbf{x})$ est supérieure au $(1-\alpha)$-quantile de la distribution $\chi^2(2)$, noté t_α. En d'autres termes les ZCAs potentielles correspondent à l'ensemble d'excursion $\{\mathbf{x} : T(\mathbf{x}) \geq t_\alpha\}$. Chaque

point $\mathbf{x}$ de cet ensemble correspond au rejet de l'hypothèse locale nulle $H_0(\mathbf{x})$: "$\mathbb{E}[Z(\mathbf{y})] = 0$, pour $\mathbf{y}$ dans un voisinage de $\mathbf{x}$". Les points adjacents de l'ensemble d'excursion forment des composantes connexes.

Afin de tester si le champ $Z(\cdot)$ est issu d'un champ aléatoire stationnaire (H_0) ou présentant des discontinuités dans la moyenne (H_1), nous devons définir un test global. Ce test consiste à déterminer la significativité des composantes connexes de l'ensemble d'excursion. Cela revient à tester simultanément les tests locaux $H_0(\mathbf{x}_i)$ versus $H_1(\mathbf{x}_i)$ où l'indice i désigne les points de l'ensemble d'excursion. Les tests locaux étant très dépendants, nous ne pouvons pas tester simultanément ces hypothèses. Aussi nous allons tester la significativité de chaque composante connexe indépendamment les unes des autres, ce qui revient à tester simultanément les tests locaux $H_0(\mathbf{x}_j)$ versus $H_1(\mathbf{x}_j)$ où les $\mathbf{x}_j$ appartiennent à la j-ième composante connexe.

Lorsque le seuil t_α est très grand, il y a avec une grande probabilité au plus une composante connexe sur $\mathcal{D}$ sous l'hypothèse nulle. Ainsi tester sur chaque composante connexe est équivalent à tester sur le domaine entier $\mathcal{D}$.

Pour tester la significativité d'une composante connexe $\mathcal{C}_{t_\alpha}$, nous prenons comme statistique de test sa taille, S_{t_α}. Dans la section 6.3 nous avons pu déterminer la distribution de la taille de $\mathcal{C}_{t_\alpha}$ (théorème 11). D'après l'équation (6.8), lorsque $t_\alpha \to \infty$ nous avons :

$$t_\alpha S_{t_\alpha} \stackrel{\mathcal{L}}{=} \pi \det\left(\mathbf{\Lambda}(\mathbf{0})\right)^{-1/2} E(2),$$

où $E(2) \sim \mathcal{E}xp(1/2)$, ce qui s'écrit aussi $\mathbb{P}\left[t_\alpha \det\left(\mathbf{\Lambda}\right)^{1/2} S_{t_\alpha}/\pi \geq u\right] = \exp\left(-u/2\right)$. La valeur critique associée à ce test est donc :

$$p = \exp\left(-t_\alpha \det\left(\mathbf{\Lambda}\right)^{1/2} S_{t_\alpha}/2\pi\right). \tag{6.11}$$

Nous dirons qu'une ZCA est significative lorsque la valeur critique est inférieure à un niveau η donné, par exemple $\eta = 5\%$.

Le niveau local α contrôlant le FDR global η donné est fonction de η, mais dépend aussi d'autres paramètres tels que la taille de la grille sur laquelle les calculs sont effectués, le paramètre de portée de la fonction de covariance, la densité d'échantillonnage, etc. Le niveau α est déterminé pour une série de M simulations Monte-Carlo sous H_0 correspondant à la configuration d'échantillonnage et à la fonction de covariance de la variable, et sera noté $\hat{\alpha}$. $1 - \hat{\alpha}$ est la valeur pour laquelle ηM simulations exactement présentent des ZCAs au niveau $\hat{\alpha}$. La détermination du niveau α est discutée plus en détails dans la section 7.2.

6.5 Illustration

Reprenons l'exemple de la section 5.3 où nous avions simulé dans un carré de côté un 100 points répartis aléatoirement et issus d'un champ gaussien centré-réduit de fonction de covariance exponentielle de portée $b = 0.1$ (figure 6.2*a*). Une discontinuité de 2.5 écarts-type avait été introduite en $x^1 = 0.4$ (figure 6.2*b*). Les ZCAs sont représentées en présence d'une discontinuité au niveau $1 - \hat{\alpha} = 0.9994$ et le seuil de significativité est $\eta = 5\%$ (figure 6.2*c*). Les détails de l'implémentation, en particulier l'approximation de

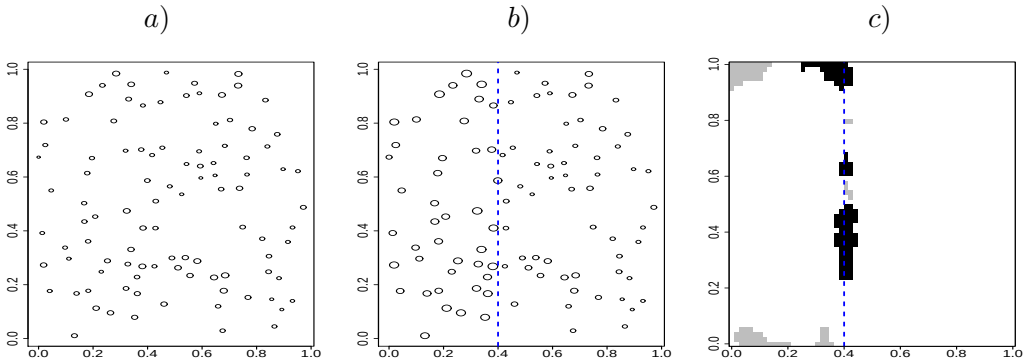


FIG. 6.2: *Echantillon de 100 points répartis aléatoirement, issus d'un processus : a) stationnaire, b) avec une discontinuité le long des pointillés. Les ZCAs significatives sont colorées en noir et les non significatives en gris : c). La grille utilisée est de dimension* 60×60.

l'équation (6.11) en pratique, sont donnés dans le chapitre 7. Dans le cas stationnaire aucune ZCA potentielle n'avait été détectée. En présence de la discontinuité, les ZCAs situées sur les bords du domaine ne sont pas significatives et les ZCAs alignées le long de la discontinuité sont pour la plupart significatives. Seules les ZCAs de petite taille ne sont pas significatives. La table 6.1 contient les surfaces et les valeurs critiques des ZCAs de la figure 6.2*c*. Dans cette table les ZCAs sont numérotées de 1 à 8. La ZCA numéro 1 correspond à la ZCA en bas à gauche de la figure 6.2*c* et la ZCA numéro 8 à la ZCA en haut à gauche. Les ZCAs 2 à 7 correspondent aux ZCAs le long de la discontinuité en partant de bas en haut. La table 6.1 illustre l'équation (6.11) : à niveau α fixé, plus la surface de la ZCA est grande, plus la valeur critique est faible.

	1	2	3	4	5	6	7	8
S	*0.0075*	*0.0028*	*0.0164*	*0.0014*	*0.0031*	*0.0006*	*0.0097*	*0.0097*
p	0.0786	0.6379	8.10^{-11}	0.1292	0.0462	0.6893	2.10^{-3}	0.6308

TAB. 6.1: *Surface S et valeur critique p des ZCAs détectées (figure 6.2c).*

La figure 6.3 représente les ZCAs obtenues sur la simulation précédente en utilisant différentes valeurs du niveau $1-\alpha$. Nous avons considéré le niveau $1-\hat{\alpha}=0.9994$ et deux autres niveaux, un plus faible et un plus élevé : $1-\alpha=0.995, 0.99995$ (figures 6.3*b*, 6.3*a* et 6.3*c* respectivement). Ces résultats montrent que la détection de ZCAs est sensible au niveau $1-\alpha$. Plus le niveau est faible, meilleure est la détection (en terme de taille et de significativité des ZCAs).

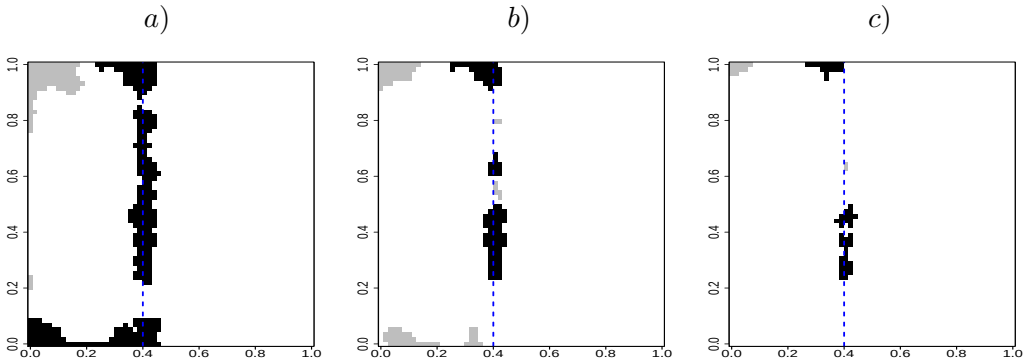


FIG. 6.3: *ZCAs (significatives en noir et non significatives en gris) obtenues pour différentes valeurs de* $1-\alpha$ *: a)* $1-\alpha=0.995$*, b)* $1-\alpha=0.9994$ *et c)* $1-\alpha=0.99995$*.*

Chapitre 7

Implémentation

Dans ce chapitre nous allons voir comment la méthode est implémentée en pratique. Les problèmes de la discrétisation du plan par une grille, de la détemination du niveau α pour cette grille et de l'estimation de la fonction de covariance seront traités successivement.

7.1 Discrétisation du plan par une grille

Comme nous l'avons vu lors des illustrations sections 5.3 et 6.5, en pratique la méthode est appliquée aux nœuds d'une grille superposée sur le domaine d'étude. En chaque pixel $[i, j]$, nous calculons dans un premier temps le gradient $W[i, j]$, sa matrice de covariance $\mathbf{\Sigma}[i, j]$ et le champ $T[i, j]$. Le gradient est multiplié par $\mathbf{\Sigma}^{-1/2}[i, j]$, ce qui nous permet d'obtenir les champs $U_1[i, j]$ et $U_2[i, j]$. Les éléments des matrices de covariance des champs $\dot{U}_k$, $\mathbf{\Lambda}_k$, $k = 1, 2$, sont calculés selon le résultat 1. Pour ces calculs nous avons besoin de la matrice de covariance entre les données, $\mathbf{C}$, et en chaque pixel du vecteur de covariance $C_Z[i, j]$ et de ses dérivées. Ces éléments sont obtenus à partir de la fonction de covariance $C_Z(\cdot)$ du champ $Z(\cdot)$. Nous définissons ensuite les Zones de Changement Abrupt potentielles au niveau $1 - \alpha$ comme l'ensemble des pixels pour lesquels la statistique $T[i, j]$ est au desssus du $(1 - \alpha)$-quantile d'une $\chi^2(2)$. Pour chaque ZCA potentielle de surface S_{t_α}, les valeurs de U_1, U_2, v, $\mathbf{\Lambda}_1$, et $\mathbf{\Lambda}_2$ sont calculées au pixel où T est maximum. La valeur critique associée au test de significativité est ensuite évaluée via :

$$p = \exp\left(-t_\alpha \det\left(\mathbf{\Lambda}\right)^{1/2} S_{t_\alpha}/2\pi\right),$$

où $\mathbf{\Lambda} = v\mathbf{\Lambda}_1 + (1 - v)\mathbf{\Lambda}_2$ et $v = U_1^2/(U_1^2 + U_2^2)$. Cette relation est une approximation numérique du théorème 11 dûe à la discrétisation du plan. Cette approximation a pu

être illustrée grâce à 1000 simulations indépendantes sous l'hypothèse nulle. Pour chaque simulation nous avons généré 100 points issus d'un champ gaussien centré-réduit, de fonction de covariance exponentielle de portée $b = 0.1$ dans un carré de côté un, et nous avons calculé la statistique $X = t_\alpha S_{t_\alpha} \det(\mathbf{\Lambda})^{1/2} / \pi$ pour chaque ZCA potentielle détectée. D'après le théorème 11, cette statistique converge vers une variable aléatoire exponentielle d'espérance 2, $E(2)$, lorsque $t_\alpha \to \infty$. La figure 7.1 montre les histogrammes de la statistique X pour une grille 30×30 et 60×60, et pour $1 - \alpha = 0.95$, 0.995, correspondant à $t_\alpha = 5.99$, 10.51. La densité de $E(2)$ a été superposée sur chaque histogramme. A côté de chaque histogramme nous avons représenté le graphe quantile-quantile, ou *qqplot* dans la suite, permettant de comparer visuellement la fonction de répartition observée en abscisse (X) à une fonction de répartition théorique en ordonnée $(E(2))$. Si la loi théorique est bien adaptée, les points du graphique sont alignés sur la droite $y = x$.

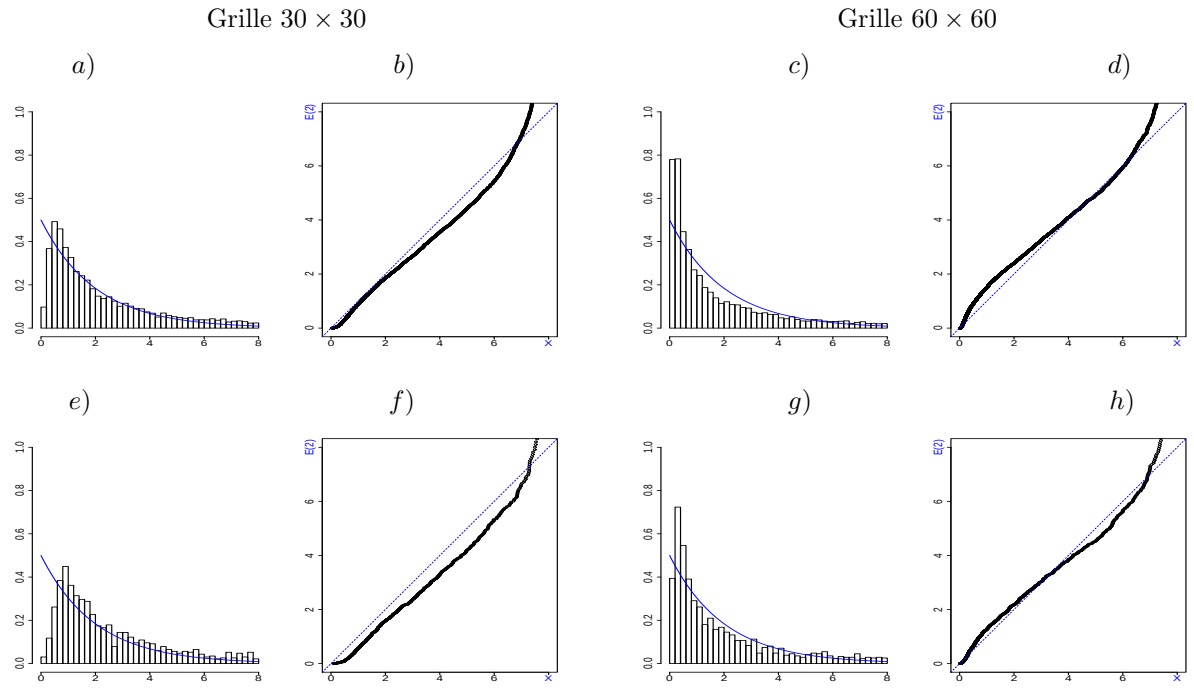


FIG. 7.1: *Histogramme et qqplot de la statistique* $X = t_\alpha S_{t_\alpha} \det(\mathbf{\Lambda})^{1/2} / \pi$ *comparée à une exponentielle d'espérance 2, dans le cas d'une grille* 30×30 *(à gauche) et* 60×60 *(à droite), et pour* $1 - \alpha = 0.95$ *(première ligne) et* $1 - \alpha = 0.995$ *(deuxième ligne).*

Un effet de discrétisation est clairement visible : la taille de l'ensemble d'excursion décroît lorsque le seuil t_α augmente. Aussi la grille doit être suffisamment fine pour détecter les petites composantes connexes lorsque $1 - \alpha$ devient grand. Cet effet est particulièrement visible sur les figures 7.1*a* et 7.1*e* sur lesquelles la combinaison d'une grille grossière et

d'un niveau élevé mène à un manque des petites valeurs de X. La figure 7.1*c* montre un excès des faibles valeurs de X et indique que la convergence en loi (6.8) ne tient pas car le seuil est trop faible.

La grille doit donc être suffisamment fine pour capturer toutes les composantes connexes de l'ensemble d'excursion et le niveau $1-\alpha$ suffisamment élevé pour que la convergence en loi tienne. Nous dirons qu'une grille est suffisamment fine, à un niveau $1-\alpha$ fixé, lorsque histogramme et densité théorique de $E(2)$ concordent bien, *i.e.* lorsque le qqplot comparant les distributions de X et de $E(2)$ s'ajuste le long de la droite $y = x$.

7.2 Détermination de α

La méthode de détection de Zones de Changement Abrupt nécessite deux niveaux de signficativité : $1-\alpha$, niveau associé au test local de détection de changement abrupt et $1-\eta$, niveau associé au test global de significativité des ZCAs potentielles. Comme la méthode de Worsley (1994) et Cao (1999) (section 4.3), il s'agit d'un problème de comparaisons multiples et la théorie développée dans le chapitre 6 permet de relier les deux niveaux de significativité. Déterminer α consiste à déterminer un niveau de confiance pour un ensemble de tests simultanés (tests multiples), *i.e.* dans notre cas le même α va être utilisé pour tester en chaque pixel $[i, j]$ de la grille d'interpolation, si la statistique $T[i, j]$ est supérieure au quantile d'ordre $1-\alpha$ d'une χ^2 à deux degrés de liberté, t_α.

Dans cette section, nous détaillons quelques méthodes existantes de contrôle du taux d'erreur de type I ou d'évaluation du niveau α et proposons différentes procédures pour déterminer α.

Méthodes existantes

Benjamini et Hochberg, 1995

Le *False Discovery Rate* (FDR) est une nouvelle approche [Benjamini et Hochberg (1995)] utilisée pour les problèmes de comparaisons multiples. Au lieu de contrôler le taux de faux positifs (ce qui est par exemple le cas de la procédure de Bonferroni), le FDR contrôle la proportion attendue de faux positifs. Le FDR est déterminé à partir des valeurs critiques observées et par conséquent s'adapte aux données.

Considérons le problème de tester simultanément N hypothèses nulles $H_{0,1}, \ldots, H_{0,N}$ et notons R le nombre d'hypothèses nulles rejetées. D'un point de vue fréquentiste la situation peut se résumer comme dans la table 7.1 [Benjamini et Hochberg (1995)]. Les N

Nombre d'hypothèses nulles	non rejets	rejets	
vraies	U	V	N_0
fausses	T	S	N_1
	$N-R$	R	N

TAB. 7.1: *Nombre d'erreurs commises lors du test de N hypothèses.*

hypothèses sont supposées connues. Les nombres N_0 et $N_1 = N - N_0$ d'hypothèses nulles vraies et fausses sont des paramètres inconnus. R est une variable aléatoire observable et S, T, U et V sont des variables aléatoires non observables. Si chaque hypothèse nulle est testée individuellement au niveau η, alors $R = R(\eta)$ croît avec η. Tester chaque hypothèse nulle individuellement au niveau η garantit que $\mathbb{E}[V/N] \leq \eta$, car $V \leq R$. De plus lorsque chaque hypothèse est testée au niveau η/N, $V \sim \mathcal{B}in(N_0, \eta/N)$ et :

$$\begin{aligned} \mathbb{P}[V \geq 1] &= 1 - \mathbb{P}[V < 1] = \mathbb{P}[V = 0] = 1 - (1 - \eta/N) \\ &\approx 1 - (1 - \eta N_0/N) = \eta N_0/N \leq \eta \end{aligned}$$

Idéalement le nombre V de faux positifs et le nombre T de faux négatifs doivent être minimisés. Le taux de faux positifs introduit est l'espérance de la proportion d'erreurs de type I parmi les hypothèses rejetées :

$$\text{FDR} = \mathbb{E}\left[\frac{V}{R}\right], \text{ si } R > 0, \ 0 \text{ si } R = 0.$$

Une procédure de tests multiples contrôle un taux d'erreur de type I au niveau η, si ce taux est inférieur ou égal à η lorsque la procédure est appliquée pour produire une liste de R hypothèses rejetées. Supposons qu'il y ait une statistique de test associée à chaque test d'hypothèses $H_{0,i}$ versus $H_{1,i}$, $i = 1, \ldots, N$. La procédure de contrôle du FDR proposée dans Benjamini et Hochberg (1995) suppose l'indépendance des N statistiques de test. Dans leur procédure, ils calculent les valeurs critiques, p_i, basées sur la statistique du test $H_{0,i}$ versus $H_{1,i}$, $i = 1, \ldots, N$, et les statistiques d'ordre $p_{(1)} \leq \cdots \leq p_{(N)}$ des valeurs critiques $(p_1, \ldots, p_N)$, et notent $H_{0,(i)}$ l'hypothèse associée à $p_{(i)}$. Pour $\eta \in (0, 1)$ un niveau donné que le FDR ne doit pas dépasser, ils définissent $\hat{k}$, par :

$$\hat{k} = \max\left\{i : p_{(i)} \leq \frac{i}{N}\eta\right\}.$$

Si un tel $\hat{k} \geq 1$ existe, alors toutes les hypothèses $H_{0,(i)}, i = 1, \ldots, \hat{k}$ sont rejetées. Si $\hat{k}$ n'existe pas, alors aucune hypothèse $H_{0,i}$, $i = 1, \ldots, N$ n'est rejetée.

Pour les $0 \leq N_0 \leq N$ valeurs critiques indépendantes correspondant aux hypothèses nulles vraies, et pour toutes les N_1 valeurs que peuvent prendre les valeurs critiques correspondant aux hypothèses nulles fausses, d'après Benjamini et Hochberg (1995) la procédure de tests multiples décrite précédemment satisfait l'inégalité :

$$\mathbb{E}\left[\frac{V}{R}\middle| P_{N_0+1} = p_1, \ldots, P_N = p_{N_1}\right] \leq \frac{N_0}{N}\eta. \tag{7.1}$$

Si nous supposons que N_1 des hypothèses sont fausses, alors quelle que soit la distribution jointe des valeurs critiques associées à ces hypothèses, l'intégration de l'inégalité (7.1) permet d'obtenir :

$$\mathbb{E}\left[\frac{V}{R}\right] \leq \frac{N_0}{N}\eta \leq \eta,$$

et le FDR est contrôlé.

Shen, Huang et Cressie, 2002

Le point faible de la procédure FDR de Benjamini et Hochberg (1995) est que bien que $\hat{k}$ dépend de toutes les valeurs critiques à travers leur statistique d'ordre, l'ampleur des plus grandes valeurs critiques n'est pas spécifiquement utilisée. L'approche proposée par Shen *et al.* (2002), appelée le *Enhanced False Discovery Rate* (EFDR), prend en compte l'information contenue dans toutes les valeurs critiques et s'avère plus puissante que le FDR standard. La puissance de la procédure EDFR ne décroît pas quand N augmente, ce qui est le cas de la procédure FDR. Cette procédure utilise le principe du FDR mais l'améliore en réduisant le nombre d'hypothèses à tester.

Shen *et al.* (2002) cherchent à detecter un signal f dans un domaine spatial, par un test d'hypothèse nulle $f = 0$ et d'alternative la non nullité par endroits. Cette méthode suppose que les données sont modélisées par un processus gaussien $Z(x) = f(x) + \epsilon(x)$, où $\epsilon(\cdot)$ est un processus stationnaire gaussien centré. Elle utilise la décomposition en ondelettes des données observées sur une grille et teste simultanément l'hypothèse nulle de nullité des cœfficients des ondelettes, versus l'alternative d'existence de cœfficients non nuls. Leur procédure EFDR de contrôle de la probabilité d'erreur de type I se décompose en trois étapes.

Etape 1 : Pour $1 \leq L \leq N$, les L tests d'hypothèses correspondant aux L plus grandes statistiques d'ordre des cœfficients des ondelettes sont sélectionnés parmi les N possibles. Les hypothèses nulles restantes sont considérées comme acceptées.

Etape 2 : La procédure FDR de Benjamini et Hochberg (1995) est ensuite appliquée aux cœfficients sélectionnés.

Etape 3 : Plusieurs valeurs de L sont choisies entre 1 et N et un critère basé sur les degrés de liberté généralisés [Ye (1998)] est évalué en utilisant les étapes 1 et 2. La valeur optimale de L est celle qui minimise ce critère.

La procédure EFDR se réduit à la procédure FDR lorsque $L = N$. Shen *et al.* (2002) ont pu montrer qu'en considérant le test de H_0 versus H_1 basé sur N tests d'hypothèses indépendants $H_{0,i}$ versus $H_{1,i}$, $i = 1, \ldots, N$, toutes les procédures basées sur les étapes 1 et 2 contrôlent le FDR à ne pas être plus grand que η. Ce résultat stipule que la procédure de contrôle du FDR, précédée de l'élimination de certaines hypothèses individuelles, a un FDR qui peut être borné par un niveau préspécifié η, sans se soucier de savoir si ces hypothèses devaient être éliminées.

Ce que nous proposons

Nous cherchons à déterminer le niveau α en fonction du FDR global η, prenant en compte la fonction de covariance de $Z(\cdot)$, le schéma d'échantillonnage et la discrétisation. Les tests locaux de détection de changement abrupt sont non indépendants, aussi nous ne pouvons utiliser des méthodes de contrôle du FDR du type celles de Benjamini et Hochberg (1995) ou de Shen *et al.* (2002).

Utilisation de la portée intégrale

Une première solution consiste à s'inspirer des méthodes de Chaudhuri et Marron (1999) et de Godtliebsen *et al.* (2001, 2002). Comme l'avons vu section 1.2, ils ont déterminé le niveau α, à niveau η fixé, en calculant préalablement le nombre de cas d'intervalles de confiance indépendants, $m(h)$ dans Chaudhuri et Marron (1999), et le nombre de groupes indépendants, l dans Godtliebsen *et al.* (2001, 2002). Les équations (1.3) et (1.5) fournissent le niveau α en fonction de η.

Afin de déterminer l'analogue du nombre de blocs indépendants nous allons utiliser la notion de portée intégrale (voir par exemple Lantuéjoul (1991)) pour une définition précise de cette notion). Dans la section 2.2, nous avons vu que pour $Z(\cdot)$ un champ aléatoire stationnaire, de moyenne m, de variance σ^2 et de fonction de corrélation $\rho(\mathbf{h})$,

$$Z(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} Z(\mathbf{x})\, \mathrm{d}\mathbf{x},$$

où $|\mathcal{D}|$ représente le volume du domaine $\mathcal{D} \subset \mathbb{R}^d$, définit un estimateur sans biais de m.

La variance de cet estimateur est :

$$\mathrm{Var}\,(Z(\mathcal{D})) = \frac{\sigma^2}{|\mathcal{D}|^2}\int_{\mathcal{D}}\int_{\mathcal{D}} \rho(\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}.$$

La propriété d'ergodicité (2.4) n'indique pas la taille que doit atteindre le domaine $\mathcal{D}$ pour que $\mathrm{Var}\,(Z(\mathcal{D}))$ devienne négligeable. Cela est précisé à l'aide de la notion de portée intégrale.

Définition 7 *La portée intégrale est la quantité définie par :*

$$A = \lim_{\mathcal{D}\to\infty} |\mathcal{D}| \frac{Var\,(Z(\mathcal{D}))}{\sigma^2}.$$

La portée intégrale représente l'échelle du phénomène, alors que $\mathcal{D}$ représente l'échelle de l'observation. Elle possède les propriétés suivantes.

Propriétés :

(i) A n'existe pas toujours, lorsqu'elle existe, elle est non négative et à la dimension d'un d-volume.

(ii) Si $A < +\infty$, lorsque $\mathcal{D}$ est grand, $\mathrm{Var}\,(Z(\mathcal{D})) \approx \sigma^2 A/|\mathcal{D}|$.

(iii) Si $A \neq 0$, il existe N tel que $|\mathcal{D}|/A \approx N$ et $\mathrm{Var}\,(Z(\mathcal{D})) \approx \sigma^2/N$ représente la variance de la moyenne de N observations indépendantes de même taille A.

Lorsque la portée intégrale est non nulle et que la fonction de corrélation est intégrable, nous avons :

$$A = \int_{\mathbb{R}^d} \rho(\mathbf{h})\,\mathrm{d}\mathbf{h} = \int_{\mathbb{R}^d} \frac{C_Z(\mathbf{h})}{C_Z(\mathbf{0})}\,\mathrm{d}\mathbf{h}.$$

Ainsi, pour un modèle de covariance exponentiel $C_Z(\mathbf{h}) = C_Z(\mathbf{0})\exp(-\mathbf{h}/b)$, $\mathbf{h} \geq \mathbf{0}$ défini dans $\mathbb{R}^2$, la portée intégrale a la forme simple suivante

$$A = 2\pi b^2. \tag{7.2}$$

D'après l'équation (1.5), nous pouvons définir le niveau α suivant :

$$\alpha_G = 1 - (1-\eta)^{1/N}, \tag{7.3}$$

où N est défini dans la propriété (iii). Le niveau α, associé à l'équation (1.3), ne peut être réellement utilisé car il correspond à un test bilatéral ($\eta = \mathbb{P}\,[\cup_i |X_i| > q_\alpha]$), alors que nous avons un test unilatéral ($\eta = \mathbb{P}\,[\cup_i T(\mathbf{x}_i) > t_\alpha]$). Le test de Chaudhuri et Marron (1999) dans le cas unilatéral entraîne un niveau α de la forme α_G associé au test unilatéral de Godtliebsen *et al.* (2001, 2002).

Calcul de α en chaque pixel de la grille d'interpolation

D'après l'équation (6.11), dire que l'exponentielle, $\exp\left(-t_\alpha \det\left(\mathbf{\Lambda}[i,j]\right)^{1/2} S_{t_\alpha}/2\pi\right)$, est égale au niveau η, revient à considérer que le seuil $t_\alpha = \chi^2_{1-\alpha}(2)$ vaut :

$$t_\alpha = -2\pi \log(\eta) \det\left(\mathbf{\Lambda}[i,j]\right)^{-1/2} / S_{t_\alpha},$$

d'où pour $F_{\chi^2(2)}$ la fonction de répartition d'une $\chi^2(2)$:

$$1 - \alpha_{[i,j]} = F_{\chi^2(2)}\left(-2\pi \log(\eta) \det\left(\mathbf{\Lambda}[i,j]\right)^{-1/2} / S_{t_\alpha}\right). \tag{7.4}$$

Dans Worsley (1994) et Cao (1999), $\mathbf{\Lambda}$ est constant sur tout le domaine, aussi il leur est possible d'obtenir le niveau α correspondant au niveau η via l'équation (7.4). Dans notre cas, $\mathbf{\Lambda}$ peut prendre une valeur différente en chaque site du domaine, et nous n'avons par conséquent pas une valeur unique de α.

Nous pouvons considérer comme niveau α, soit un niveau différent en chaque pixel de la grille d'interpolation, $\alpha_{[i,j]}$, soit le α médian, α_{med}, ou moyen, α_{moy}, des $\alpha_{[i,j]}$. Le choix de α est discuté un peu plus loin.

Simulations Monte-Carlo

Une autre solution consiste à déterminer le niveau α par des simulations de type Monte-Carlo. Pour M simulations d'un champ gaussien sous l'hypothèse nulle, conditionnel à la même fonction de covariance et au même schéma d'échantillonnage que la variable, le niveau correct $\hat{\alpha}$ est la plus grande valeur de α telle que le nombre de simulations ayant des ZCAs significatives au dessus du seuil t_α soit inférieur au produit entre le niveau global η et le nombre de simulations :

$$\hat{\alpha} = \sup\{\alpha : M_\alpha \leq \eta M\},$$

où M_α est le nombre de simulations (parmi M) ayant des ZCAs significatives au niveau α.

Choix du niveau adéquat

Nous avons calculé le niveau $1 - \alpha_{[i,j]}$ (équation 7.4) en chaque pixel d'une grille 30×30 et d'une grille 60×60, pour 100 simulations sous l'hypothèse nulle d'absence de discontinuité, en utilisant un modèle de covariance exponentiel de portée $b = 0.1$. Les valeurs obtenues de $1 - \alpha_{[i,j]}$ sont résumées dans la table 7.2. Il apparaît que les valeurs

de $\alpha_{[i,j]}$ sont très petites et qu'elles dépendent de la discrétisation, ce qui va dans le sens des résultats de la section 7.1 : plus la discrétisation est fine, plus le niveau $1-\alpha_{[i,j]}$ est grand (tend vers 1).

	Min.	1^{er} quartile	Médiane	Moyenne	$3^{\text{ème}}$ quartile	Max.
grille 30×30	0.03631	1	1	0.9988	1	1
grille 60×60	0.08122	1	1	0.9999	1	1

TAB. 7.2: *Valeurs de $1-\alpha_{[i,j]}$ obtenues en chaque pixel de la grille d'interpolation pour un ensemble de 100 simulations.*

La figure 7.2 représente le niveau $1-\alpha_G$ (équation 7.3) en fonction du paramètre de portée d'un modèle de covariance exponentiel, pour un domaine de surface unité. Cette figure montre que plus la portée est grande, moins il y a de groupes indépendants dans le domaine, et plus le niveau $1-\alpha_G$ est faible. Considérant un modèle de covariance exponentiel de portée $b = 0.1$ et un niveau global $\eta = 0.05$, le niveau $1-\alpha_G$ est égal à $(1-\eta)^{1/(1/(2\pi b^2))} = 0.9968$.

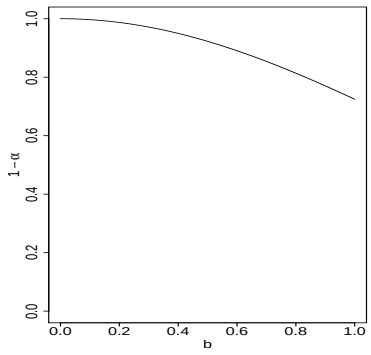


FIG. 7.2: *Evolution du niveau $1-\alpha_G$ en fonction de la portée b d'un modèle de covariance exponentiel.*

Pour chacune des 100 simulations, nous avons déterminé le niveau $1-\hat{\alpha}$ en fonction de la discrétisation. Le résumé statistique de ces valeurs est donné dans la table 7.3.

	Min.	1^{er} quartile	Médiane	Moyenne	$3^{\text{ème}}$ quartile	Max.
grille 30×30	0.9973	0.9989	0.9991	0.9990	0.9994	0.9996
grille 60×60	0.9987	0.9992	0.9994	0.9993	0.9995	0.9998

TAB. 7.3: *Valeurs de $1-\hat{\alpha}$ pour une discrétisation 30×30 et 60×60 pour un ensemble de 100 simulations utilisant un modèle de covariance exponentiel de portée $b = 0.1$.*

Ces valeurs montrent que le niveau $1-\alpha$ est plus élevé lorsque la discrétisation est plus fine, de manière à détecter les petites ZCAs.

Nous avons calculé pour l'ensemble des 100 simulations sous l'hypothèse nulle, la statistique $X = t_\alpha S_{t_\alpha} \det(\mathbf{\Lambda})^{1/2} / \pi$. La figure 7.3 montre les histogrammes de la statistique X pour une grille 30×30 (à gauche) et 60×60 (à droite). La densité de $E(2)$ a été superposée sur les histogrammes. A côté de chaque histogramme nous avons représenté le qqplot permettant de comparer les distributions de X (en abscisse) et de $E(2)$ (en ordonnée). La première ligne fournit les résultats obtenus avec $1-\alpha_{[i,j]}$. Il apparaît qu'avec ce niveau les petites valeurs de X sont largement sur-estimées et les grandes valeurs sous-estimées. Il est donc exclus d'utiliser $1-\alpha_{[i,j]}$ comme niveau local. Considérer le niveau α_{moy} (deuxième ligne), moyenne des $\alpha_{[i,j]}$ sur le domaine, améliore les résultats, mais les petites ZCAs restent sur-estimées (figure 7.3*g*). La troisième ligne correspond au cas $1-\alpha_G$. Le niveau α_G prend en compte le schéma d'échantillonnage et la fonction de covariance de la variable considérée au travers de la portée, b, du modèle de covariance. Cependant, il y a un inconvénient à considérer α_G : il est indépendant de la grille utilisée pour appliquer la méthode. Or nous avons vu dans la section 7.1, qu'il y a un lien entre discrétisation et niveau α à considérer. Le fait que α_G soit indépendant de la discrétisation peut avoir des conséquences sur la détection des ZCAs : pour une discrétisation grossière les petites ZCAs seront sous-estimées et pour une discrétisation fine elles seront sur-estimées. Ce phénomène est légèrement visible sur les figures 7.3*i* et 7.3*k*. La dernière ligne fournit les résultats pour $1-\hat{\alpha}$ obtenu par des simulations de type Monte-Carlo. Les figures 7.3*n* et 7.3*p* montrent, par l'alignement des points sur la droite $y = x$ que X a la même distribution que $E(2)$. Considérer $1-\hat{\alpha}$ comme niveau local est un bon choix, et fournit un bon compromis entre un niveau local $1-\alpha$ assez élevé pour que les résultats mathématiques soient valables, mais pas trop élevé pour pouvoir détecter et visualiser des ZCAs.

La figure 7.4 représente l'application de la méthode à l'exemple des sections 5.3 et 6.5 pour différentes valeurs du niveau α : $1-\alpha \in \left\{1-\alpha_{[i,j]}, 1-\alpha_{moy}, 1-\alpha_G, 1-\hat{\alpha}\right\}$ et pour différentes discrétisations : grille 30×30 (première ligne) et 60×60 (deuxième ligne). Cette figure illustre les résultats précédents : le niveau $1-\alpha_{[i,j]}$ est trop élevé en terme de détection de ZCAs (figures 7.4*a* et 7.4*e*) et considérer sa moyenne corrige un peu le problème (figures 7.4*b* et 7.4*f*). Les niveaux $1-\alpha_G$ (figures 7.4*c* et 7.4*g*) et $1-\hat{\alpha}$ (figures 7.4*d* et 7.4*h*) fournissent des résultats assez similaires. Les ZCAs sont un peu plus grandes dans le cas de $1-\alpha_G$ du fait que le niveau soit légérement inférieur à $1-\hat{\alpha}$.

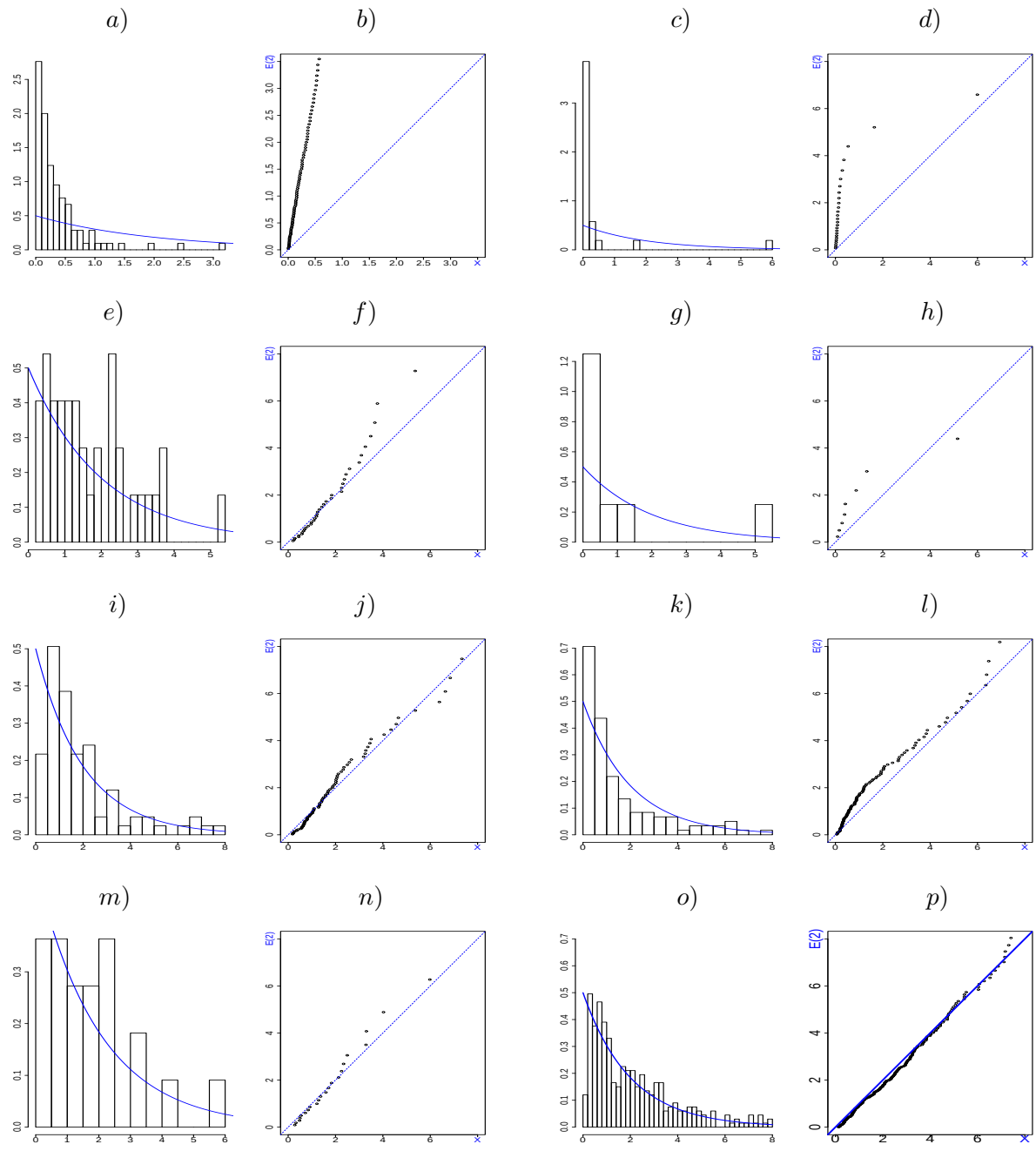


FIG. 7.3: *Histogramme et qqplot de la statistique* $X = t_\alpha S_{t_\alpha} \det(\mathbf{\Lambda})^{1/2}/\pi$ *comparée à une exponentielle d'espérance 2,* $E(2)$*, pour un ensemble de 100 simulations, dans le cas d'une grille* 30×30 *(à gauche) et* 60×60 *(à droite), et pour* $1 - \alpha_{[i,j]}$ *(première ligne),* $1 - \alpha_{moy}$ *(deuxième ligne),* $1 - \alpha_G$ *(troisième ligne) et* $1 - \hat{\alpha}$ *(dernière ligne).*

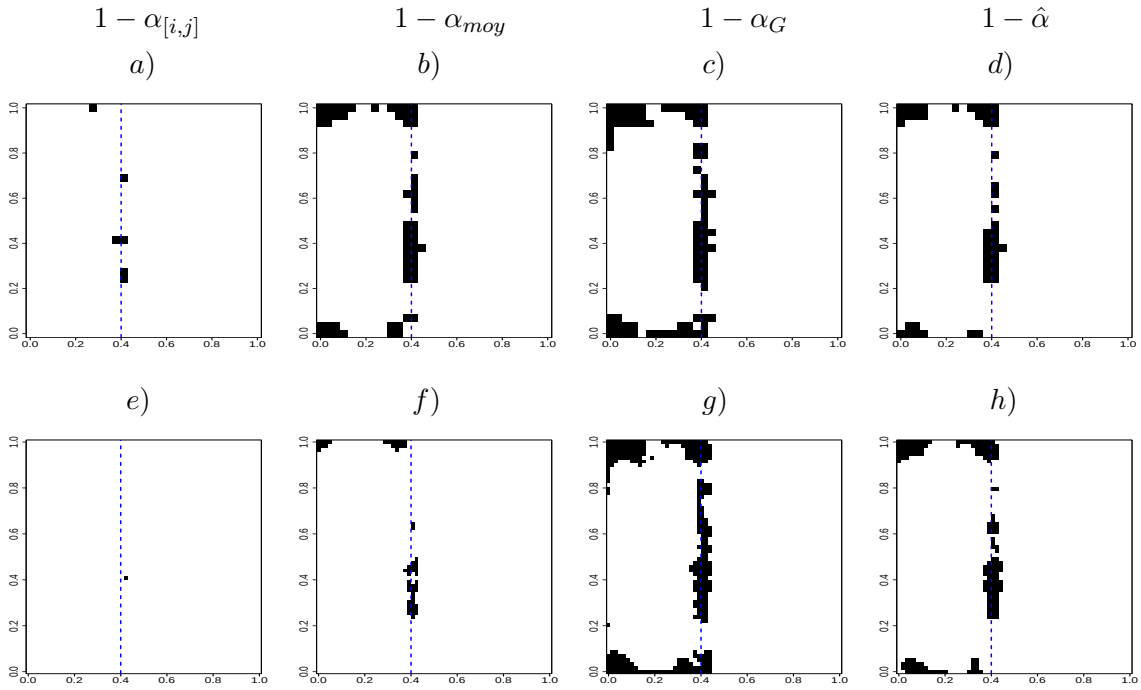


FIG. 7.4: *ZCAs potentielles obtenues pour la simulation utilisée pour illustrer la méthode, en fonction du niveau* $1-\alpha$ *:* $1-\alpha_{[i,j]}$, $1-\alpha_{moy}$, $1-\alpha_G$ *et* $1-\hat{\alpha}$ *(de gauche à droite), et de la discrétisation :* 30×30 *(première ligne) et* 60×60 *(deuxième ligne).*

Les effets de bord qui apparaissent sur ces figures sont dûs au fait que le champ est supposé de moyenne nulle à l'extérieur du rectangle $[0, 0.4]\times[0, 1]$. Nous verrons section 8.1 que cet effet dépend de l'interpolateur spatial choisi. Les figures 7.4*c* et 7.4*g* et les figures 7.4*d* et 7.4*h* montrent que les niveaux $1-\alpha_G$ et $1-\hat{\alpha}$ sont insensibles à la discrétisation.

D'après ce que nous venons de voir, parmi les différents niveaux $1-\alpha$ proposés, seuls deux d'entre eux peuvent être retenus. Le meilleur choix consiste à considérer le niveau $1-\hat{\alpha}$ obtenu par des simulations de type Monte-Carlo. En effet, ce niveau répond aux exigences requises : il est fonction du niveau global η, prend en compte la fonction de covariance de $Z(\cdot)$, le schéma d'échantillonnage et la discrétisation. Le second choix possible est le niveau $1-\alpha_G$ déterminé à l'aide de la portée intégrale. Ce niveau satisfait presque toutes les exigences sauf qu'il ne dépend pas de la discrétisation. Cela ne pose pas de problème important pour les discrétisations que nous avons considéré et peut par conséquent être utilisé en première approximation car il nécessite bien moins de calculs que le niveau $1-\hat{\alpha}$.

7.3 Estimation de la covariance sous l'alternative

Dans tout ce qui précède nous avons supposé que la fonction de covariance était connue, mais ce n'est pas le cas en pratique, aussi elle doit être estimée. Elle est estimée en utilisant le variogramme $\gamma(\mathbf{h}) = \mathbb{E}\left[(Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h}))^2\right]/2$. Pour un champ stationnaire d'ordre 2, ce qui par hypothèse est le cas du champ $Z(\cdot)$ que nous considérons, variogramme et covariance sont reliés par : $\gamma(\mathbf{h}) = C_Z(\mathbf{0}) - C_Z(\mathbf{h})$ (section 2.1). L'estimation du variogramme correspond à un ajustement au sens des moindres carrés du variogramme expérimental :

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \left(Z(\mathbf{x}_i) - Z(\mathbf{x}_j)\right)^2,$$

où la somme est prise sur $N(\mathbf{h}) = \{(i,j) : \mathbf{x}_i - \mathbf{x}_j \simeq \mathbf{h}\}$ et $|N(\mathbf{h})|$ est le nombre d'éléments distincts de $N(\mathbf{h})$. Notons σ^2 la variance du champ aléatoire : $\sigma^2 = \mathrm{Var}\left(Z(\cdot)\right) = C_Z(\mathbf{0})$.

Sous l'hypothèse alternative, la présence de discontinuités de l'espérance du champ implique quelques difficultés dans l'estimation du variogramme. Les différences $Z(\mathbf{x}_i) - Z(\mathbf{x}_j)$ interviennent dans le calcul du variogramme expérimental. Si les points $\mathbf{x}_i$ et $\mathbf{x}_j$ sont de part et d'autre de la discontinuité, cela entraîne un biais dans l'estimation des paramètres du variogramme. En particulier la variance estimée $\hat{\sigma}^2$ est surestimée. Le champ $T(\cdot)$ étant inversément proportionnel à σ^2, cela mène à de petites ZCAs potentielles et par conséquent à de grandes valeurs critiques et donc à une perte de puissance de la méthode.

Afin de résoudre cette difficulté, nous proposons une procédure itérative dans laquelle les paramètres de la fonction de covariance, le niveau α et les ZCAs sont estimés à chaque itération. La procédure est la suivante :

(i) *Première itération :* Les paramètres de la fonction de covariance sont estimés globalement, *i.e.* en utilisant l'ensemble des points de données. Le niveau α correspondant et un premier ensemble de ZCAs sont alors estimés à partir de ces paramètres.

(ii) *Itérations suivantes :* Les paramètres de la fonction de covariance sont ré-estimés, mais les couples $\{Z(\mathbf{x}_i), Z(\mathbf{x}_j)\}$ dont le segment $[\mathbf{x}_i, \mathbf{x}_j]$ intersecte une ZCA potentielle au niveau α sont éliminés de la procédure d'estimation du variogramme. Si la portée de la fonction de covariance change, le niveau α doit être recalculé. Les ZCAs sont alors ré-estimées à l'aide des nouveaux paramètres. La procédure est ré-itérée jusqu'à la convergence.

(iii) *Convergence :* La convergence est atteinte à la k-ième itération lorsque l'ensemble de ZCAs détectées est identique entre deux itérations consécutives : $\mathrm{ZCAs}^{(k)} = \mathrm{ZCAs}^{(k+1)}$.

La figure 7.5 illustre les configurations possibles de couples $\{Z(\mathbf{x}_i), Z(\mathbf{x}_j)\}$ qui peuvent être éliminés de la procédure d'estimation du variogramme. C'est la cas si :

- les points $\mathbf{x}_i$ et $\mathbf{x}_j$ sont disposés de part et d'autre de la ZCA de sorte que le segment $[\mathbf{x}_i, \mathbf{x}_j]$ intersecte la ZCA (figure 7.5*a*),
- un des points appartient à la ZCA (figure 7.5*b*),
- les deux points appartiennent à la ZCA (figure 7.5*c*).

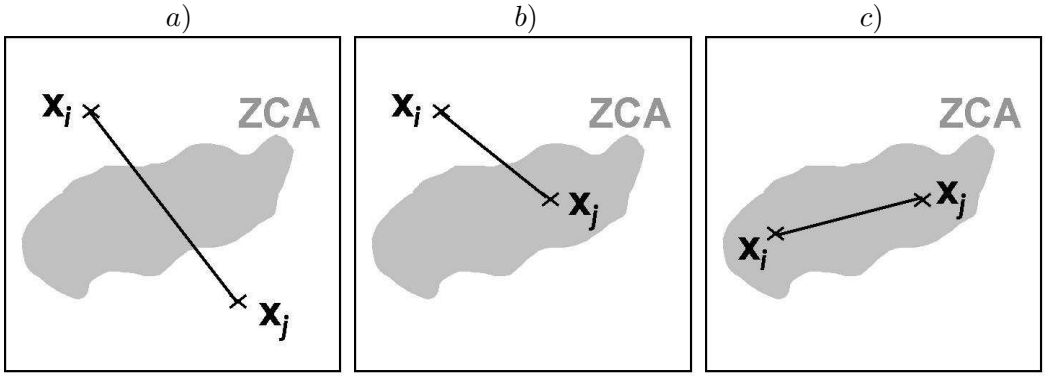


FIG. 7.5: *Cconfigurations de points $\mathbf{x}_i$, $\mathbf{x}_j$ pour lesquels les couples $\{Z(\mathbf{x}_i),\ Z(\mathbf{x}_j)\}$ ne sont pas pris en compte dans le calcul du variogramme expérimental.*

Nous n'avons pas de preuve formelle de la convergence de la procédure, mais elle est très souvent observée en moins de cinq itérations. Cette procédure fait décroître la variance estimée et accroître les ZCAs détectées, essentiellement en terme de surface. Sous l'hypothèse nulle, les ZCAs significatives sont détectées avec une probabilité η. Lorsque des ZCAs sont détectées sous cette hypothèse, il n'y a que très peu de points qui sont éliminés de la procédure d'estimation de la fonction de covariance. Sur toutes les simulations testées sous l'hypothèse nulle, les ZCAs étaient inchangées dès la deuxième itération.

Illustration

Nous avons appliqué la procédure itérative à l'exemple utilisé dans les sections 5.3 et 6.5. Nous avions généré dans un carré de côté un, 100 points issus d'un champ gaussien centré-réduit, de fonction de covariance exponentielle de portée $b = 0.1$. Une discontinuité de 2.5 écarts-type avait été introduite en $x = 0.4$ (figure 7.6*a*). Les ZCAs détectées sont représentées sur la figure 7.6*d*. Les paramètres de la fonction de covariance estimés à partir de tous les points d'échantillonnage sont les suivants : $b = 0.275$ et $\sigma^2 = 2.99$. Le variogramme correspondant est représenté figure 7.6*b*. Le niveau α correspondant est, via

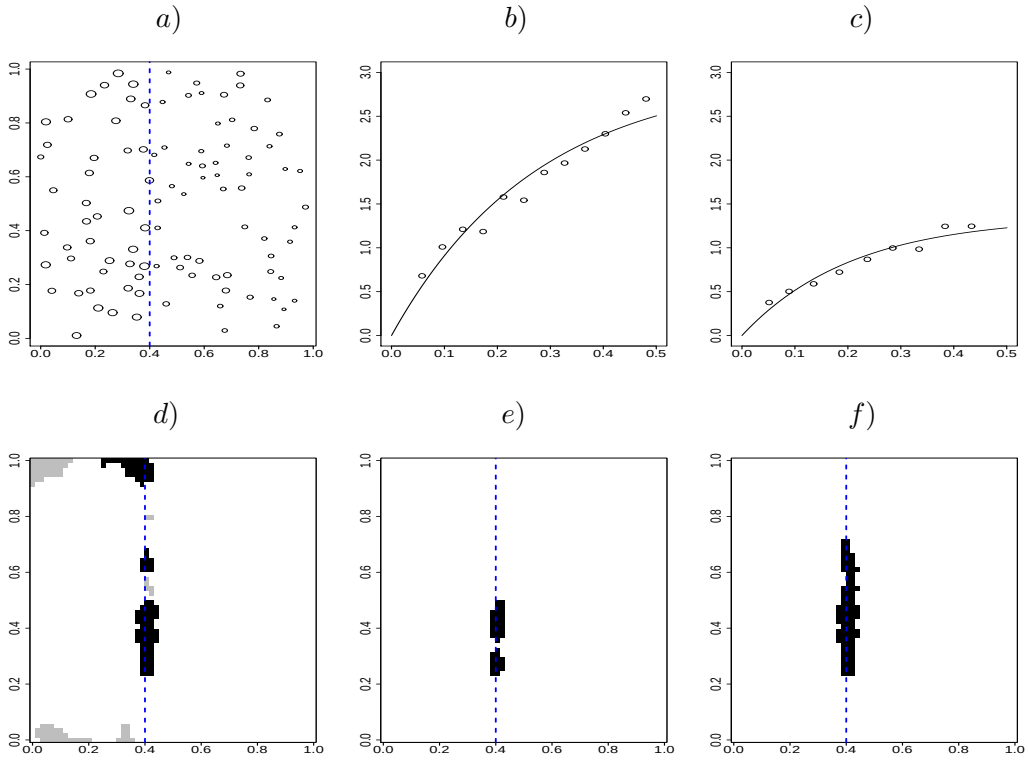


FIG. 7.6: *Application de la procédure itérative sur l'exemple utilisé pour illustrer la méthode. a) Echantillon de points considéré avec une discontinuité* $a = 2.5$ *introduite en* $x^1 = 0.4$*. Variogrammes obtenus à la première itération, b), et à la convergence, c). ZCAs détectées (significatives en noir, non significatives en gris) en considérant la covariance connue, d), à la première itération, e) et à la convergence, f).*

les simulations de type Monte-Carlo, $1 - \hat{\alpha} = 0.9984$, et le premier ensemble de ZCAs est illustré figure 7.6*e*. La convergence a été atteinte à la cinquième itération. Les paramètres de la fonction de covariance à la convergence sont : $b = 0.208$ et $\sigma^2 = 1.35$ (figure 7.6*c*). Au niveau $1 - \hat{\alpha} = 0.9994$ correspondant, les ZCAs sont illustrées figure 7.6*f*. Les figures 7.6*b* et 7.6*c* montrent que la variance estimée décroît fortement entre la première et la dernière itération, ce qui illustre l'erreur commise dans l'estimation de la variance. Les figures 7.6*e* et 7.6*f* montrent l'amélioration en terme de détection de ZCAs qu'apporte la procédure itérative. La figure 7.6 montre que la procédure itérative a éliminé les effets de bord. Cela peut être dû au fait que dans la procédure itérative nous estimons les paramètres de la fonction de covariance, alors que dans la procédure de simulations nous les supposons connus.

Nous avons également testé si la procédure itérative a tendance à accroître les ZCAs dûes à des valeurs aberrantes. Pour cela nous avons considéré l'échantillon de points de

l'exemple précédent sous l'hypothèse nulle et nous avons ajouté 4 écarts-type à la valeur d'un des points du centre du domaine. L'échantillon de points est représenté sur la figure 7.7*a*. Sur cette figure chaque cercle est proportionnel à la valeur du point et le point marqué comme valeur abérante est représenté par un cercle plein. La procédure itérative

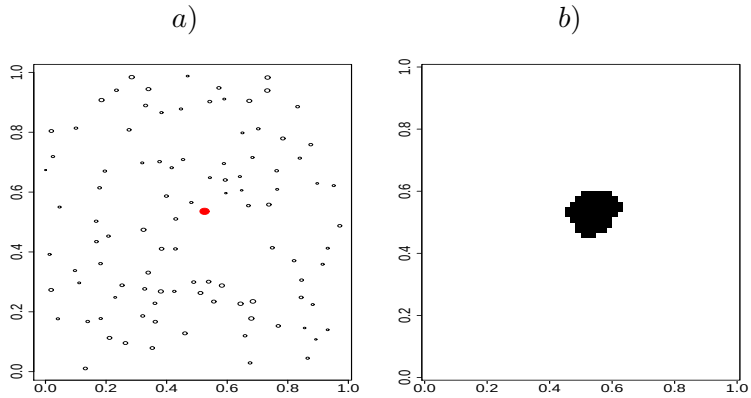


FIG. 7.7: *a*) *Echantillon de points (proportionnels à leur valeur), avec introduction d'une valeur abérante (cercle plein). b*) *ZCAs (significatives en noir) obtenues à partir de cet échantillon.*

a été appliquée à cet échantillon. Lors de la première itération, les paramètres estimés sont les suivants : portée $b = 0.1$ et palier $\sigma^2 = 1.52$. Nous avons déterminé les ZCAs au niveau $1 - \hat{\alpha} = 0.9994$ (figure 7.7*b*). Une ZCA significative est détectée au centre du domaine, correspondant à la valeur abérante. La procédure converge dès la première itération : les paramètres estimés du variogramme sont inchangés à la deuxième estimation, les ZCAs restent donc les mêmes.

Cet exemple présente un double intérêt. D'une part il montre que la méthode permet d'isoler les valeurs aberrantes et d'autre part que la procédure itérative n'amplifie pas la surface des ZCAs liées à des valeurs aberrantes.

Chapitre 8

Discussion

Nous avons proposé une méthode de détection, *i.e.* d'estimation et de test, de Zones de Changement Abrupt pour des données spatiales de schéma d'échantillonnage irrégulier et de faible densité. La méthode est basée sur les ensembles d'excursion d'un champ de χ^2 non stationnaire pour des seuils élevés, reliés au gradient prédit. Afin de construire un test global pour rejeter l'hypothèse d'absence de discontinuité, nous avons pu montrer de nouveaux résultats sur le champ de χ^2 non stationnaire, $T(\cdot)$, que nous considérons. Nous avons en particulier montré que la loi des excès du champ de χ^2 en un maximum, $\mathbb{P}\left[T(\mathbf{x}) > t + u \mid T(\mathbf{x}) > t,\, T(\mathbf{x}) \text{ est un maximum}\right]$, est exponentielle et nous avons pu établir la distribution asymptotique de la taille d'une composante connexe de l'ensemble d'excursion lorsque le seuil tend vers l'infini. Nous avons également proposé un algorithme contrôlant le *global False Discovery Rate*.

Nous allons discuter des extensions possibles de ce travail et comparer la méthode à d'autres mentionnées dans l'introduction de cette partie. Nous avons dû faire diverses hypothèses pour établir les résultats. Nous avons supposé que le champ aléatoire est gaussien, d'espérance connue, stationnaire d'ordre 2 et de fonction de covariance régulière (hypothèses $\mathcal{H}_1$ et $\mathcal{H}_2$, section 5.1). Voyons comment ces hypothèses peuvent être allégées.

8.1 Espérance inconnue

Nous avons dans la section 5.1 choisi comme interpolateur du champ aléatoire $Z(\cdot)$ le krigeage simple, en supposant que le champ était centré. En général l'espérance n'est pas nulle et est inconnue. Dans ce cas le prédicteur optimal est le krigeage ordinaire (section

3.1), noté $Z^*_{KO}(\mathbf{x})$:

$$Z^*_{KO}(\mathbf{x}) = C'(\mathbf{x})\mathbf{C}^{-1}\mathbf{Z} + \left(1 - C'(\mathbf{x})\mathbf{C}^{-1}\mathbf{1}\right)\frac{\mathbf{1}'\mathbf{C}^{-1}\mathbf{Z}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}}, \tag{8.1}$$

où $\mathbf{1} = (1, \ldots, 1)'$ est de dimension n. Comparé au prédicteur de krigeage simple (5.1), il y a un second terme dans (8.1), correspondant au filtrage de la moyenne inconnue. Le gradient interpolé est en utilisant le krigeage ordinaire :

$$W_{KO}(\mathbf{x}) = \partial C'(\mathbf{x})\mathbf{K}^{-1}\mathbf{Z}, \text{ avec } \mathbf{K}^{-1} = \mathbf{C}^{-1}\left(\mathbf{Id} - \frac{\mathbf{1}\mathbf{1}'\mathbf{C}^{-1}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}}\right).$$

Si la fonction de covariance du champ $Z(\cdot)$ est deux fois différentiable à l'origine, $W_{KO}(\mathbf{x})$ est le prédicteur optimal du gradient de $Z(\mathbf{x})$, autrement nous considérerons, toujours d'après Chilès et Delfiner (1999), que $W_{KO}(\mathbf{x})$ est un prédicteur du gradient local. Sa matrice de covariance est :

$$\mathbf{\Sigma}_{KO}(\mathbf{x}) = \mathbb{E}\left[W_{KO}(\mathbf{x})W'_{KO}(\mathbf{x})\right] = \partial C'(\mathbf{x})\mathbf{K}^{-1}\mathbf{C}^{-1}\mathbf{K}^{-1}\partial C(\mathbf{x}).$$

Comme $\mathbf{K}^{-1}\mathbf{C}^{-1}\mathbf{K}^{-1} = \mathbf{K}^{-1}$, nous avons :

$$\mathbf{\Sigma}_{KO}(\mathbf{x}) = \partial C'(\mathbf{x})\mathbf{K}^{-1}\partial C(\mathbf{x}).$$

Le champ,

$$T(\mathbf{x}) = W'_{KO}(\mathbf{x})\mathbf{\Sigma}^{-1}_{KO}(\mathbf{x})W_{KO}(\mathbf{x}),$$

est alors un champ de $\chi^2(2)$. Ainsi, le cas où l'espérance du champ $Z(\cdot)$ est inconnue est formellement identique à celui décrit dans les chapitres 5 et 6. Seul $\mathbf{C}^{-1}$ est remplacé par $\mathbf{K}^{-1}$ dans le calcul de $W(\mathbf{x})$, $U(\mathbf{x})$, $\mathbf{\Sigma}(\mathbf{x})$, $\mathbf{\Lambda}(\mathbf{x})$ et $T(\mathbf{x})$. Les propositions et théorèmes établis restent donc valides.

Nous avons illustré, figure 8.1, les ZCAs (significatives en noir et non significatives en gris) au niveau $1-\hat{\alpha}$, obtenues sur l'exemple utilisé pour illustrer la méthode, avec soit le krigeage simple (figure 8.1*a*), soit le krigeage ordinaire (figure 8.1*b*) comme interpolateur. Que nous utilisions l'un ou l'autre des interpolateurs, la discontinuité est bien détectée. Lorsque nous utilisons le krigeage ordinaire, l'effet de bord est inexistant. Cela s'explique par le fait qu'avec le krigeage ordinaire, la moyenne n'est pas fixée (à zéro dans cet exemple), mais est évaluée localement en chaque point de la grille d'interpolation.

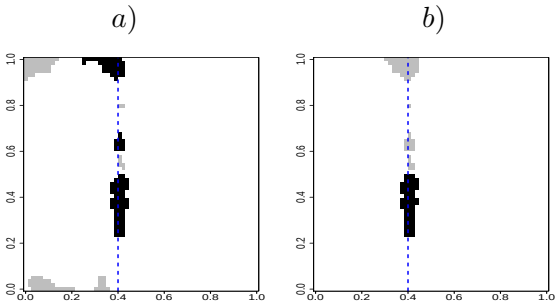


FIG. 8.1: *ZCAs (significatives en noir et non significatives en gris) au niveau* $1-\hat{\alpha}$*, obtenues en utilisant le krigeage simple (*$1-\hat{\alpha}=0.9994$*), a), ou le krigeage ordinaire (*$1-\hat{\alpha}=0.9993$*), b), comme interpolateur, à partir de l'exemple utilisé pour illustrer la méthode.*

8.2 Modèles de covariance plus généraux

Il est possible d'alléger les hypothèses $\mathcal{H}_1$ et $\mathcal{H}_2$ (section 5.1). L'hypothèse $\mathcal{H}_1$ peut en particulier être remplacée par :

$\mathcal{H}_1'$: $Z(\cdot)$ est stationnaire au sens intrinsèque (section 2.1), *i.e.*

$$\mathbb{E}\left[Z(\mathbf{x})-Z(\mathbf{y})\right]=0 \text{ et } \mathrm{Var}\left(Z(\mathbf{x})-Z(\mathbf{y})\right)=2\gamma(\mathbf{x}-\mathbf{y}), \forall \mathbf{x},\mathbf{y}\in\mathcal{D}.$$

Dans cette hypothèse moins restrictive la variance du champ $Z(\cdot)$ n'est pas nécessairement finie, la fonction de covariance n'existe pas forcément et le variogramme remplace la fonction de covariance. Il est bien connu en géostatisque que dans ce cas le prédicteur optimal de $Z(\cdot)$ en un point non échantillonné $\mathbf{x}$ est le krigeage intrinsèque. Il s'agit de l'équivalent du krigeage ordinaire dans lequel la fonction de covariance est remplacée par le variogramme (section 3.2). Ainsi sous l'hypothèse $\mathcal{H}_1'$ la méthode de détection de ZCAs reste valable pourvu que nous utilisions le krigeage intrinsèque comme prédicteur. Notons que le test d'existence de discontinuités prend alors la forme suivante :

H_0 : $\mathbb{E}\left[Z(\mathbf{x})-Z(\mathbf{y})\right]=\mathbf{0}$, $\forall \mathbf{x},\mathbf{y}\in\mathcal{D}$,

H_1 : il existe des courbes Γ telles que $\Gamma\cap\{(\mathbf{x},\mathbf{y})\in\mathcal{D}\times\mathcal{D} : \mathbb{E}\left[Z(\mathbf{x})-Z(\mathbf{y})\right]\neq\mathbf{0}\}\neq\emptyset$.

L'hypothèse $\mathcal{H}_2$ concernait la régularité de la fonction de covariance autour de $\mathbf{0}$. Pour simplifier nous avons supposé qu'elle était indéfiniment différentiable pour tout $\mathbf{h}$ tel que $\|\mathbf{h}\|>0$. Cette hypothèse est peu restrictive et est vérifiée par la plupart des fonctions de covariance usuelles. En réalité, seules les dérivées quatrièmes de la fonction de covariance sont utilisées, aussi il est suffisant de supposer la différentiabilité à l'ordre quatre.

Le modèle de covariance sphérique ne vérifie pas non plus cette hypothèse moins restrictive et par conséquent ne peut pas, en théorie, être utilisé. Cependant l'ensemble des points

où la différentiabilité n'est pas vérifiée, comme au point de raccordement $\mathbf{h} = b$, est un ensemble de mesure nulle dans le plan et considérer la différentiabilité presque sûre suffit. Aussi, cette hypothèse ne constitue pas une limite sérieuse pour l'utilisation de notre méthode.

8.3 Comparaison avec d'autres méthodes

Comparée à l'approche non paramétrique développée dans Hall *et al.* (2001) (section 1.2), la méthode que nous proposons ne nécessite pas de point de départ. Nous ne faisons donc pas l'hypothèse implicite qu'il existe une discontinuité sur le domaine, mais nous fournissons simultanément un test pour rejeter l'hypothèse nulle d'absence de ZCAs et une estimation de ces ZCAs en cas de rejet. Nous assignons également une valeur critique à chaque ZCA et de cette façon nous discriminons entre un gradient local significativement élevé et un gradient local non significatif. Notre méthode détecte, *i.e.* estime et teste, une classe plus large de courbes de discontinuité, mais à un coût d'un modèle de champ aléatoire plus particulier. Elle est à la fois plus générale et plus particulière que celle développée dans Hall *et al.* (2001). D'une part elle est plus particulière dans le sens qu'elle est reliée à l'hypothèse de gaussiannité des données. Cependant si les données ne sont pas gaussiennes, elles peuvent être transformées en utilisant une forme paramétrique (log transformation ou Cox transformation) ou une transformation non paramétrique. Cela bien sûr n'assure pas la gaussianité du champ aléatoire, mais pourrait être suffisant pour que la méthode soit utilisable en pratique (ce travail reste à effectuer). D'autre part cette hypothèse rend possible la prise en compte de la corrélation entre les points d'échantillonnage et la détection de discontinuités pour des échantillons de faible densité.

Il existe une relation étroite entre les méthodes paramétriques d'estimation par noyau et l'interpolation en utilisant une fonction de covariance. Les deux méthodes sont linéaires et le paramètre de largeur de la bande est similaire à la portée de la fonction de covariance. Dans Chaudhuri et Marron (1999, 2000) et Godtliebsen *et al.* (2001, 2002) la significativité des caractéristiques est évaluée à toutes les échelles, *i.e.* pour toutes les largeurs de bande. L'hypothèse est que la donnée $Z(\mathbf{x}_i)$ est issue d'un signal f bruité : $Z(\mathbf{x}_i) = f(\mathbf{x}_i)+\epsilon_i$, où ϵ_i est un bruit gaussien indépendant. Sous l'hypothèse nulle d'un signal constant, l'ensemble de données est un échantillon de points indépendants et identiquement distribués d'une variable aléatoire. La significativité des caractéristiques du signal inconnu est évaluée grâce à un estimateur à noyau de f. Concernant les applications que nous avons à l'esprit, l'hypothèse d'un bruit indépendant n'est pas réaliste et nous considérons que ϵ est une

fonction aléatoire. Plus précisément nous supposons qu'il s'agit d'un champ aléatoire stationnaire dont la fonction de covariance est à estimer. Comme nous l'avons vu dans la section 8.2, nous pouvons considérer que ϵ est un champ aléatoire intrinsèque avec un variogramme. Ayant estimé les paramètres de la fonction de covariance à partir des données, il n'est pas nécessaire de détecter les caractéristiques significatives à toutes les échelles. Nous préférons plutôt estimer les paramètres de la fonction de covariance qui ajustent au mieux les données. Par conséquent notre méthode nécessite moins de calculs et est plus simple à interpréter visuellement. De plus, comme nous le verrons dans la section 12.1, la méthode est robuste par rapport à une erreur d'estimation de la portée de la fonction de covariance.

8.4 Amélioration de l'algorithme

Dans la procédure itérative (section 7.3), nous éliminons du calcul du variogramme expérimental les couples $\{Z(\mathbf{x}_i), Z(\mathbf{x}_j)\}$ tels que le segment $[\mathbf{x}_i, \mathbf{x}_j]$ intersecte une ZCA au niveau $1-\alpha$ (figure 7.5).

Jusqu'à présent nous avons calculé et visualisé les ZCAs au niveau $1-\alpha = 1-\hat{\alpha}$. Cependant, rien ne nous empêche d'utiliser un niveau $1-\alpha$ plus faible afin d'obtenir une meilleure visualisation des ZCAs. La figure 8.2*a* représente les ZCAs (significatives en noir) au niveau $1-\hat{\alpha}$ obtenues à la première itération de l'exemple utilisé pour illustrer la procédure itérative. Sur la figure 8.2*b*, nous avons représenté les ZCAs significatives de la figure 8.2*a* à un niveau plus faible, $1-\alpha_0 = 0.99$. Nous pouvons constater que la représentation à un niveau plus faible entraîne que la ZCA est plus grande. L'augmentation de sa surface n'est dûe qu'à un effet d'un meilleur alignement le long de la discontinuité, la discontinuité est mieux délimitée.

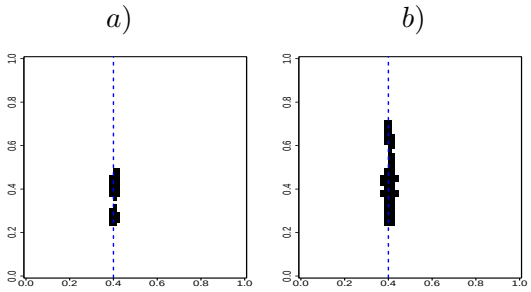


FIG. 8.2: *a) ZCAs significatives (en noir) au niveau* $1-\hat{\alpha}$. *b) ZCAs significatives au niveau* $1-\hat{\alpha}$ *représentées au niveau* $1-\alpha_0 = 0.99$.

Nous proposons d'améliorer la procédure itérative de la façon suivante. Au lieu de retirer du calcul du variogramme expérimental les couples $\{Z(\mathbf{x}_i), Z(\mathbf{x}_j)\}$ tels que le segment $[\mathbf{x}_i, \mathbf{x}_j]$ intersecte une ZCA au niveau $1-\hat{\alpha}$, nous éliminons les couples $\{Z(\mathbf{x}_i), Z(\mathbf{x}_j)\}$ tels que le segment $[\mathbf{x}_i, \mathbf{x}_j]$ intersecte une ZCA au niveau $1-\alpha_0$ ($\leq 1-\hat{\alpha}$) contenant une ZCA significative au niveau $1-\hat{\alpha}$ (figure 8.3).

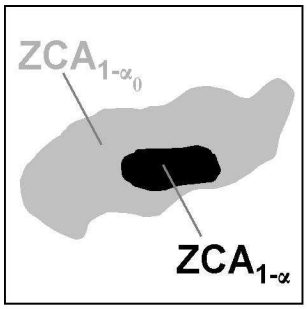


FIG. 8.3: *Exemple de ZCA au niveau $1-\alpha_0$ (en gris) contenant une ZCA significative au niveau $1-\alpha$ (en noir).*

Voici un exemple sur simulation de l'amélioration de la procédure itérative. Considérons l'exemple précédent où nous avions simulé dans le carré unitaire 100 points répartis aléatoirement et issus d'un champ gaussien centré réduit de fonction de covariance exponentielle de portée $b = 0.1$. Une discontinuité $a = 2.5$ avait été introduite en $x^1 = 0.4$. Les paramètres (portée b et variance σ^2) des fonctions de covariance estimées à la première itération et à la convergence en fonction du niveau $1-\alpha_0$ sont reportés dans la table 8.1.

Itération 1		Convergence			
b	σ^2	Itération	b	σ^2	$1-\alpha_0$
0.275	2.99	5	0.21	1.35	$1-\hat{\alpha} = 0.9994$
		4	0.21	1.32	0.99
		4	0.18	1.26	0.95
		3	0.15	1.13	0.90

TAB. 8.1: *Portée b et variance σ^2 estimées à la première itération et à la convergence pour différentes valeurs de $1-\alpha_0$.*

Les variogrammes correspondants sont illustrés sur la première ligne de la figure 8.4. La table 8.1 montre que plus nous considérons une valeur faible pour le niveau $1-\alpha_0$, plus le modèle obtenu à la convergence se rapproche du modèle initial (modèle exponentiel de portée $b = 0.1$ et de variance $\sigma^2 = 1$). Notons que ces paramètres varient peu lorsque nous considérons $1-\alpha_0 \leq 0.99$, ce qui montre que l'utilisation de petites valeurs du

niveau $1-\alpha_0$ n'entraîne pas de changement dans la détection des ZCAs (du fait que nous imposons que les ZCAs au niveau $1-\alpha_0$ doivent contenir une ZCA significative au niveau $1-\hat{\alpha}$). Cela est illustré sur la deuxième ligne de la figure 8.4, où nous avons représenté les ZCAs obtenues à la première itération et à la convergence en fonction du niveau $1-\alpha_0$.

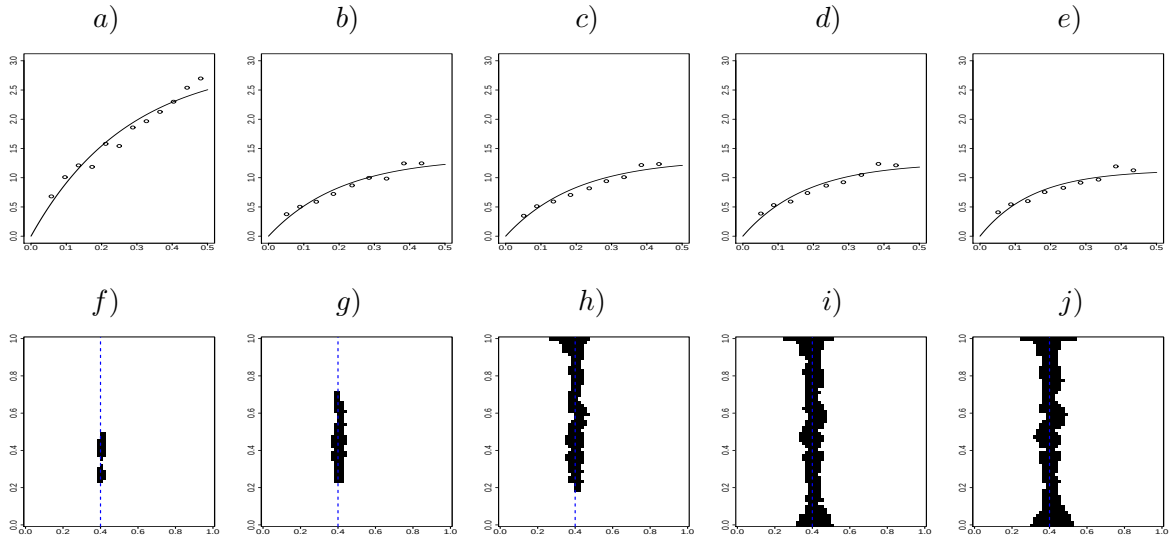


FIG. 8.4: *Illustration de l'amélioration de la procédure itérative. Variogrammes et ZCAs à la première itération : a) et f), à la convergence pour $1-\hat{\alpha}$: b) et g), pour $1-\alpha_0 = 0.99$: c) h), pour $1-\alpha_0 = 0.95$: d) et i), pour $1-\alpha_0 = 0.90$: e) et j).*

Remarquons que considérer le niveau $1-\alpha_0$ dans la procédure itérative n'a pas pour conséquence d'amplifier la taille des ZCAs dûes à des valeurs aberrantes. En effet, nous avons repris l'exemple de la deuxième illustration de la section 7.3 et appliqué la procédure itérative avec $1-\alpha_0 \in \{0.99, 0.95\}$. Quelquesoit $1-\alpha_0$, la procédure converge à la première itération.

8.5 Schéma d'échantillonnage

La méthode n'impose pas de schéma d'échantillonnage particulier et s'applique aussi bien à des échantillons de points régulièrement spatialisés qu'irrégulièrement. Cependant la densité locale de l'échantillon a quelques conséquences sur la méthode. Lorsque la densité locale est très élevée, il peut arriver que le gradient prédit soit artificiellement élevé, ce qui mène à des ZCAs statistiquement significatives mais sans signification physique. De telles ZCAs sont de très petites tailles. C'est en particulier le cas en présence de valeurs

aberrantes. La méthode peut donc être utilisée pour détecter les valeurs aberrantes dans un contexte spatial, par exemple si il y a de petites ZCAs isolées contenant un seul point d'échantillonnage.

Inversément, dans une partie non échantillonnée, en fonction de la portée, la matrice de covariance du gradient prédit peut être très faible ou singulière ce qui mène à l'impossibilité d'estimer les ZCAs. Mais comme nous l'avons vu section 5.2, nous excluons les sites où la matrice de covariance $\mathbf{\Sigma}$ n'est pas inversible.

La densité locale d'échantillonnage a une conséquence directe sur la puissance de la méthode. Lorsque l'échantillon est localement peu dense, il n'est pas possible de détecter des changements abrupts car nous ne pouvons pas définir le type de transition (régulière ou forte) correspondant au gradient prédit. Cela nous mène à considérer la puissance du test local de détection. La puissance globale a été explorée dans une étude de simulation (chapitre 11). En spécifiant l'hypothèse alternative, nous avons pu approcher la puissance locale, *i.e.* la probabilité de détecter en $\mathbf{x}$ une courbe de discontinuité passant par $\mathbf{x}$. Ce travail est détaillé dans la partie suivante.

Troisième partie

Puissance du test de détection

Introduction

La méthode d'estimation des zones où une variable faiblement échantillonnée dans le plan varie brusquement, décrite dans la partie précédente, suppose que la variable est spatialisée et faiblement échantillonnée. Elle repose sur un test dont l'hypothèse nulle est la stationnarité du champ aléatoire $Z(\cdot)$ modélisant cette variable et l'alternative est l'existence de courbes régulières Γ sur lesquelles $\mathbb{E}[Z(\cdot)]$ est discontinue. Comme dans toute analyse géostatistique cette méthode requiert un bon compromis entre une bonne couverture du domaine et un échantillonnage de toutes les distances afin de parvenir à une bonne estimation du modèle de covariance. Rappelons que la méthode est basée sur l'étude du gradient local. Celui-ci devant être bien estimé, il est nécessaire d'avoir un échantillon suffisamment dense en tout point. En effet la densité locale d'échantillonnage a une conséquence directe sur la puissance de la méthode : il est difficile de détecter des changements abrupts lorsque l'échantillon est trop clairsemé. Le manque d'échantillons ne permet pas de tester si le gradient local prédit correspond à une transition régulière ou localement forte. Cela mène à nous poser la question de la puissance du test. Cependant considérer la puissance globale en fonction de la densité de points (comme dans la section 11.2) n'est pas pleinement satisfaisant. En effet cela fournit la probabilité qu'une discontinuité soit détectée sachant son existence, mais cela ne nous indique pas si toutes les zones du domaine sont suffisamment échantillonnées. Aussi nous calculons la puissance du test de détection en chaque point $\mathbf{x}$ du domaine $\mathcal{D}$. En pratique, cette puissance sera calculée en chaque pixel de la grille d'interpolation. L'analyse de la puissance montre qu'elle n'est pas constante sur le domaine. La cartographie de la puissance permet d'identifier les zones de faible puissance qui définissent les lieux où l'échantillonnage n'est pas approprié à l'estimation des ZCAs.

Le calcul de la puissance est complexe d'une part parce qu'il faut se placer sous l'hypothèse alternative d'existence d'une discontinuité alors que l'on ne connaît pas la forme de cette discontinuité, d'autre part parce que les tests locaux ne sont pas indépendants.

Nous considérons que le point $\mathbf{x}$ est le centre d'une fenêtre mobile et qu'il appartient à une courbe de discontinuité Γ supposée assez régulière, pouvant être localement approchée par sa tangente en $\mathbf{x}$, de direction inconnue puisque la forme de la discontinuité est indéterminée. Les hypothèses sont alors la stationnarité du champ $Z(\cdot)$ (H_0) et l'existence d'une courbe de discontinuité de $\mathbb{E}[Z(\cdot)]$ représentée par une droite passant par $\mathbf{x}$ (H_1) telle que de part et d'autre le champ $Z(\cdot)$ ait pour différence de moyenne une constante a. Ainsi, sous l'alternative, le champ $Z(\cdot)$ a la forme suivante :

$$Z(\mathbf{x}) = C'(\mathbf{x})\mathbf{C}^{-1}\left[Z + A(\mathbf{x})\right],$$

avec $A(\mathbf{x})$, vecteur de longueur n, d'éléments $\pm a/2$ selon que le point de donné est d'un côté ou de l'autre de la discontinuité. Sous l'alternative, le gradient $W(\mathbf{x})$ se décompose alors en la somme du gradient sous l'hypothèse nulle et d'une quantité déterministe :

$$\begin{aligned} W(\mathbf{x}) &= \partial C'(\mathbf{x})\mathbf{C}^{-1}Z + \partial C'(\mathbf{x})\mathbf{C}^{-1}A(\mathbf{x}) \\ &= W_{H_0}(\mathbf{x}) + k_a(\mathbf{x}). \end{aligned} \tag{8.2}$$

Le calcul de la puissance au centre d'une fenêtre mobile a nécessité diverses étapes de construction. Nous avons dans un premier temps cherché à calculer la puissance en un point x appartenant à la droite $\mathbb{R}$ (chapitre 9). Ce travail a en particulier permis d'étudier le comportement de la puissance entre deux points de $\mathbb{R}$. La puissance en un point $\mathbf{x}$ de $\mathcal{D} \subset \mathbb{R}^2$ est ensuite calculée en considérant que la discontinuité est une droite passant par $\mathbf{x}$ telle que de part et d'autre le champ ait pour différence de moyenne une constante a (chapitre 10). Le calcul de cette puissance nous permet de montrer que la non indépendance des tests locaux n'est pas négligeable.

Chapitre 9

Calcul de la puissance pour un champ sur $\mathbb{R}$

Dans cette partie, nous allons calculer la puissance du test de détection pour un champ sur $\mathbb{R}$. La puissance est calculée en un point $x \in \mathbb{R}$ pour différents interpolateurs du champ : krigeage simple et ordinaire (section 3.1).
Nous analysons ensuite deux cas particuliers. Dans le premier cas nous supposons que le modèle de covariance est exponentiel et dans le second sphérique. Nous étudions à travers ces deux cas l'évolution de la puissance le long d'un segment, *i.e.* entre deux points de données.

9.1 Calcul de la puissance

Soit $Z(\cdot) = \{Z(x),\ x \in \mathbb{R}\}$ un champ aléatoire gaussien de moyenne nulle et de variance unité. $Z(\cdot)$ est supposé stationnaire et de fonction de covariance, $C_Z(h)$ indéfiniment différentiable pour tout $|h| > 0$.

Considérons le modèle simple suivant. Soient $x_1 = 0$ et $x_2 = d$ deux points de $\mathbb{R}$ et $\mathbf{Z} = (Z_1, Z_2)'$ la représentation du champ $Z(\cdot)$ en x_1 et x_2. Nous allons calculer la puissance en un point x appartenant au segment $[x_1, x_2]$. Nous introduisons une discontinuité en ajoutant une constante a_i aux valeurs Z_i. Nous prendrons souvent $a_1 = -a_2$. Ainsi, en présence d'une discontinuité le vecteur de données devient : $\mathbf{Z} + A(x) = (Z_1 + a_1, Z_2 + a_2)'$. Afin de calculer la puissance du test de détection de ZCAs, nous cherchons à écrire la statistique $T(x)$ en présence d'une discontinuité, ce qui revient à décomposer $T(x)$ en fonction de la statistique sous l'hypothèse nulle, $T_{H_0}(x)$, et d'une constante $\tau(x)$. Rappe-

lons que $T(x)$ s'écrit $T(x) = W'(x)\Sigma^{-1}(x)W(x)$. Comme $W(x) \sim \mathcal{N}(0, \Sigma(x))$, nous avons $T(x) \sim \chi^2(1)$. Afin d'évaluer $T(x)$, il est nécessaire de déterminer le prédicteur $Z^*(x)$ de $Z(x)$, le gradient de ce prédicteur, $W(x) = \partial Z^*(x)$ ainsi que la variance de $W(x)$. Ces quantités ont été évaluées dans la section 3.1 et sont résumées dans la table 9.1.

	Krigeage simple	**Krigeage ordinaire**
$Z^*(x) =$	$C'(x)\mathbf{C}^{-1}\mathbf{Z}$	$C'(x)\mathbf{C}^{-1}\mathbf{Z} + \left(1 - C'(x)\mathbf{C}^{-1}\mathbf{1}\right)\frac{\mathbf{1'C^{-1}Z}}{\mathbf{1'C^{-1}1}}$
$W(x) =$	$\partial C'(x)\mathbf{C}^{-1}\mathbf{Z}$	$\partial C'(x)\mathbf{K}^{-1}\mathbf{Z}$
$\Sigma(x) =$	$\partial C'(x)\mathbf{C}^{-1}\partial C(x)$	$\partial C'(x)\mathbf{K}^{-1}\partial C(x)$
		où $\mathbf{K}^{-1} = \mathbf{C}^{-1}\left(\mathbf{Id} - \frac{\mathbf{11'C^{-1}}}{\mathbf{1'C^{-1}1}}\right)$

TAB. 9.1: *Prédicteur de $Z(x)$: $Z^*(x)$, gradient de $Z^*(x)$: $W(x)$ et variance de $W(x)$: $\Sigma(x)$, en fonction de l'interpolateur de $Z(x)$.*

Lemme 6 *Pour $x \in \mathbb{R}$, en présence d'une discontinuité la statistique $T(x)$ se décompose de la façon suivante :*

$$T(x) = \left(\sqrt{T_{H_0}(x)} + \tau(x)\right)^2,$$

où $\tau(x) = k_a(x)/\Sigma^{1/2}(x)$ avec $k_a(x) = \partial C'(x)\mathbf{G}^{-1}A(x)$, $\mathbf{G}^{-1} = \mathbf{C}^{-1}$ dans le cas du krigeage simple et $\mathbf{G}^{-1} = \mathbf{K}^{-1}$ dans le cas du krigeage ordinaire.

Le résultat est immédiat d'après (8.2) et $T(x) = W^2(x)/\Sigma(x)$.
Le lemme 6 permet d'évaluer la puissance du test de détection en un point x de $\mathbb{R}$.

Proposition 11 *Le taux de détection d'une discontinuité en un point $x \in \mathbb{R}$ est :*

$$1 - \beta(x) = 1 - \Phi\left(\sqrt{t_\alpha} - \tau(x)\right) + \Phi\left(-\sqrt{t_\alpha} - \tau(x)\right), \tag{9.1}$$

où Φ est la fonction de répartition d'une loi normale $\mathcal{N}(0,1)$, $t_\alpha = \chi^2_{1-\alpha}(1)$ et $\tau(x)$ est défini dans le lemme 6.

Preuve :

$$\begin{aligned}
1 - \beta(x) &= \mathbb{P}[\text{ rejeter } H_0 \mid H_1 \text{ vraie }] \\
&= \mathbb{P}_{H_1}[T(x) \geq t_\alpha] \\
&= \mathbb{P}_{H_1}\left[\left(\sqrt{T_{H_0}(x)} + \tau(x)\right)^2 \geq t_\alpha\right], \text{ d'après le lemme 6} \\
&= \mathbb{P}_{H_1}\left[\sqrt{T_{H_0}(x)} \geq \sqrt{t_\alpha} - \tau(x)\right] + \mathbb{P}_{H_1}\left[\sqrt{T_{H_0}(x)} \leq -\sqrt{t_\alpha} - \tau(x)\right] \\
&= 1 - \Phi\left(\sqrt{t_\alpha} - \tau(x)\right) + \Phi\left(-\sqrt{t_\alpha} - \tau(x)\right),
\end{aligned}$$

où Φ est la fonction de répartition d'une loi normale $\mathcal{N}(0,1)$. En effet, puisque nous avons $W_{H_0}(x) \sim \mathcal{N}(0, \Sigma(x))$, il s'en suit que $\sqrt{T_{H_0}(x)} \sim \mathcal{N}(0,1)$. □

9.2 Pour un modèle de covariance exponentiel

Toujours dans le cadre où nous calculons la puissance en $x \in [x_1, x_2]$, considérons comme modèle de covariance, un modèle exponentiel de portée b :

$$C_Z(h) = \exp(-h/b) \text{ si } h \geq 0.$$

La covariance entre x et les points de données est $C(x) = \left(e^{-x/b}, e^{-(d-x)/b}\right)'$ et sa dérivée par rapport à x s'écrit $\partial C(x) = \left(-e^{-x/b}/b, e^{-(d-x)/b}/b\right)'$.

La matrice de covariance entre les points d'échantillonnage est $\mathbf{C} = \begin{pmatrix} 1 & e^{-d/b} \\ e^{-d/b} & 1 \end{pmatrix}$,

par conséquent $\mathbf{C}^{-1} = \frac{1}{1-e^{-2d/b}} \begin{pmatrix} 1 & -e^{-d/b} \\ -e^{-d/b} & 1 \end{pmatrix}$.

Avec le krigeage simple

Compte tenu du modèle de covariance et de l'introduction d'une discontinuité (a_1, a_2), nous pouvons écrire :

$$\begin{aligned}
Z^*(x) &= C'(x)\mathbf{C}^{-1}\mathbf{Z} \\
&= \left\{\sinh\left(\tfrac{d-x}{b}\right)(Z_1 + a_1) + \sinh\left(\tfrac{x}{b}\right)(Z_2 + a_2)\right\} / \ \sinh\left(\tfrac{d}{b}\right), \\
W(x) &= \partial C'(x)\mathbf{C}^{-1}\mathbf{Z} \\
&= \left\{-\cosh\left(\tfrac{d-x}{b}\right) Z_1 + \cosh\left(\tfrac{x}{b}\right) Z_2\right\} / \left(b \sinh\left(\tfrac{d}{b}\right)\right) \\
&\quad + \left\{-\cosh\left(\tfrac{d-x}{b}\right) a_1 + \cosh\left(\tfrac{x}{b}\right) a_2\right\} / \left(b \sinh\left(\tfrac{d}{b}\right)\right) \\
&= W_{H_0}(x) + k_a(x), \\
\Sigma(x) &= \partial C'(x)\mathbf{C}^{-1}\partial C(x) = \left\{\cosh\left(\tfrac{d-x}{b}\right) e^{-x/b} + \cosh\left(\tfrac{x}{b}\right) e^{-(d-x)/b}\right\} / b^2 \sinh\left(\tfrac{d}{b}\right),
\end{aligned}$$

où les fonctions sinus et cosinus hyperboliques sont définies par :

$$\cosh(x) = \frac{e^x + e^{-x}}{2}, \ \sinh(x) = \frac{e^x - e^{-x}}{2}, \ x \in \mathbb{R}.$$

Avec le krigeage ordinaire

Les champs $Z^*(x)$, $W(x)$ et la variance $\Sigma(x)$ prennent dans ce cas la forme suivante :

$$
\begin{aligned}
Z^*(x) &= C'(x)\mathbf{C}^{-1}\mathbf{Z} + (1 - \mathbf{C}'(\mathbf{x})\mathbf{C}^{-1}\mathbf{1})\,\tfrac{\mathbf{1}'\mathbf{C}^{-1}\mathbf{Z}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}} \\
&= \left[\tfrac{1}{2} + \tfrac{e^{-x/b} - e^{-(d-x)/b}}{2\gamma(d)}\right](Z_1 + a_1) + \left[\tfrac{1}{2} - \tfrac{e^{-x/b} - e^{-(d-x)/b}}{2\gamma(d)}\right](Z_2 + a_2), \\
&\quad \text{avec } \gamma(d) = 1 - e^{-d/b}, \\
W(x) &= \partial C'(x)\mathbf{K}^{-1}\mathbf{Z} \\
&= (Z_2 - Z_1)\left(e^{-x/b} + e^{-(d-x)/b}\right)/2b\gamma(d) + (a_2 - a_1)\left(e^{-x/b} + e^{-(d-x)/b}\right)/2b\gamma(d) \\
&= W_{H_0}(x) + k_a(x), \\
\Sigma(x) &= \partial C'(x)\mathbf{K}^{-1}\partial C(x) = \left(e^{-x/b} + e^{-(d-x)/b}\right)^2/2b^2\gamma(d).
\end{aligned}
$$

Dans ce cas la statistique $T(x)$ ne dépend plus de x :

$$T(x) = \frac{W^2(x)}{\Sigma(x)} = \frac{\left[(Z_2 - Z_1) + (a_2 - a_1)\right]^2}{2\gamma(d)}.$$

Aussi la puissance en x (9.1) peut s'écrire :

$$1 - \beta(x) = 1 - \Phi\left(\sqrt{t_\alpha} - \frac{a_2 - a_1}{\sqrt{2\gamma(d)}}\right) + \Phi\left(-\sqrt{t_\alpha} - \frac{a_2 - a_1}{\sqrt{2\gamma(d)}}\right), \tag{9.2}$$

où Φ est la fonction de répartition d'une loi normale $\mathcal{N}(0,1)$. Remarquons que la puissance est aussi indépendante de x.

Illustration

Nous avons calculé la puissance en faisant varier le point x entre $x_1 = 0$ et $x_2 = 1$ pour différentes valeurs des paramètres b, a_1 et a_2.
Nous avons considéré $b \in \{0.05, 0.1, 0.2, 0.4, 0.8\}$. La figure 9.1 représente la puissance pour le krigeage simple (sur les deux premières lignes) en fonction de la distance x. Pour $a_1 = -a_2$, avec $a_1 \in \{1, 1.5, 2\}$ (figures 9.1a, 9.1b, 9.1c respectivement), ces résultats montrent que quelle que soit la valeur de b, la puissance est maximale à mi-distance de x_1 et de x_2, ce qui est intuitif puisque les valeurs affectées à x_1 et x_2 sont les mêmes au signe près. Il apparaît également que plus la portée, b, est grande, plus la puissance est élevée et plus elle a tendance à devenir constante le long du segment $[x_1, x_2]$ (elle devient quasi constante pour $b = 3d/2, \forall a_1$).
Pour $(a_1, a_2) = (2, 0), (2, -1)$ et $(3, -1)$ (figures 9.1d, 9.1e, 9.1f respectivement), la puissance atteint son maximum pour $x \in [x_i, d/2]$ où x_i est tel que $|a_i| = \max\{|a_1|, |a_2|\}$. Elle est également plus élevée sur cet intervalle que sur l'autre. La puissance augmente là aussi avec la portée.

En utilisant le krigeage ordinaire (dernière ligne) et $a_1 = -a_2,\ a_1 \in \{1, 1.5, 2\}$ (figures 9.1g, 9.1h, 9.1i respectivement), les résultats montrent que quelle que soit la valeur de b, la puissance est constante le long du segment $[x_1, x_2]$, ce qui illustre l'équation (9.2).

Pour un champ $Z(\cdot)$ de fonction de covariance exponentielle, des problèmes numériques surviennent lorsque le rapport d/b est supérieur à 200. Dans ce cas $\Sigma(x)$ atteint des valeurs inférieures à 10^{-82}. $\Sigma(x)$ étant alors très proche de 0, nous ne pouvons plus écrire $T(x) = W'(x)\Sigma^{-1}(x)W(x)$. Ce phénomène n'apparaîtra pas en pratique dans notre cas. En effet, avoir une portée très petite devant la taille du domaine signifie qu'il y a indépendance entre les points de données. Cela nous ramène à avoir un échantillon de très faible densité, ce qui n'est pas le cas des applications possibles pour la méthode car l'interpolation n'est pas réalisable. Comme nous le verrons par la suite lors de l'application et des simulations, pour un domaine carré de côté d, les échantillons considérés auront une portée pratique de l'ordre de $d/3$.
Intéressons-nous à l'évolution de $\Sigma(x)$ et plus particulièrement à la valeur de son minimum en fonction du rapport d/b. Dans le cas du krigeage simple, nous avons vu que $\Sigma(x) = \left\{\cosh\left(\frac{d-x}{b}\right) e^{-x/b} + \cosh\left(\frac{x}{b}\right) e^{-(d-x)/b}\right\} / \left(b^2 \sinh\left(\frac{d}{b}\right)\right)$. Cette fonction est positive, décroissante pour $x \in [x_1, d/2]$, et croisssante pour $x \in [d/2, x_2]$. Elle atteint son minimum en $x = d/2$. Dans ce cas,

$$\Sigma(d/2) = \frac{1}{b^2 \sinh(d/b)}.$$

Si $b \gg d$, alors le rapport $d/b \to 0$, $\sinh(d/b) \to 0$ et $\Sigma(d/2) \to \infty$. Au contraire, lorsque $d \gg b$, le rapport d/b devient très grand, nous avons $\sinh(d/b) \to \infty$ et $\Sigma(d/2) \to 0$.

9.3 Pour un modèle de covariance sphérique

Considérons maintenant le modèle de covariance sphérique,

$$C_Z(h) = \begin{cases} 1 & \text{si } h = 0 \\ 1 - \frac{3}{2}\frac{h}{b} + \frac{1}{2}\left(\frac{h}{b}\right)^3 & \text{si } 0 < h < b \\ 0 & \text{si } b \geq h \end{cases} \tag{9.3}$$

Pour ce modèle nous avons, $C(x) = (c(x), c(d-x))'$, où $c(x) = C_Z(x)$. Alors
$\partial C(x) = (\partial c(x), \partial c(d-x))'$, $\mathbf{C} = \begin{pmatrix} 1 & c(d) \\ c(d) & 1 \end{pmatrix}$ et $\mathbf{C}^{-1} = \frac{1}{1-c^2(d)} \begin{pmatrix} 1 & -c(d) \\ -c(d) & 1 \end{pmatrix}$.

Avec le krigeage simple

Pour ce modèle et l'introduction d'une discontinuité (a_1, a_2), nous avons :

$$\begin{aligned} Z^*(x) &= \{(c(x) - c(d)c(d-x))(Z_1 + a_1) \\ &\quad + (c(d-x) - c(d)c(x))(Z_2 + a_2)\} / (1 - c^2(d)), \\ W(x) &= \{(\partial c(x) - c(d)\partial c(d-x)) Z_1 + (\partial c(d-x) - c(d)\partial c(x)) Z_2\} / (1 - c^2(d)) \\ &\quad + \{(\partial c(x) - c(d)\partial c(d-x)) a_1 + (\partial c(d-x) - c(d)\partial c(x)) a_2\} / (1 - c^2(d)) \\ &= W_{H_0}(x) + k_a(x) \end{aligned}$$

$$\Sigma(x) = \{(\partial c(x))^2 - 2c(d)\partial c(x)\partial c(d-x) + (\partial c(d-x))^2\} / (1 - c^2(d)).$$

Lorsque $b < d/2$ il est impossible de calculer la puissance sur l'intervalle $[b, d-b]$, car dans ce cas $Z^*(x) = W(x) = \Sigma(x) = 0$.

Avec le krigeage ordinaire

Les champs $Z^*(x)$ et $W(x)$ et la variance $\Sigma(x)$ ont la forme suivante,

$$\begin{aligned} Z^*(x) &= C'(x)\mathbf{C}^{-1}\mathbf{Z} + (1 - \mathbf{C}'(\mathbf{x})\mathbf{C}^{-1}\mathbf{1}) \frac{\mathbf{1}'\mathbf{C}^{-1}\mathbf{Z}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}} \\ &= \{(c(x) - c(d)c(d-x) + (1 - c^2(d) - c(x) + c(d)c(d-x) + c(d)c(x) \\ &\quad - c(d-x))/2)(Z_1 + a_1) + (c(d-x) - c(d)c(x) + (1 - c^2(d) - c(x) \\ &\quad + c(d)c(d-x) + c(d)c(x) - c(d-x))/2)(Z_2 + a_2)\} / (1 - c^2(d)), \\ W(x) &= \partial C'(x)\mathbf{K}^{-1}\mathbf{Z} \\ &= \{(Z_2 - Z_1)(\partial c(d-x) - \partial c(x)) + (a_2 - a_1)(\partial c(d-x) - \partial c(x))\} / 2(1 - c(d)) \\ &= W_{H_0}(x) + k_a(x), \\ \Sigma(x) &= \partial C'(x)\mathbf{K}^{-1}\partial C(x) \\ &= (\partial c(d-x) - \partial c(x))^2 / 2(1 - c(d)) \end{aligned}$$

Lorsque la portée est supérieure à $d/2$ et lorsque $d - x > b$ ou $x > b$, nous avons d'après la relation (9.3), $Z^*(x) = W(x) = \mathbf{\Sigma}(x) = 0$. Ainsi, l'utilisation du modèle sphérique se limite aux cas où $b < d/2$ et $x \in [b-d, b]$.

L'expression de $T(x)$ est alors :

$$T(x) = \frac{W^2(x)}{\Sigma(x)} = \frac{[(Z_2 - Z_1) + (a_2 - a_1)]^2}{2(1 - c(d))}.$$

Elle ne dépend pas de x. Aussi la puissance en x (9.1) est aussi indépendante de x et s'écrit :

$$1 - \beta(x) = 1 - \Phi\left(\sqrt{t_\alpha} - \frac{a_2 - a_1}{\sqrt{2(1 - c(d))}}\right) + \Phi\left(-\sqrt{t_\alpha} - \frac{a_2 - a_1}{\sqrt{2(1 - c(d))}}\right), \quad (9.4)$$

où Φ est la fonction de répartition d'une loi normale $\mathcal{N}(0, 1)$.

Illustration

La puissance a été calculée en déplaçant le point x de $x_1 = 0$ à $x_2 = 1$ et en faisant varier les paramètres b, a_1 et a_2. Nous avons considéré $b \in \{0.5, 0.6, 0.8, 1, 1.2\}$. La figure 9.2 représente la puissance en fonction de la distance x. Le krigeage simple a été utilisé sur les deux premières lignes. Pour $a_1 = -a_2,\ a_1 \in \{1, 1.5, 2\}$ (figures 9.2a, 9.2b, 9.2c respectivement), les résultats montrent que, quelle que soit la valeur de b, la puissance est maximale à mi-distance de x_1 et de x_2 (sauf pour $b = d/2$ auquel cas la puissance est constante le long de $[x_1, x_2]$. Il apparaît également que plus la portée b est grande, plus la puissance est élevée et plus elle a tendance à devenir constante le long du segment $[x_1, x_2]$. Nous avons également considéré le cas où les a_i sont différents en valeur absolue, $(a_1, a_2) = (2, 0), (2, -1), (3, -1)$ (figures 9.2d, 9.2e, 9.2f respectivement). Dans ce cas, la puissance atteint son maximum pour $x \in [x_i, d/2]$ où x_i est tel que $|a_i| = \max\{|a_1|, |a_2|\}$. Elle est également plus élevée sur cet intervalle que sur l'autre. La puissance augmente là aussi avec la portée.
Dans le cas du krigeage ordinaire (dernière ligne) avec, $a_1 = -a_2,\ a_1 \in \{1, 1.5, 2\}$ (figures 9.2g, 9.2h, 9.2i respectivement), la puissance est constante, son expression (équation (9.4)) ne dépendant pas de x.

Conclusion

Dans les deux exemples de fonction de covariance que nous avons utilisé, nous pouvons constater que la puissance entre deux points $x_1, x_2 \in \mathbb{R}$ évolue de la même façon. En utilisant le krigeage simple comme interpolateur, la puissance est maximale en le centre du segment $[x_1, x_2]$ lorsque les deux points ont le même poids en valeur absolue et est maximale entre le centre de $[x_1, x_2]$ et le point où le poids est le plus élevé en valeur absolue. Dans ce cas, plus la portée diminue, plus le maximum tend à se rapprocher du centre du segment $[x_1, x_2]$. Elle est constante lorsque nous utilisons le krigeage ordinaire, ce que nous savions de part les équations (9.2) et (9.4), et est égale à la valeur maximale obtenue pour le krigeage simple. Aussi vaut-il mieux ignorer la connaissance de la moyenne théorique et privilégier le krigeage ordinaire.

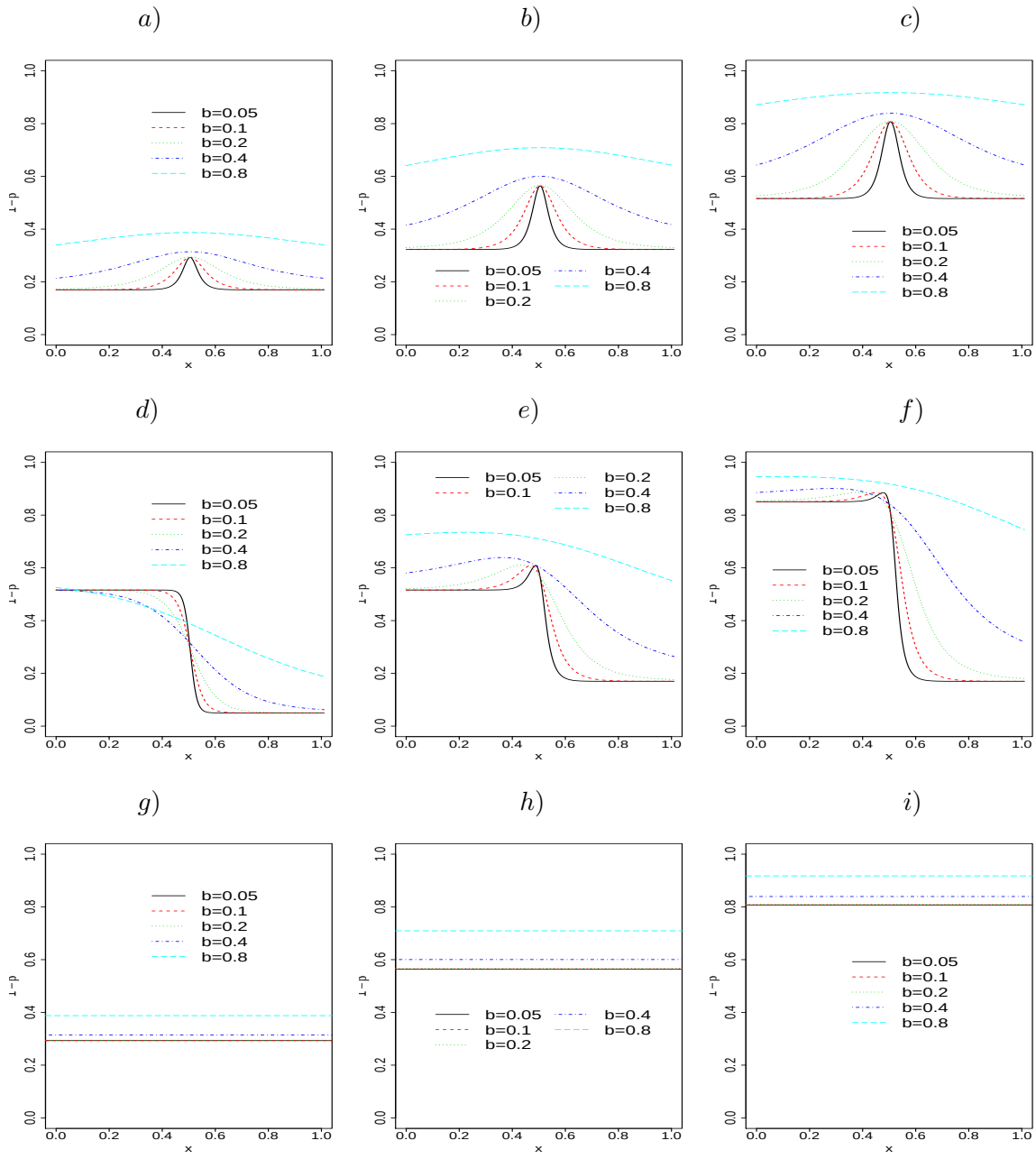


FIG. 9.1: *Puissance en* $x \in [0,1]$ *dans le cas d'un modèle de covariance exponentiel de portée* $b \in \{0.05, 0.1, 0.2, 0.4, 0.8\}$, *pour le krigeage simple (deux premières lignes) avec a)* $(a_1, a_2) = (1, -1)$, *b)* $(1.5, -1.5)$ *et c)* $(2, -2)$, *d)* $(a_1, a_2) = (2, 0)$, *e)* $(2, -1)$ *et f)* $(3, -1)$ *et pour le krigrage ordinaire (dernière ligne) avec g)* $(a_1, a_2) = (1, -1)$, *h)* $(1.5, -1.5)$ *et i)* $(2, -2)$.

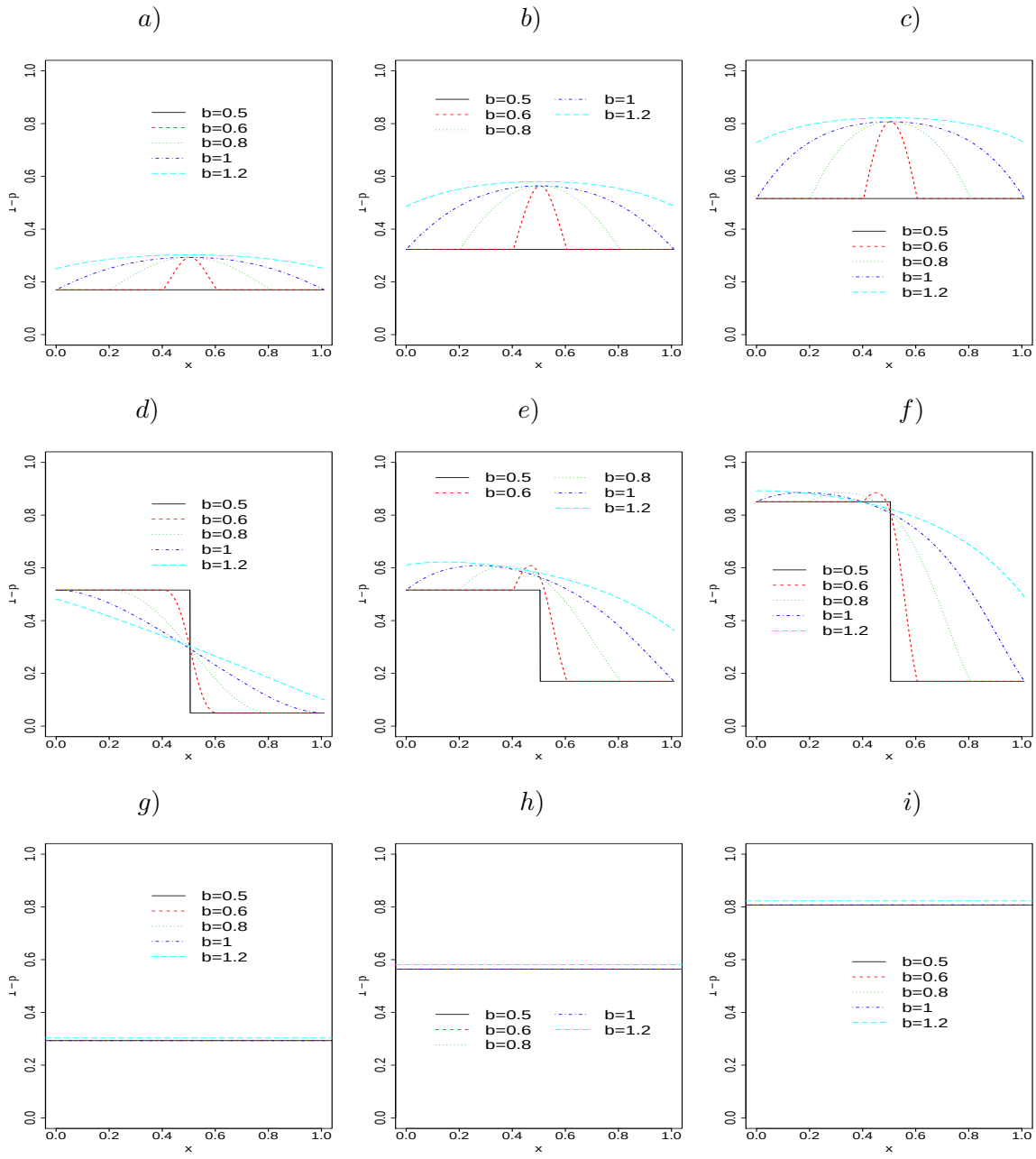


FIG. 9.2: *Puissance en* $x \in [0,1]$ *dans le cas d'un modèle de covariance sphérique de portée* $b \in \{0.05, 0.1, 0.2, 0.4, 0.8\}$, *pour le krigeage simple (deux premières lignes) avec a)* $(a_1, a_2) = (1, -1)$, *b)* $(1.5, -1.5)$ *et c)* $(2, -2)$, *d)* $(a_1, a_2) = (2, 0)$, *e)* $(2, -1)$ *et f)* $(3, -1)$ *et pour le krigrage ordinaire (dernière ligne) avec g)* $(a_1, a_2) = (1, -1)$, *h)* $(1.5, -1.5)$ *et i)* $(2, -2)$.

Chapitre 10

Calcul de la puissance pour un champ spatial sur $\mathbb{R}^2$

Nous nous intéressons maintenant au calcul de la puissance en un point $\mathbf{x}$ du domaine d'étude, $\mathcal{D} \subset \mathbb{R}^2$. Pour cela il est nécessaire de définir l'hypothèse alternative d'existence d'une discontinuité. Une fois l'hypothèse alternative spécifiée, nous pouvons calculer la puissance en $\mathbf{x} \in \mathcal{D}$ en tenant compte de l'information dans un voisinage de $\mathbf{x}$. Dans un premier temps nous calculons la puissance ponctuelle en ne tenant compte que de l'information en $\mathbf{x}$, puis la puissance conditionnelle locale pour laquelle nous conditionnons par rapport à la valeur de la statistique de test aux pixels définissant une fenêtre centrée en $\mathbf{x}$.

10.1 Spécification de l'hypothèse alternative

Les Zones de Changement Abrupt peuvent avoir une géométrie complexe dans le plan. La méthode de détection de ZCAs laisse une certaine liberté sur leur forme. Afin de pouvoir spécifier l'hypothèse alternative, nous allons supposer que les courbes de discontinuités Γ sont assez régulières et qu'elles ne s'intersectent pas.

L'hypothèse de régularité de Γ permet de supposer que dans un petit voisinage de $\mathbf{x}$, la courbe de discontinuité est une droite car elle peut être localement approchée par sa tangente en $\mathbf{x}$. Ainsi nous définissons la discontinuité par une droite passant par $\mathbf{x}$ telle que de part et d'autre le champ aléatoire ait pour moyenne $\pm a/2$, où a est une constante (figure 10.1). Les hypothèses sont donc les suivantes :

$H_0(\mathbf{x})$: $Z(\cdot)$ est stationnaire,

$H_1(\mathbf{x})$: $\mathbb{E}[Z(\cdot)]$ présente une discontinuité représentée par une droite passant par $\mathbf{x}$.

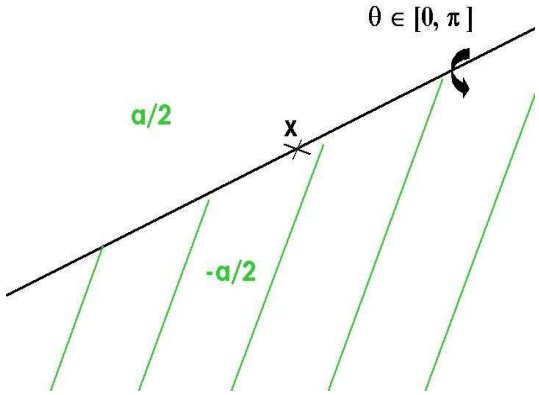


FIG. 10.1: *Modèle utilisé pour l'alternative dans le calcul de la puissance en* $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^2$.

Sous l'alternative, le gradient $W(\mathbf{x})$ peut s'écrire :

$$W(\mathbf{x}) = W_{H_0}(\mathbf{x}) + k_a(\mathbf{x}), \text{ avec } k_a(\mathbf{x}) = \partial C'(\mathbf{x})\mathbf{C}^{-1}A(\mathbf{x}),$$

où $W_{H_0}(\mathbf{x}) = \partial C'(\mathbf{x})\mathbf{C}^{-1}\mathbf{Z}$ est le gradient sous l'hypothèse de stationnarité et $A(\mathbf{x})$ est un vecteur de dimension n, d'éléments $\pm a/2$ selon que le point de donnée est d'un côté ou de l'autre de la discontinuité.

Comme les tests de détection de changement abrupts ne sont pas indépendants et que la puissance doit tenir compte de toute l'information du domaine, le calcul de la puissance est complexe. Aussi nous allons calculer la puissance en un point $\mathbf{x}$ où $\mathbf{x}$ est le centre d'une fenêtre $\mathcal{F}_k \subseteq \mathcal{D}$. Cela permet d'obtenir une minoration de la puissance $1-\beta$. Pour une taille croissante de fenêtre, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_k \subseteq \cdots \subseteq \mathcal{D}$, nous avons :

$$1-\beta(\mathcal{F}_0) \leq 1-\beta(\mathcal{F}_1) \leq \cdots \leq 1-\beta(\mathcal{F}_k) \leq \cdots \leq 1-\beta. \tag{10.1}$$

En effet, plus la fenêtre est grande, plus il y a d'information prise en compte dans le calcul de la puissance. En effet, notons $H_{1,\mathcal{F}_k}(\mathbf{x})$ l'hypothèse alternative en considérant une fenêtre $\mathcal{F}_k$: $\mathbb{E}[Z(\cdot)]$ présente une discontinuité représentée par une droite passant par $\mathbf{x}$, où $\mathbf{x}$ est le centre d'une fenêtre $\mathcal{F}_k$ contenant $n_{\mathcal{F}_k}$ points (*i.e.* $n_{\mathcal{F}_k}$ pixels). La puissance en $\mathbf{x}$ est la probabilité de rejeter l'hypothèse nulle de stationnarité sous $H_{1,\mathcal{F}_k}(\mathbf{x})$, ce qui revient à dire qu'il existe au moins un point de $\mathcal{F}_k$, $\mathbf{x}_i$, tel que $T(\mathbf{x}_i) > t_\alpha$. Aussi, la

puissance s'écrit :

$$\begin{aligned} 1-\beta(\mathcal{F}_k) &= \mathbb{P}_{H_{1,\mathcal{F}_k}(\mathbf{x})}\left[\cup_{i=1}^{n_{\mathcal{F}_k}} (\text{rejeter } H_0(\mathbf{x}_i))\right] \\ &= 1-\mathbb{P}_{H_{1,\mathcal{F}_k}(\mathbf{x})}\left[\cap_{i=1}^{n_{\mathcal{F}_k}} (\text{accepter } H_0(\mathbf{x}_i))\right] \\ &= 1-\mathbb{P}_{H_{1,\mathcal{F}_k}(\mathbf{x})}\left[\cap_{i=1}^{n_{\mathcal{F}_k}} (T(\mathbf{x}_i)<t_\alpha)\right] \\ &= 1-\mathbb{P}_{H_{1,\mathcal{F}_k}(\mathbf{x})}\left[T(\mathbf{x}_1)<t_\alpha,\ldots,T(\mathbf{x}_{n_{\mathcal{F}_k}})<t_\alpha\right]. \end{aligned}$$

Donc plus la fenêtre est grande plus la contrainte "$T(\mathbf{x}_1) < t_\alpha, \ldots, T(\mathbf{x}_{n_{\mathcal{F}_k}}) < t_\alpha$" est forte et plus la puissance est élevée. Nous discuterons à la fin de cette partie de la taille optimale de la fenêtre pour laquelle nous avons une bonne approximation.

Dans ce qui précède nous avons supposé que la discontinuité pouvait être localement approchée par sa tangente. Comme nous ne connaissons pas l'orientation de la discontinuité, nous utilisons la formule des probabilités totales pour calculer la puissance en $\mathbf{x}$. L'hypothèse alternative est fonction de l'orientation de la discontinuité : $H_1(\mathbf{x}) = H_1(\mathbf{x}, \theta)$ est l'existence d'une discontinuité linéaire passant par $\mathbf{x}$ et d'angle θ (figure 10.1). Nous supposons une distribution uniforme de l'orientation de la discontinuité. La puissance en $\mathbf{x}$ s'écrit alors :

$$1-\beta(\mathcal{F}_k) = \frac{1}{\pi}\int_0^\pi \left[1-\beta(\mathcal{F}_k(\theta))\right] \mathrm{d}\theta = \frac{1}{\pi}\int_0^\pi \mathbb{P}_{H_1(\mathbf{x},\theta)}[\text{rejeter } H_0(\mathbf{x})]\,\mathrm{d}\theta.$$

10.2 Puissance ponctuelle

Afin de calculer la puissance sur une fenêtre minimale, *i.e.* $\mathcal{F}_0 = \{\mathbf{x}\}$, nous allons utiliser le fait que la statistique $T(\mathbf{x})$ se décompose en la somme du carré de deux champs gaussiens $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$, non indépendants, non stationnaires, de variance unité, définis dans la proposition 3. Ces champs sont centrés en l'absence de discontinuité. En revanche, en présence d'une discontinuité leur espérance est fonction de $k_a(\mathbf{x})$ et est notée $\mu_i(\mathbf{x}; a, \theta)$, $i = 1, 2$.
La puissance en $\mathbf{x}$ est définie par : $1-\beta(\mathcal{F}_0) = \frac{1}{\pi}\int_0^\pi \left[1-\beta(\mathcal{F}_0(\theta))\right] \mathrm{d}\theta$, avec :

$$\begin{aligned} 1-\beta(\mathcal{F}_0(\theta)) &= \mathbb{P}\left[\text{ rejet de } H_0(\mathbf{x}) \mid H_1(\mathbf{x},\theta) \text{ vraie }\right] \\ &= \mathbb{P}\left[T(\mathbf{x}) > t_\alpha \mid \exists \text{ une discontinuité passant par } \mathbf{x},\ \text{d'angle } \theta\right] \\ &= \mathbb{P}\left[T(\mathbf{x}) > t_\alpha \mid\ W(\mathbf{x}) = W_{H_0}(\mathbf{x}) + k_a(\mathbf{x})\right]. \end{aligned} \tag{10.2}$$

Proposition 12 *Le taux de détection d'une discontinuité en un point* $\mathbf{x} \in \mathcal{D}$ *tenant compte de l'information contenue dans la fenêtre* $\mathcal{F}_0 = \{\mathbf{x}\}$ *est :*

$$1 - \beta(\mathcal{F}_0) = \frac{1}{\pi} \int_0^{\pi} \mathbb{P}_{H_1(\mathbf{x},\theta)}[T(\mathbf{x}) \geq t_\alpha]\, d\theta, \text{ avec } \mathbb{P}_{H_1(\mathbf{x},\theta)}\left[T(\mathbf{x}) \geq t_\alpha\right] = 1 - F_{\chi^2(2,\,\mu_\theta(\mathbf{x}))}(t_\alpha),$$

où $F_{\chi^2(2,\,\mu_\theta(\mathbf{x}))}$ *est la fonction de répartition d'une* χ^2 *décentrée à deux degrés de liberté et de paramètre de décentrage* $\mu_\theta(\mathbf{x}) = \sum_{i=1}^{2} \mu_i^2(\mathbf{x}; a, \theta)$.

Preuve : $1 - \beta(\mathcal{F}_0) = \frac{1}{\pi} \int_0^{\pi} \left[1 - \beta(\mathcal{F}_0(\theta))\right] \mathrm{d}\theta$, avec :

$$\begin{aligned}
1 - \beta(\mathcal{F}_0(\theta)) &= \mathbb{P}\left[U_1^2(\mathbf{x}) + U_2^2(\mathbf{x}) > t_\alpha \mid W_i(\mathbf{x}) = W_{i,H_0}(\mathbf{x}) + k_{a,i}(\mathbf{x}), i = 1, 2\right] \text{ d'après l'équation (10.2)} \\
&= \mathbb{P}_{H_1(\mathbf{x},\theta)}\left[U_1^2(\mathbf{x}) + U_2^2(\mathbf{x}) > t_\alpha\right], \\
&\quad \text{avec } U_1(\mathbf{x}) = \frac{W_{1,H_0}(\mathbf{x})}{\sigma_1(\mathbf{x})} + \mu_1(\mathbf{x}; a, \theta), \\
&\qquad U_2(\mathbf{x}) = \left\{ \frac{W_{2,H_0}(\mathbf{x})}{\sigma_2(\mathbf{x})} - \rho(\mathbf{x}) \frac{W_{1,H_0}(\mathbf{x})}{\sigma_1(\mathbf{x})} \right\} / \sqrt{1 - \rho^2(\mathbf{x})} + \mu_2(\mathbf{x}; a, \theta), \\
&\qquad \mu_1(\mathbf{x}; a, \theta) = \frac{k_{a,1}(\mathbf{x})}{\sigma_1(\mathbf{x})} \\
&\qquad \text{et } \mu_2(\mathbf{x}; a, \theta) = \left\{ \frac{k_{a,2}(\mathbf{x})}{\sigma_2(\mathbf{x})} - \rho(\mathbf{x}) \frac{k_{a,1}(\mathbf{x})}{\sigma_1(\mathbf{x})} \right\} / \sqrt{1 - \rho^2(\mathbf{x})}.
\end{aligned}$$

Nous avons sous $H_0(\mathbf{x})$: $U_i(\mathbf{x}) \sim \mathcal{N}(0, 1)$, ainsi sous $H_1(\mathbf{x})$: $U_i(\mathbf{x}) \sim \mathcal{N}(\mu_i(\mathbf{x}; a, \theta), 1)$.

$$\begin{aligned}
1 - \beta(\mathcal{F}_0(\theta)) &= \mathbb{P}_{H_1(\mathbf{x},\theta)}\left[U_1^2(\mathbf{x}) + U_2^2(\mathbf{x}) > t_\alpha\right], \text{ avec } U_i(\mathbf{x}) \sim \mathcal{N}(\mu_i(\mathbf{x}; a, \theta), 1) \\
&= 1 - \mathbb{P}_{H_1(\mathbf{x},\theta)}\left[U_1^2(\mathbf{x}) + U_2^2(\mathbf{x}) < t_\alpha\right], \text{ avec } U_i(\mathbf{x}) \sim \mathcal{N}(\mu_i(\mathbf{x}; a, \theta), 1) \\
&= 1 - F_{\chi^2(2,\,\mu_\theta(\mathbf{x}))}(t_\alpha), \text{ avec } \mu_\theta(\mathbf{x}) = \sum_{i=1,2} \mu_i^2(\mathbf{x}; a, \theta).
\end{aligned}$$

□

En pratique, la puissance en un point $\mathbf{x}$ sera plus simplement calculée comme étant la moyenne des puissances en $\mathbf{x}$ pour n_θ directions :

$$1 - \beta(\mathcal{F}_0) = \frac{1}{n_\theta} \sum_{p=1}^{n_\theta} \mathbb{P}_{H_1(\mathbf{x},\theta_p)}\left[T(\mathbf{x}) > t_\alpha\right].$$

Nous prendrons en particulier $n_\theta = 4$ et $\theta \in \{0, \pi/4, \pi/2, 3\pi/4\}$ pour l'angle formé avec l'axe des abscisses. Comme nous allons le voir dans l'illustration suivante, ces 4 directions suffisent pour avoir une bonne approximation de la puissance.

Illustration

Nous avons simulé 1000 échantillons de 100 points répartis aléatoirement dans un carré de côté un. Ces points sont issus d'un champ aléatoire gaussien standard de fonction de covariance exponentielle, $\exp(-\|\mathbf{h}\|/b)$, de portée $b = 0.1$. Une discontinuité est introduite le long de la droite $y = 0.5$ et nous considérons plusieurs valeurs de la discontinuité : $a \in \{0, 2, 3, 4\}$. La figure 10.2*a* est un exemple d'échantillon simulé où la discontinuité vaut $a = 2.5$. Nous avons calculé pour chacune de ces simulations la puissance le long de la discontinuité, en utilisant le niveau $1 - \alpha = 0.999$ et une grille 60×60. La figure 10.2*b* représente la moyenne des puissances le long de la discontinuité pour chaque valeur de a.

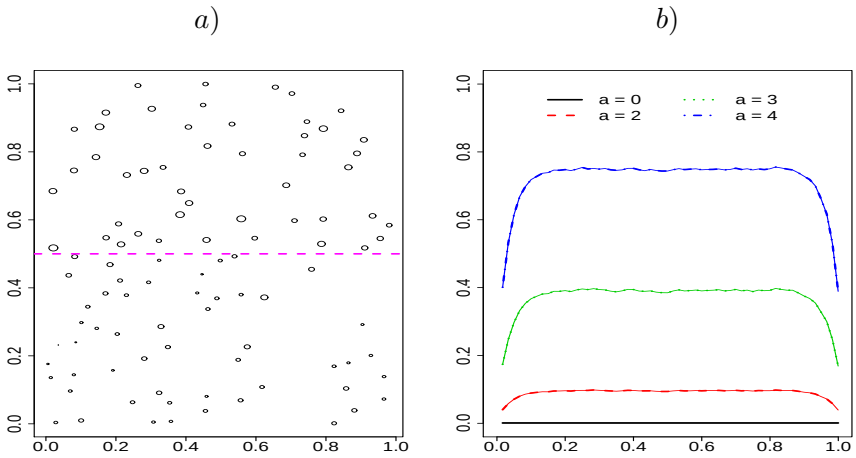


FIG. 10.2: *a) Exemple d'échantillon de 100 points répartis aléatoirement dans un carré de côté un, issus d'un champ gaussien standard de covariance exponentielle de portée $b = 0.1$. Une discontinuité est introduite en $y = 0.5$. b) Puissance moyenne en un point sur 1000 échantillons du type a), le long de la discontinuité, pour $a \in \{0, 2, 3, 4\}$.*

La puissance se révèle constante le long de la discontinuité aux effets de bord près. Elle est de l'ordre de 0.09, 0.38, 0.75 pour respectivement $a = 2, 3, 4$.

Nous avons calculé la puissance au pixel central de la grille 60×60, pour 100 simulations du même type que ce qui précède, en considérant $1 - \alpha = 1 - \hat{\alpha}$ et en faisant varier le nombre de directions, $n_\theta \in \{1, 2, \ldots, 32\}$, et la valeur de la discontinuité $a \in \{0, 2, 3, 4\}$. Les résultats sont représentés sur la figure 10.3. Cette figure montre que la puissance est quasi constante au delà de 4 directions de la discontinuité. La puissance reste proche de la droite en pointillés correspondant à la valeur de la puissance pour $n_\theta = 4$.

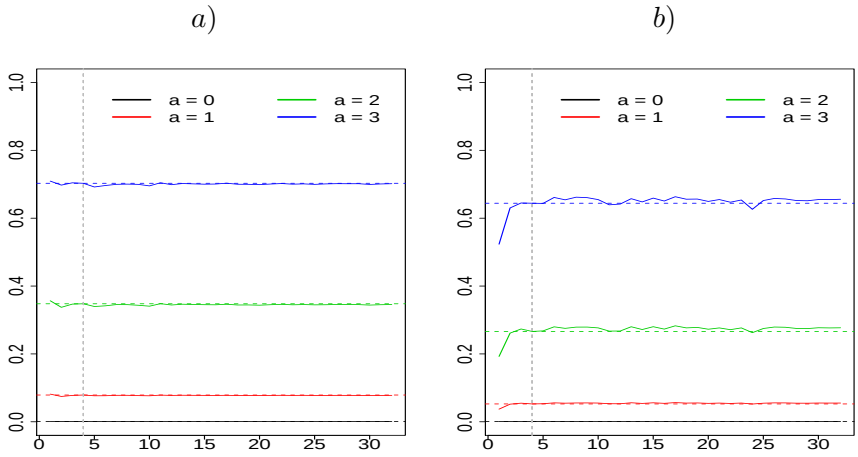


FIG. 10.3: *Trait plein : puissance en fonction du nombre de directions et de la valeur de la discontinuité. Les échantillons considérés sont constitués de 100 points répartis aléatoirement dans un carré de côté un, issus d'un champ gaussien standard de covariance exponentielle de portée* $b = 0.1$. *Pointillés : les droites horizontales correspondent à la valeur de la puissance pour 4 directions, et la droite verticale à* $n_\theta = 4$. *a) moyenne de 100 échantillons, b) pour un échantillon.*

Avant de calculer la puissance en un point $\mathbf{x}$ connaissant l'information sur une fenêtre $\mathcal{F}_k$, nous allons nous intéresser à la puissance en deux points $\mathbf{x}$ et $\mathbf{y}$ sachant que la discontinuité passe par ces deux points. Cela va en particulier nous permettre de montrer que nous ne pouvons pas considérer l'indépendance entre les tests locaux.

10.3 Non indépendance des tests locaux

Dans cette section nous considérons l'hypothèse alternative suivante, $H_1(\mathbf{x}, \mathbf{y})$: il existe une discontinuité représentée par une droite passant par deux points $\mathbf{x}$ et $\mathbf{y}$. Notons $H_0(\mathbf{x})$ l'hypothèse nulle définie pour le calcul de la puissance en un point $\mathbf{x}$. Ici la discontinuité est représentée par une droite passant par $\mathbf{x}$ et $\mathbf{y}$ telle que de part et d'autre de cette droite le champ $Z(\cdot)$ ait pour moyenne $\pm a/2$. La puissance s'écrit alors :

$$\begin{aligned} 1 - \beta(\mathbf{x}, \mathbf{y}) &= \mathbb{P}_{H_1(\mathbf{x},\mathbf{y})}\left[\text{rejeter } H_0\right] \\ &= 1 - \mathbb{P}_{H_1(\mathbf{x},\mathbf{y})}\left[\text{ non rejet de } H_0(\mathbf{x}) \text{ et de } H_0(\mathbf{y})\right] \\ &= 1 - \mathbb{P}_{H_1(\mathbf{x},\mathbf{y})}\left[T(\mathbf{x}) < t_\alpha, T(\mathbf{y}) < t_\alpha\right]. \end{aligned} \tag{10.3}$$

Proposition 13 *Le taux de détection en* $\mathbf{x}$ *et* $\mathbf{y}$ *d'une droite passant par* $\mathbf{x}$ *et* $\mathbf{y}$ *est :*

$$1 - \beta(\mathbf{x}, \mathbf{y}) = 1 - \frac{1}{4} \iint_0^{t_\alpha} \iint_0^{2\pi} f_{\mathbf{V}}(\mathbf{v})\, d\theta_{\mathbf{x}}\, d\theta_{\mathbf{y}}\, dt_{\mathbf{x}}\, dt_{\mathbf{y}},$$

avec $\mathbf{V} = (U_{H_1,1}(\mathbf{x}), U_{H_1,2}(\mathbf{x}), U_{H_1,1}(\mathbf{y}), U_{H_1,2}(\mathbf{y}))'$ *où* $U_{H_1,i}(\mathbf{x}) \sim \mathcal{N}(\mu_i(\mathbf{x}; a), 1)$
et $\mathbf{v} = (\sqrt{t_\mathbf{x}}\cos\theta_\mathbf{x}, \sqrt{t_\mathbf{x}}\sin\theta_\mathbf{x}, \sqrt{t_\mathbf{y}}\cos\theta_\mathbf{y}, \sqrt{t_\mathbf{y}}\sin\theta_\mathbf{y})'$.

Preuve : La puissance en $\mathbf{x}$ et $\mathbf{y}$ s'écrit,

$$\begin{aligned}
1-\beta(\mathbf{x},\mathbf{y}) &= 1 - \iint_0^{t_\alpha} \mathbb{P}_{H_1}\left[T(\mathbf{x}) = t_\mathbf{x}, T(\mathbf{y}) = t_\mathbf{y}\right] \mathrm{d}t_\mathbf{x}\,\mathrm{d}t_\mathbf{y}, \text{ d'après (10.3)} \\
&= 1 - \iint_0^{t_\alpha} \mathbb{P}_{H_1}\left[U_1^2(\mathbf{x}) + U_2^2(\mathbf{x}) = t_\mathbf{x}, U_1^2(\mathbf{y}) + U_2^2(\mathbf{y}) = t_\mathbf{y}\right] \mathrm{d}t_\mathbf{x}\,\mathrm{d}t_\mathbf{y} \\
&= 1 - \frac{1}{4}\iint_0^{t_\alpha}\iint_0^{2\pi} \mathbb{P}_{H_1}\left[U_1(\mathbf{x}) = \sqrt{t_\mathbf{x}}\cos\theta_\mathbf{x}, U_2(\mathbf{x}) = \sqrt{t_\mathbf{x}}\sin\theta_\mathbf{x},\right. \\
&\qquad\qquad \left. U_1(\mathbf{y}) = \sqrt{t_\mathbf{y}}\cos\theta_\mathbf{y}, U_2(\mathbf{y}) = \sqrt{t_\mathbf{y}}\sin\theta_\mathbf{y}\right] \mathrm{d}\theta_\mathbf{x}\,\mathrm{d}\theta_\mathbf{y}\,\mathrm{d}t_\mathbf{x}\,\mathrm{d}t_\mathbf{y}
\end{aligned}$$

d'après l'équation paramétrique d'un cercle (10.5). Les paramètres $\theta_\mathbf{x},\ \theta_\mathbf{y}, t_\mathbf{x},$ et $t_\mathbf{y}$ varient de façon uniforme et sont indépendants des orientations.

$$\begin{aligned}
&= 1 - \frac{1}{4}\iint_0^{t_\alpha}\iint_0^{2\pi} f_\mathbf{V}(\mathbf{v})\,\mathrm{d}\theta_\mathbf{x}\,\mathrm{d}\theta_\mathbf{y}\,\mathrm{d}t_\mathbf{x}\,\mathrm{d}t_\mathbf{y}, \\
&\quad \text{avec } \mathbf{V} = (U_1(\mathbf{x}), U_2(\mathbf{x}), U_1(\mathbf{y}), U_2(\mathbf{y}))'_{H_1} \\
&\quad \text{et } \mathbf{v} = (\sqrt{t_\mathbf{x}}\cos\theta_\mathbf{x}, \sqrt{t_\mathbf{x}}\sin\theta_\mathbf{x}, \sqrt{t_\mathbf{y}}\cos\theta_\mathbf{y}, \sqrt{t_\mathbf{y}}\sin\theta_\mathbf{y})'.
\end{aligned}$$

Comme sous l'alternative $U_i(\mathbf{x}) \sim \mathcal{N}(\mu_i(\mathbf{x}; a), 1)$, la densité de $\mathbf{V}$ a la forme suivante,

$$f_\mathbf{V}(\mathbf{v}) = \frac{1}{(2\pi)^2 \det\left(\mathbf{\Sigma}_\mathbf{V}\right)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{v} - m_\mathbf{V})'\mathbf{\Sigma}_\mathbf{V}^{-1}(\mathbf{v} - m_\mathbf{V})\right\},$$

avec $m_\mathbf{V} = (\mu_1(\mathbf{x}; a), \mu_2(\mathbf{x}; a), \mu_1(\mathbf{y}; a), \mu_2(\mathbf{y}; a))'$ et pour :

$$\left.\begin{aligned}
c_{11}(\mathbf{x},\mathbf{y}) &= \rho_{11}(\mathbf{x},\mathbf{y}) \\
c_{12}(\mathbf{x},\mathbf{y}) &= \{\rho_{12}(\mathbf{x},\mathbf{y}) - \rho(\mathbf{y})\rho_{11}(\mathbf{x},\mathbf{y})\}/\sqrt{1-\rho^2(\mathbf{y})} \\
c_{21}(\mathbf{x},\mathbf{y}) &= \{\rho_{21}(\mathbf{x},\mathbf{y}) - \rho(\mathbf{x})\rho_{11}(\mathbf{x},\mathbf{y})\}/\sqrt{1-\rho^2(\mathbf{x})} \\
c_{22}(\mathbf{x},\mathbf{y}) &= \{\rho_{22}(\mathbf{x},\mathbf{y}) - \rho(\mathbf{y})\rho_{21}(\mathbf{x},\mathbf{y}) - \rho(\mathbf{x})\rho_{12}(\mathbf{x},\mathbf{y}) \\
&\quad +\rho(\mathbf{x})\rho(\mathbf{y})\rho_{11}(\mathbf{x},\mathbf{y})\}\,/\sqrt{(1-\rho^2(\mathbf{x}))(1-\rho^2(\mathbf{y}))} \\
c_{ij}(\mathbf{x},\mathbf{y}) &= c_{ji}(\mathbf{y},\mathbf{x}),\ c_{ii}(\mathbf{x},\mathbf{x}) = 1 \text{ et } c_{12}(\mathbf{x},\mathbf{x}) = 0,
\end{aligned}\right\} \tag{10.4}$$

nous avons : $\mathbf{\Sigma}_\mathbf{V} = \begin{pmatrix} 1 & 0 & c_{11}(\mathbf{x},\mathbf{y}) & c_{12}(\mathbf{x},\mathbf{y}) \\ 0 & 1 & c_{21}(\mathbf{x},\mathbf{y}) & c_{22}(\mathbf{x},\mathbf{y}) \\ c_{11}(\mathbf{x},\mathbf{y}) & c_{21}(\mathbf{x},\mathbf{y}) & 1 & 0 \\ c_{12}(\mathbf{x},\mathbf{y}) & c_{22}(\mathbf{x},\mathbf{y}) & 0 & 1 \end{pmatrix}$. □

En pratique, la puissance en $\mathbf{x}$ et $\mathbf{y}$ peut être calculée via l'approximation numérique :

$$
\begin{aligned}
&1-\beta(\mathbf{x},\mathbf{y}) \\
&\approx 1-\tfrac{1}{4}\frac{1}{(2\pi)^2\det(\mathbf{\Sigma_V})^{1/2}}\frac{(2\pi)^2 t_\alpha^2}{n_{\theta_\mathbf{x}} n_{\theta_\mathbf{y}} n_{t_\mathbf{x}} n_{t_\mathbf{y}}}\sum_{\theta_\mathbf{x}}\sum_{\theta_\mathbf{y}}\sum_{t_\mathbf{x}}\sum_{t_\mathbf{y}}\exp\left\{-\tfrac{1}{2}(\mathbf{v}-m_\mathbf{V})'\mathbf{\Sigma_V}^{-1}(\mathbf{v}-m_\mathbf{V})\right\} \\
&= \quad 1-\frac{t_\alpha^2}{4\det(\mathbf{\Sigma_V})^{1/2} n_{\theta_\mathbf{x}} n_{\theta_\mathbf{y}} n_{t_\mathbf{x}} n_{t_\mathbf{y}}}\sum_{\theta_\mathbf{x}}\sum_{\theta_\mathbf{y}}\sum_{t_\mathbf{x}}\sum_{t_\mathbf{y}}\exp\left\{-\tfrac{1}{2}(\mathbf{v}-m_\mathbf{V})'\mathbf{\Sigma_V}^{-1}(\mathbf{v}-m_\mathbf{V})\right\},
\end{aligned}
$$

où $n_{\theta_\mathbf{x}}$, $n_{\theta_\mathbf{y}}$, $n_{t_\mathbf{x}}$ et $n_{t_\mathbf{y}}$ représentent respectivement les dimensions des vecteurs de paramètres $\theta_\mathbf{x}$, $\theta_\mathbf{y}$, $t_\mathbf{x}$ et $t_\mathbf{y}$.
Nous pouvons également calculer la puissance en $\mathbf{x}$ et $\mathbf{y}$ via une approximation du type Monte-Carlo. Nous générons n_e vecteurs multigaussiens $\mathbf{V}$, de moyenne $m_\mathbf{V}$ et de matrice de covariance $\mathbf{\Sigma_V}$ et nous dénombrons les fois où $T(\mathbf{x}) < t_\alpha$ et $T(\mathbf{y}) < t_\alpha$. Cela se traduit par l'approximation,

$$
\begin{aligned}
1-\beta(\mathbf{x},\mathbf{y}) &= 1-\mathbb{P}_{H_1(\mathbf{x},\mathbf{y})}\left[T(\mathbf{x})<t_\alpha, T(\mathbf{y})<t_\alpha\right] \\
&\approx 1-\frac{1}{n_e}\sum_{i=1}^{n_e}\mathbf{1}_{\{T(\mathbf{x})<t_\alpha\}}\mathbf{1}_{\{T(\mathbf{y})<t_\alpha\}}.
\end{aligned}
$$

Les deux approximations conduisent aux mêmes résultats. Cependant, pour des questions de rapidité de calculs nous utiliserons par la suite l'approximation du type Monte-Carlo.

Illustration

Nous avons simulé 1000 échantillons de 100 points répartis aléatoirement dans un carré de côté un, issus d'un champ aléatoire gaussien centré-réduit de fonction de covariance exponentielle de portée 0.1. Une discontinuité a été introduite en $y = 0.5$ (figure 10.4*a*). De part et d'autre de cette droite, le champ a pour moyenne $\pm a/2$, $a \in \{0, 2, 3, 4\}$. L'erreur de type I, α, du test de détection de ruptures est fixée à $\alpha = 0.001$. Nous avons calculé la puissance en deux points de la discontinuité pour différentes distances et pour les différentes valeurs de a. Les couples de points pour lesquels la puissance est calculée sont placés symétriquement autour du centre du domaine (figure 10.4*a*). La figure 10.4*b* représente la puissance (moyenne prise sur les 1000 échantillons) en deux points en fonction de la distance entre les points (trait plein) et de la discontinuité. Lorsque la distance tend vers zéro, la puissance entre les deux points diminue et tend (par valeurs supérieures) vers la puissance au centre du domaine. Cela s'explique par le fait que plus les points sont proches, plus ils sont corrélés. Les courbes en pointillés représentent la puissance que nous obtiendrions si les tests étaient indépendants. Dans ce cas la puissance en deux

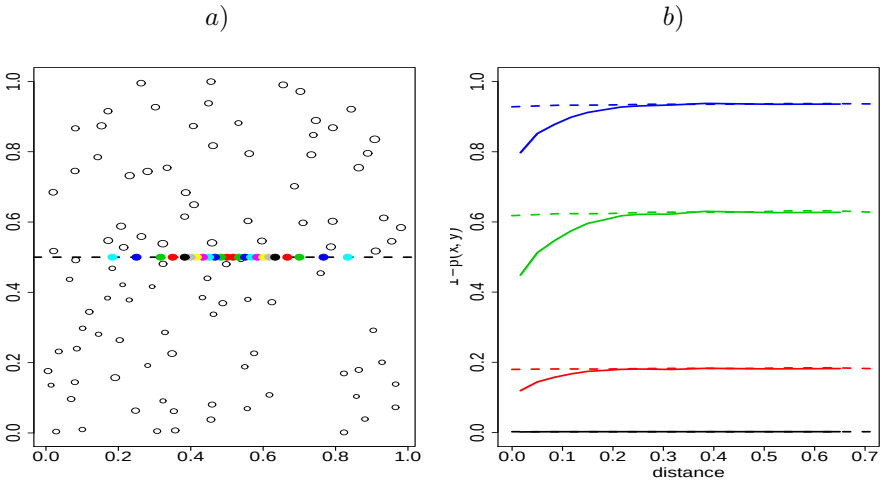


FIG. 10.4: *a) Discontinuité et couples de points pour lesquels la puissance en deux points est calculée, avec un modèle de covariance exponentiel de portée* $b = 0.1$ *b) Trait plein : puissance moyenne en deux points sur 1000 échantillons. Pointillés : puissance en deux points en supposant les tests indépendants.* $a = 0$ *(noir),* $a = 2$ *(rouge),* $a = 3$ *(vert) et* $a = 4$ *(bleu).*

points s'écrirait,

$$\begin{aligned} 1 - \beta(\mathbf{x}, \mathbf{y}) &= 1 - \mathbb{P}_{H_1(\mathbf{x},\mathbf{y})}\left[T(\mathbf{x}) < t_\alpha, T(\mathbf{y}) < t_\alpha\right] \\ &= 1 - \mathbb{P}_{H_1(\mathbf{x},\mathbf{y})}\left[T(\mathbf{x}) < t_\alpha\right] \mathbb{P}\left[T(\mathbf{y}) < t_\alpha\right] \\ &= 1 - \beta(\mathbf{x})\beta(\mathbf{y}) \end{aligned}$$

Pour chaque valeur de a, l'écart entre la courbe en trait plein et la courbe en pointillés indique l'erreur que nous aurions commise si nous avions supposé l'indépendance des tests locaux. Cette erreur est non négligeable lorsque la distance est inférieure à 0.3, et le devient lorsque la distance est supérieure à 0.3. Au delà de cette distance les tests peuvent être considérés comme indépendants. Cette distance correspond à la portée pratique du modèle de covariance. En effet pour un modèle de covariance exponentiel de portée b, la portée pratique est $3b$ (section 3.3).

10.4 Puissance conditionnelle locale

Considérons maintenant que $\mathbf{x}$ est le centre d'une fenêtre quelconque, $\mathcal{F}$, de $\mathcal{D}$ (figure 10.5). Dire que l'hypothèse nulle est rejetée revient à dire qu'il existe une discontinuité qui passe par au moins un des points de $\mathcal{F}$, *i.e.* l'hypothèse nulle de la puissance (ponctuelle) en ce point est rejetée.

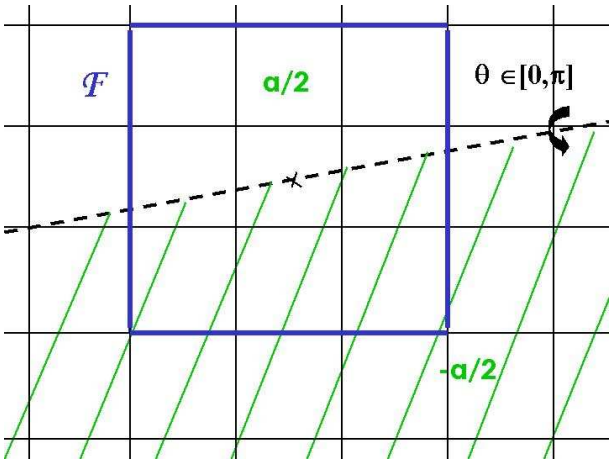


FIG. 10.5: *Modèle de discontinuité utilisé pour l'alternative dans le calcul de la puissance conditionnelle locale.*

Le calcul de la puissance en $\mathbf{x}$, centre d'une fenêtre $\mathcal{F}$, utilise la forme paramétrique de l'équation d'un cercle :

$$U_1^2(\mathbf{x}) + U_2^2(\mathbf{x}) = t_{\mathbf{x}} \text{ s'écrit } U_1(\mathbf{x}) = \sqrt{t_{\mathbf{x}}}\cos(\omega_{\mathbf{x}}), U_2(\mathbf{x}) = \sqrt{t_{\mathbf{x}}}\sin(\omega_{\mathbf{x}}). \tag{10.5}$$

Notons $n_{\mathcal{F}}$ le nombre de pixels contenus dans la fenêtre $\mathcal{F}$. La puissance au centre $\mathbf{x}$ de la fenêtre $\mathcal{F}$ est définie par : $1-\beta(\mathcal{F}) = \frac{1}{\pi}\int_0^\pi [1-\beta(\mathcal{F}(\theta))]\, \mathrm{d}\theta$, avec :

$$\begin{aligned} 1-\beta(\mathcal{F}(\theta)) &= \mathbb{P}_{H_1(\theta)}\left[\cup_{i=1}^{n_{\mathcal{F}}} (\text{rejeter } H_0(\mathbf{x}_i))\right] \qquad (10.6)\\ &\quad \text{où } H_0(\mathbf{x}_i) \text{ est l'hypothèse locale nulle,} \\ &= 1-\mathbb{P}_{H_1(\theta)}\left[\cap_{i=1}^{n_{\mathcal{F}}} (\text{accepter } H_0(\mathbf{x}_i))\right] \\ &= 1-\mathbb{P}_{H_1(\theta)}\left[\cap_{i=1}^{n_{\mathcal{F}}} (T(\mathbf{x}_i) < t_\alpha)\right] \\ &= 1-\mathbb{P}_{H_1(\theta)}\left[T(\mathbf{x}_1) < t_\alpha, ..., T(\mathbf{x}_{n_{\mathcal{F}}}) < t_\alpha\right]. \qquad (10.7) \end{aligned}$$

Proposition 14 *Le taux de détection d'une discontinuité en un point $\mathbf{x} \in \mathcal{D}$ tenant compte de l'information contenue dans une fenêtre $\mathcal{F} \subseteq \mathcal{D}$ est,*

$$1-\beta(\mathcal{F}) = \frac{1}{\pi}\int_0^\pi \mathbb{P}_{H_1(\theta)}[\textit{rejeter } H_0]\, d\theta,$$

avec :

$$\mathbb{P}_{H_1(\theta)}\left[\textit{rejeter } H_0\right] = 1 - \frac{1}{2^{n_{\mathcal{F}}}}\int_{\omega\in[0,2\pi]^{n_{\mathcal{F}}}}\int_{\mathbf{t}\in[\mathbf{t}_\alpha,\infty)^{\mathbf{n}_{\mathcal{F}}}} f_{\mathbf{V}_\theta}(\mathbf{v})\, d\mathbf{t}\, d\omega,$$

où $\mathbf{V}_\theta$ est un vecteur gaussien de longueur $2n_{\mathcal{F}}$ constitué des champs U_1 et U_2 en $\mathbf{x}_1, ..., \mathbf{x}_{n_{\mathcal{F}}}$ sous l'alternative et $\mathbf{t} = (t_{\mathbf{x}_1}, ..., t_{\mathbf{x}_{n_{\mathcal{F}}}})'$, $\omega = (\omega_1, ..., \omega_{n_{\mathcal{F}}})'$ sont des paramètres issus de l'équation (10.5).

Preuve : La puissance s'écrit : $1-\beta(\mathcal{F}) = \frac{1}{\pi}\int_0^\pi [1-\beta(\mathcal{F}(\theta))]\,\mathrm{d}\theta$, avec :

$$1-\beta(\mathcal{F}(\theta)) \;=\; 1-\frac{1}{2^{n_\mathcal{F}}}\int_\theta\int_{\mathbf{t}} f_{\mathbf{V}_\theta}(\mathbf{v})\,\mathrm{d}\theta\,\mathrm{d}\mathbf{t},\ \text{d'après (10.7)}$$

où $\mathbf{V}_\theta = (U_{H_1,1}(\mathbf{x}_1), U_{H_1,2}(\mathbf{x}_1), ..., U_{H_1,1}(\mathbf{x}_{n_\mathcal{F}}), U_{H_1,2}(\mathbf{x}_{n_\mathcal{F}}))'$, $\theta = (\theta_1, ..., \theta_{n_\mathcal{F}})'$,
$\mathbf{v} = \left(\sqrt{t_1}\cos\theta_1, \sqrt{t_1}\sin\theta_1, ..., \sqrt{t_{n_\mathcal{F}}}\cos\theta_{n_\mathcal{F}}, \sqrt{t_{n_\mathcal{F}}}\sin\theta_{n_\mathcal{F}}\right)'$ et $\mathbf{t} = (t_1, ..., t_{n_\mathcal{F}})'$.
La densité de $\mathbf{V}_\theta$, $f_{\mathbf{V}_\theta}(\mathbf{v})$, s'écrit

$$f_{\mathbf{V}_\theta}(\mathbf{v}) = \frac{1}{(2\pi)^{n_\mathcal{F}/2}\det\left(\mathbf{\Sigma}_{\mathbf{V}_\theta}\right)^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{v}-\mathbf{m}_{\mathbf{V}_\theta})'\mathbf{\Sigma}_{\mathbf{V}_\theta}^{-1}(\mathbf{v}-\mathbf{m}_{\mathbf{V}_\theta})\right\},$$

avec $\mathbf{m}_\mathbf{v} = (\mu_1(\mathbf{x}_1;a,\theta), \mu_2(\mathbf{x}_1;a,\theta), ..., \mu_1(\mathbf{x}_{n_\mathcal{F}};a,\theta), \mu_2(\mathbf{x};_{n_\mathcal{F}};a,\theta))'$ et

$$\mathbf{\Sigma}_{\mathbf{V}_\theta} \;=\; \begin{pmatrix} 1 & 0 & c_{11}(\mathbf{x}_1,\mathbf{x}_2) & \dots & c_{11}(\mathbf{x}_1,\mathbf{x}_{n_\mathcal{F}}) & c_{12}(\mathbf{x}_1,\mathbf{x}_{n_\mathcal{F}}) \\ 0 & 1 & c_{21}(\mathbf{x}_1,\mathbf{x}_2) & \dots & c_{21}(\mathbf{x}_1,\mathbf{x}_{n_\mathcal{F}}) & c_{22}(\mathbf{x}_1,\mathbf{x}_{n_\mathcal{F}}) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{11}(\mathbf{x}_1,\mathbf{x}_{n_\mathcal{F}}) & c_{12}(\mathbf{x}_1,\mathbf{x}_{n_\mathcal{F}}) & c_{11}(\mathbf{x}_2,\mathbf{x}_{n_\mathcal{F}}) & \dots & 1 & 0 \\ c_{12}(\mathbf{x}_1,\mathbf{x}_{n_\mathcal{F}}) & c_{22}(\mathbf{x}_1,\mathbf{x}_{n_\mathcal{F}}) & c_{12}(\mathbf{x}_2,\mathbf{x}_{n_\mathcal{F}}) & \dots & 0 & 1 \end{pmatrix},$$

avec $c_{ij}(\mathbf{x}_k,\mathbf{x}_l)$, $i,j\in\{1,2\}$, $k,j\in\{1,\dots,n_\mathcal{F}\}$ définis dans les équations (10.4).

□

De même que pour le calcul de puissance en deux points, $1-\beta(\mathcal{F})$ est calculée en pratique via une approximation de type Monte-Carlo :

$$1-\beta(\mathcal{F}) \approx 1-\frac{1}{n_e}\sum_{i=1}^{n_e}\mathbf{1}_{\{T(\mathbf{x}_1)<t_\alpha\}}...\mathbf{1}_{\left\{T(\mathbf{x}_{n_\mathcal{F}})<t_\alpha\right\}}.$$

Pour cela nous générons n_e vecteurs multi-gaussiens $\mathbf{V}_\theta$, de matrice de covariance $\mathbf{\Sigma}_{\mathbf{V}_\theta}$ et de moyenne $\mathbf{m}_{\mathbf{V}_\theta}$, et nous dénombrons les fois où $T(\mathbf{x}_i) < t_\alpha$, $i\in\{1,...,n_\mathcal{F}\}$.

Nous avons vu qu'en pratique la méthode de détection de ZCAs est appliquée aux nœuds d'une grille régulière superposée sur le domaine d'étude. Par conséquent, nous allons considérer que $\mathcal{F}$ est un carré de côté $2k+1$ pixels, $k\in\mathbb{N}$, de sorte que son centre corresponde au pixel central (figure 10.5).

En pratique le calcul de $1-\beta(\mathcal{F})$ est limité par la matrice $\mathbf{\Sigma}_{\mathbf{V}_\theta}$. En effet, $\mathbf{\Sigma}_{\mathbf{V}_\theta}$ est de dimension $2(2k+1)^2\times 2(2k+1)^2$ et il nous faut connaître tous ses éléments et évaluer son inverse. D'après la table 10.1, $\mathbf{\Sigma}_{\mathbf{V}_\theta}$ devient très rapidement lourde à calculer car sa

dimension croît très vite avec k. Aussi, ne pouvant calculer qu'une approximation de la puissance en $\mathbf{x}$, il sera nécessaire de tester la qualité de l'approximation de la puissance tenant compte de l'information contenue dans une fenêtre $\mathcal{F}_k$, k restant assez petit.

k	1	2	3	4	5
dimension de $\mathcal{F}$	3×3	5×5	7×7	9×9	11×11
dimension de $\mathbf{\Sigma}_{\mathbf{V}_\theta}$	18×18	50×50	98×98	162×162	242×242

k	6	7	8	9
dimension de $\mathcal{F}$	13×13	15×15	17×17	19×19
dimension de $\mathbf{\Sigma}_{\mathbf{V}_\theta}$	338×338	450×450	578×578	722×722

TAB. 10.1: *Dimension de la matrice de covariance $\mathbf{\Sigma}_{\mathbf{V}_\theta}$ en fonction de la taille de la fenêtre.*

Illustration

Nous avons simulé un échantillon de 100 points dans un carré de côté un, issus d'un champ gaussien standard de fonction de covariance exponentielle de portée $b = 0.05$. Ces points ne sont pas répartis aléatoirement sur le domaine : la densité est plus élevée dans la moitié supérieure et est très faible dans la partie droite (figure 10.6*a*). La puissance a

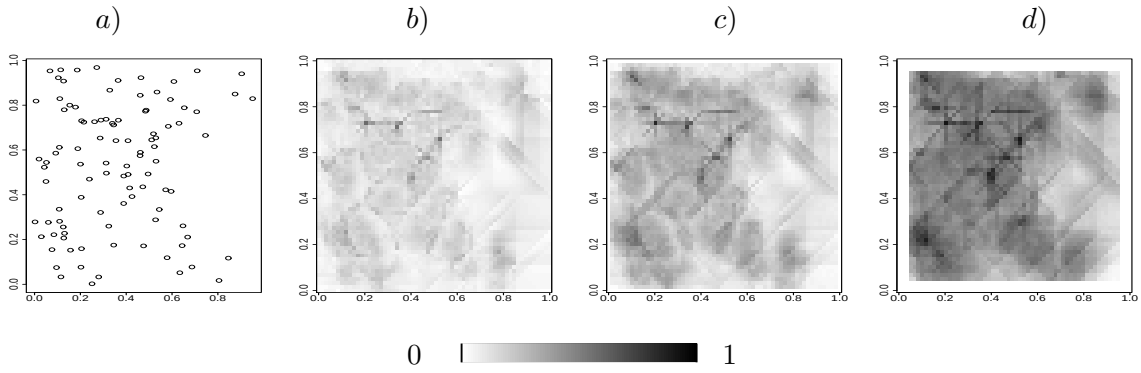


FIG. 10.6: *a) Echantillon de points présentant une densité variable sur le domaine ; b), c), d) Puissance calculée pour une discontinuité $a = 2.5$ et pour différentes valeurs du paramètre k : $k = 0, 1, 3$ respectivement.*

ensuite été calculée en considérant une discontinuité d'une valeur de 2.5 écarts-type et différentes tailles de fenêtre : 1×1, *i.e.* $k = 0$ (figure 10.6*b*), 3×3, *i.e.* $k = 1$ (figure 10.6*c*) et 7×7, *i.e.* $k = 3$ (figure 10.6*d*). Les pixels ne correspondant pas au centre d'une fenêtre de dimension $(2k + 1) \times (2k + 1)$ contenue dans le domaine sont représentés en

blanc (bords des figures 10.6*c* et 10.6*d*). La figure 10.6 illustre la relation (10.1), *i.e.* plus la fenêtre est grande, plus la puissance est élevée. Elle montre également que la puissance de la méthode est liée à la densité locale de l'échantillon. Nous pouvons constater, sur les figures 10.6*b*, 10.6*c* et 10.6*d*, qu'à la région de faible densité d'échantillonnage correspond toujours une faible puissance.

Choix de la taille de la fenêtre

En calculant la puissance en $\mathbf{x}$ en tenant compte de l'information contenue dans un voisinage de $\mathbf{x}$ nous avons obtenu une approximation de la puissance en $\mathbf{x}$. Discutons maintenant de la détermination de la taille de la fenêtre qui rendrait cette approximation optimale.

Sous l'alternative, *i.e.* pour $a \neq 0$, la puissance augmente avec la taille de la fenêtre. Ce phénomène est illustré dans la figure 10.7. Cette figure représente la puissance calculée en le centre d'un carré de côté un, pour un échantillon de 100 points répartis aléatoirement et issus d'un champ gaussien centré-réduit de modèle de covariance exponentiel de portée $b = 0.05$ et pour différentes valeurs de la discontinuité : $a \in \{0, 2, 2.5, 3\}$, sur une grille 60×60. Nous nous attendions *a priori* à ce que la puissance se stabilise à partir d'une certaine

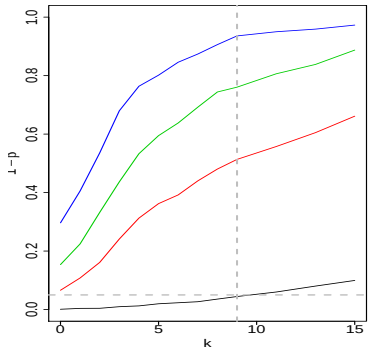


FIG. 10.7: *Courbe de puissance (en noir pour une discontinuité de $a = 0$, en rouge pour $a = 2$, en vert pour $a = 2.5$ et en bleu pour $a = 3$) en fonction du paramètre k (demie longueur de la fenêtre carrée), calculée en le centre d'un carré de côté un dans lequel ont été générés 100 points répartis aléatoirement, issus d'un champ gaussien de covariance exponentielle de portée $b = 0.05$, pour $1 - \hat{\alpha} = 0.999$. La droite horizontale correspond à $y = \eta$, et la droite verticale à $x =$ portée pratique.*

taille de fenêtre, par exemple lorsque la distance entre le centre de la fenêtre et un de ses bords attendrait la portée pratique (distance au-delà de laquelle il n'y a plus de corrélation entre les données) représentée par la droite verticale. Cela ne semble cependant pas être

le cas. Pourquoi ? Pour évaluer la puissance en $\mathbf{x}$, centre d'une fenêtre $\mathcal{F}$, nous regardons dans un *voisinage de* $\mathbf{x}$ si il y a des changements abrupts, sachant qu'une discontinuité linéaire passe par $\mathbf{x}$ (équation (10.6)). Cela engendre différents problèmes. Lorsque la taille de la fenêtre augmente, la part d'aléatoire contenue dans la fenêtre aussi. Ainsi, plus la taille de la fenêtre est grande, plus il y a de chances de détecter une discontinuité et pas nécessairement et exclusivement celle introduite. Ensuite, parce qu'une discontinuité est plus probablement détectée là où l'échantillon de points est dense. Et enfin, parce que plus la taille de la fenêtre augmente, moins l'hypothèse alternative est pertinente car nous perdons l'aspect local. En effet, nous avons supposé que la courbe de discontinuité pouvait localement, *i.e.* au centre $\mathbf{x}$ de la fenêtre, être approchée par sa tangente en $\mathbf{x}$. En faisant croître la taille de la fenêtre nous imposons que la discontinuité soit une droite du type de celle utilisée pour les simulations (voir exemple sections 5.3 et 6.5). Par conséquent, nous calculons la puissance pour un modèle particulier de courbes de discontinuité. Nous sommes donc confronté à un dilemme similaire à celui type biais/variance, rencontré en modélisation. L'erreur de modélisation est la somme du carré du biais et de la variance de tous les modèles possibles. La difficulté dans la minimisation de cette erreur est que biais et variance sont antagonistes : la diminution de l'un équivaut à l'augmentation de l'autre, aussi il faut faire certains compromis. Dans notre cas le problème est que nous ne pouvons pas évaluer le biais et donc pas, ou difficilement, spécifier de compromis.

Considérer une discontinuité représentée par une droite passant par $\mathbf{x}$ centre d'une fenêtre $\mathcal{F}$ implique un biais important par rapport au vrai modèle de discontinuité lorsque la taille de $\mathcal{F}$ devient grande. C'est ce que nous illustrons sur les figures 10.8*a* et 10.8*b*. Sur chacune de ces figures nous avons représenté en rouge le modèle de discontinuité que nous utilisons pour calculer la puissance en $\mathbf{x}$ et en vert différentes configurations possibles de discontinuités réelles. Les pointillés désignent les zones où nous commettons une erreur par rapport au vrai modèle. Deux fenêtres centrées en $\mathbf{x}$ et de taille différente sont représentées en gris. Pour une fenêtre de petite taille l'erreur est minime, mais pour une fenêtre de grande taille l'erreur est non négligeable. Comme nous ne connaissons pas le vrai modèle, il nous est impossible de déterminer l'écart par rapport au vrai modèle.

Nous calculons la puissance en considérant le niveau local $1 - \hat{\alpha}$. Dans ce cas, sous l'hypothèse nulle, il y a une grande probabilité d'avoir au plus une composante connexe sur $\mathcal{D}$ (section 6.4) et le problème d'agrégation des tests locaux ne se pose pas. Ainsi, pour une discontinuité $a = 0$, la puissance va se rapprocher du niveau global η lorsque la taille de la fenêtre augmente. Nous choisissons comme valeur de k (demi longueur de la fenêtre

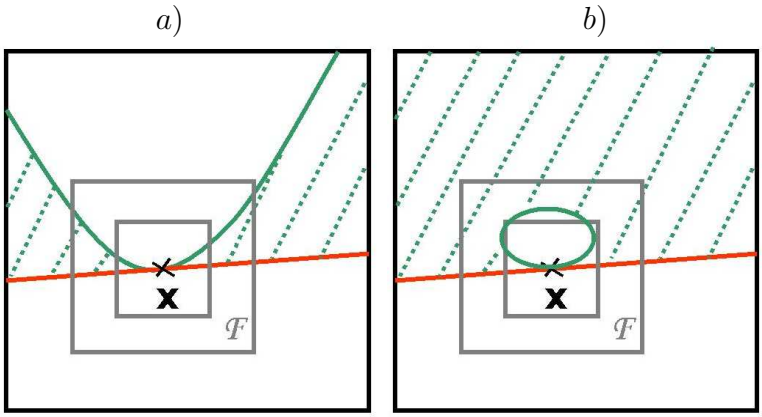


FIG. 10.8: *Erreur commise (pointillés verts) en fonction de la taille de la fenêtre (en gris), en considérant une discontinuité représentée par une droite passant par* $\mathbf{x}$ *(en rouge) au lieu du vrai modèle de discontinuité (en vert), modèle de type : a) parabole, b) elliptique.*

carrée), celle pour laquelle la puissance atteint le niveau η, ce qui correspond en fait à la portée pratique. La figure 10.7 illustre ce phénomène et représente la courbe de puissance (en noir pour $a = 0$) en fonction de k et la droite horizontale correspond à $y = \eta = 0.05$. Dans ce cas la portée pratique vaut $3b = 0.15$, ce qui correspond à approximativement $k = 9$ (droite verticale). La courbe de puissance intersecte la droite $y = \eta$ entre $k = 9$ et $k = 10$, *i.e.* en une valeur proche de celle de la portée pratique.

Remarquons tout de même que ce n'est pas vrai de façon générale, cela dépend de la densité de points d'échantillonnage dans le voisinage de $\mathbf{x}$. C'est ce que nous illustrons dans la figure 10.9. Nous avons utilisé la même procédure de simulation que pour l'exemple illustré dans la figure 10.7, sauf que l'échantillon n'est pas réparti aléatoirement sur le domaine ; il est de plus forte densité dans la moitié supérieure et présente une zone très faiblement échantillonnée dans la partie droite (figure 10.9*a*). Nous avons calculé la puissance en fonction du paramètre k en différents sites du domaine (points rouge, vert et bleu de la figure 10.9*a*) et pour différentes valeurs de la discontinuité $a = 0, 2, 2.5, 3$ (figures 10.9*b*, 10.9*c*, 10.9*d* et 10.9*e* respectivement). Chaque courbe de puissance est représentée de la même couleur que le point en lequel elle a été calculée. Il apparaît clairement que plus l'échantillon de points est dense dans le voisinage de $\mathbf{x}$, plus la puissance est élevée. De plus, nous avons tracé dans le cas $a = 0$ (figure 10.9*b*) les droites $x =$ portée pratique et $y = \eta$. Nous pouvons constater que lorsque l'échantillon est dense dans le voisinage de $\mathbf{x}$, la puissance vaut η exactement en $k =$ portée pratique. En revanche, moins il est dense, plus k doit être grand pour que la puissance atteingne η.

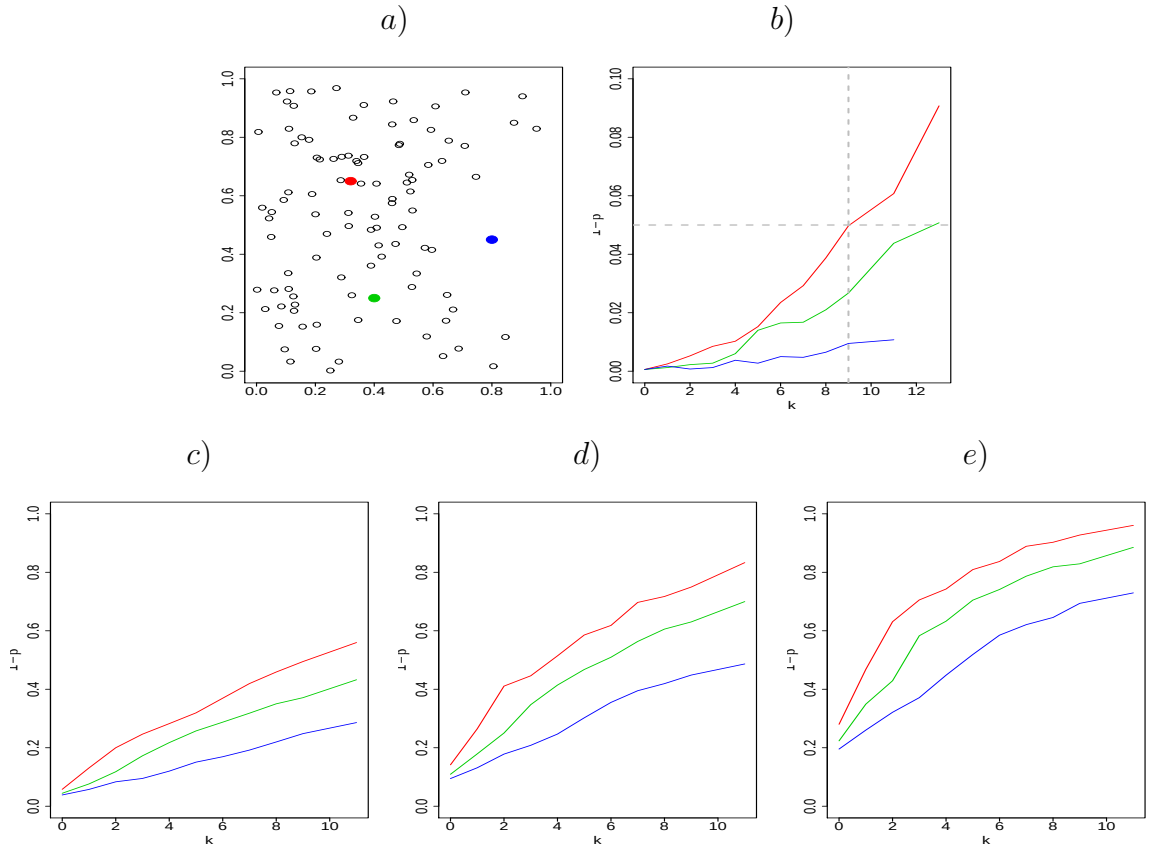


FIG. 10.9: *Courbe de puissance en fonction du paramètre k (demi longueur de la fenêtre carrée), calculée en différents points (a) : points rouge, vert et bleu) d'un carré de côté un dans lequel ont été générés 100 points issus d'un champ gaussien de covariance exponentielle de portée* $b = 0.05$*, pour* $1 - \hat{\alpha} = 0.9994$ *et différentes valeurs de la discontinuité : b)* $a = 0$*, c)* $a = 2$*, d)* $a = 2.5$ *et e)* $a = 3$ *. Sur la figure b) la droite horizontale correspond à* $y = \eta$*, et la droite verticale à* $x =$ *portée pratique.*

Rappelons qu'un des objectifs premiers du calcul de la puissance locale est qu'elle permet de déterminer si l'échantillon de points considéré est localement suffisamment dense. Aussi une première approche pourrait consister à considérer une fenêtre carrée de demi longueur $k =$ portée pratique. En effet, nous venons de voir que lorsque l'échantillon de points est localement dense, nous avons pour $k =$ portée pratique, $1 - \beta = \eta$ sous l'hypothèse nulle. Dans le cas d'un échantillon localement peu dense nous aurons, en utilisant la même taille de fenêtre, une puissance plus faible et c'est bien ce que nous souhaitons mettre en évidence. Le problème avec cette approche c'est que tout dépend de la portée pratique. Si elle est grande, la fenêtre sera grande et nous seront confrontés au problème mentionné précédemment sur la perte de l'aspect local et la considération d'une alternative erronée. De plus, moins l'échantillon est dense, plus la portée pratique

est grande.

Tout ce que nous pouvons faire à ce stade, c'est borner par valeur supérieure la taille de la fenêtre en considérant $k \leq$ portée pratique. Une étude sur simulations pourrait nous ouvrir d'autres pistes pour trouver un compromis entre la taille de la fenêtre et la conservation de l'hypothèse alternative. Ce travail reste à faire.

Discussion

Dans cette partie nous avons évalué une approximation de la puissance du test de détection de changement abrupt défini dans la section 5.2, en calculant la puissance au centre $\mathbf{x}$ d'une fenêtre contenue dans le domaine $\mathcal{D}$. Pour déterminer cette puissance nous avons calculé la probabilité qu'il y ait au moins un point de la fenêtre qui présente une discontinuité sous l'hypothèse d'existence d'une discontinuité représentée par une droite passant par $\mathbf{x}$:

$$1 - \beta(\mathcal{F}) = \mathbb{P}_{H_1}\left[\cup_{i=1}^{n_{\mathcal{F}}} (\text{rejeter } H_0(\mathbf{x}_i))\right].$$

Comme évoqué dans le chapitre précédent, le problème soulevé par la considération d'une telle discontinuité est que lorsque la taille de la fenêtre devient grande, ce modèle n'est plus valable dans un cadre général car il impose une forme de discontinuité particulière et ne cherche plus à approcher la discontinuité par sa tangente en $\mathbf{x}$. Cela soulève donc le problème de la pertinence de H_1 pour de grandes fenêtres, et par conséquent de la taille et du type de modèle de discontinuité à considérer. Il est difficilement envisageable de prendre en compte la taille de la discontinuité, compte tenu du fait que nous ne connaissons déjà pas la forme et l'orientation de celle-ci.

Nous aurions pu adopter d'autres façons de calculer la puissance. Par exemple, au lieu de chercher si un point de la fenêtre présente un changement abrupt en considérant comme alternative l'existence d'une droite passant par le centre $\mathbf{x}$ de la fenêtre, nous aurions pu calculer la probabilité que $\mathbf{x}$ présente un changement abrupt sachant qu'une discontinuité traverse la fenêtre. Ou encore, au lieu de considérer que la discontinuité peut localement être approchée par une droite, nous aurions pu considérer un ensemble de formes simples que la discontinuité pourrait prendre (mais cela impose de connaître la taille de la discontinuité). L'inconvénient commun de ces deux alternatives est que cela représente en pratique des temps de calcul non négligeables, étant plus complexes que l'approche que nous avons adoptée (qui elle-même prend du temps).

Bien qu'en règle générale l'hypothèse alternative que nous avons considéré n'est valable

que pour des fenêtres de petite taille, notre approche est correcte pour la classe particulière de modèles de type “faille”.

Quatrième partie

Validation de la méthode

Introduction

La méthode de détection de Zones de Changement Abrupt désormais développée, nous devons la valider. C'est l'objectif de cette partie, où la validation est faite via une étude sur simulations.

La puissance de la méthode, *i.e.* le taux de détection de ZCAs singificatives en présence d'une discontinuité, est étudiée dans le chapitre 11 en fonction de la densité de points. Nous avons vu dans la section 7.3 que la fonction de covariance du champ aléatoire $Z(\cdot)$ était en théorie supposée connue et qu'en pratique elle devait être estimée. Dans le chapitre 12, nous étudions la sensibilité de la méthode par rapport à une erreur dans l'estimation de la forme paramétrique et du paramètre de portée de la fonction de covariance. Nous nous intéressons également à la sensibilité de la méthode par rapport à la discrétisation. Dans la section 8.1, nous avons discuté de l'interpolateur à utiliser, selon que l'espérance de $Z(\cdot)$ était connue ou non. Toujours dans le chapitre 12, nous comparons les résultats de la méthode en utilisant le krigeage simple et le krigeage ordinaire comme interpolateurs.

Notre procédure de simulation est la suivante. Dans un carré unitaire, nous générons un vecteur $\mathbf{Z}$ de n points répartis aléatoirement, $n \in \{50, 100, 200\}$, et issus d'un champ aléatoire gaussien $\mathcal{N}(0,1)$ de fonction de covariance exponentielle :

$$C_Z(\mathbf{h}) = \exp(-\mathbf{h}/b),\ \mathbf{h} \geq 0,$$

de portée $b \in \{0.05, 0.07, 0.1, 0.13, 0.2\}$, ou de fonction de covariance sphérique :

$$C_Z(\mathbf{h}) = 1 - \frac{3}{2}\frac{\mathbf{h}}{b} + \frac{1}{2}\left(\frac{\mathbf{h}}{b}\right)^3 \text{ si } 0 \leq \mathbf{h} < b \text{ et } 0 \text{ si } b \geq \mathbf{h},$$

de portée $b = 0.3$. Nous avons introduit une discontinuité le long de la ligne $x^1 = 0.4$ en ajoutant une constante $a \in \{0, 1, 1.5, 2, 2.5, 3\}$ aux points d'échantillonnage situés dans la zone $x^1 < 0.4$. Le cas $a = 0$ correspond à H_0 (absence de discontinuité). Les calculs ont ensuite été effectués pour diverses discrétisations : grille 15×15, 30×30 et 60×60.

L'illustration de la section 6.5 montre que sous l'hypothèse nulle aucune ZCA potentielle n'est détectée et qu'en présence d'une discontinuité de $a = 2.5$, des ZCAs significatives apparaissent le long de la discontinuité et des non significatives sur les bords de la partie gauche du domaine. Nous avons appliqué la méthode sur une autre simulation dont les points d'échantillonnage proportionnels à leur valeur pour une discontinuité $a = 2.5$ sont représentés sur la figure 10.10*a*. Sur cette simulation nous voyons apparaître (figure 10.10*b*) des ZCAs significatives liées à la discontinuité introduite en $x^1 = 0.4$: celles alignées le long de $x^1 = 0.4$ et celles dûes aux effets de bord (l'extérieur du domaine étant supposé de moyenne nulle et l'intérieur du rectangle $[0, 0.4] \times [0, 1]$ de moyenne a). Une ZCA significative est également présente dans la partie nord-est du domaine, qui est indépendante de la discontinuité. Aussi, dans la suite nous appellerons *vraie ZCA* une

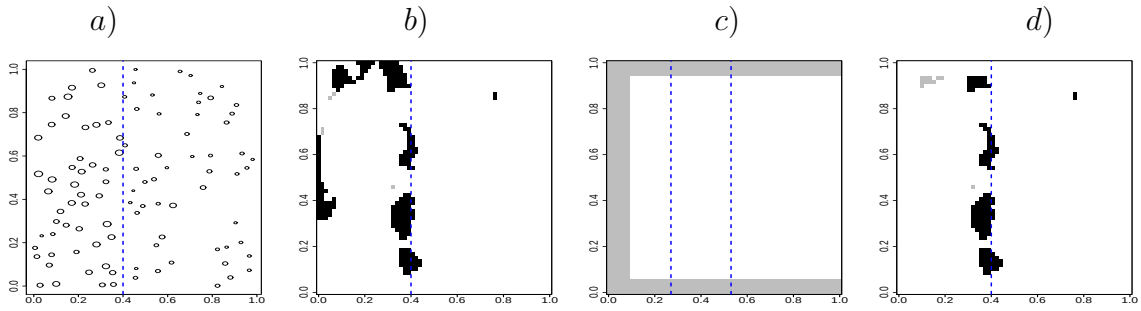


FIG. 10.10: *a) : Echantillon de points issus d'un processus avec une discontinuité le long des pointillés ; b) : ZCAs (significatives en noir) détectées ; c) : les pointillés délimitent la zone permettant de définir les vraies ZCAs et la zone grise représente les bords que nous éliminons ; d) : ZCAs (significatives en noir) obtenues après élimination des bords.*

ZCA significative dont le maximum, de coordonnées $[i, j]$, est tel que i appartient à l'intervalle $[0.27, 0.53]$ (figure 10.10*c*, pointillés). Si le maximum ne vérifie pas cette condition, nous parlerons d'*autre ZCA*.

Plus a diffère de zéro, plus nous nous éloignons de l'hypothèse nulle, et plus des effets de bord peuvent apparaître, comme sur la figure 10.10*b*. Par conséquent dans la suite lorsque nous évaluerons les ZCAs, nous éliminerons le bord : les ZCAs seront déterminées sur le domaine $x^1 \times x^2 \in [0.1, 1] \times [0.06, 0.96]$, *i.e.* en ne considérant pas la zone grisée de la figure 10.10*c*. Cela fournit des résultats du type de ceux de la figure 10.10*d*. Les ZCAs dûes au bord sont bien éliminées et les ZCAs dûes à la discontinuité à l'intérieur du domaine sont conservées. Il peut arriver, et c'est le cas dans cet exemple, qu'une ZCA significative près du bord et sur la discontinuité devienne non significative à cause de la réduction de sa surface lors de l'élimination du bord.

Chapitre 11

Puissance

11.1 Puissance globale

Nous avons calculé sur une grille 60×60 le nombre de simulations (sur 100) ayant des vraies ZCAs et le nombre de simulations ayant des autres ZCAs, pour différentes valeurs a de la discontinuité et pour un modèle de covariance exponentiel de portée $b = 0.1$. Nous avons considéré comme niveau local $1 - \alpha = 0.995$, 0.999 et $1 - \hat{\alpha}$ pour un échantillon de $n = 100$ points. Les résultats sont reportés dans la table 11.1. Le taux de détection de

		Vraies ZCAs		
a	$1 - \alpha =$	0.995	0.999	$1 - \hat{\alpha}$
1		20	6	5
1.5		45	23	17
2		75	53	45
2.5		90	78	73
3		100	93	90
		Autres ZCAs		
1		14	4	3
2		19	3	3
3		23	10	8

TAB. 11.1: *Nombre de simulations (sur 100) ayant des vraies ZCAs et/ou des autres ZCAs pour différentes valeurs de* $1 - \alpha$ *(grille* 60×60*,* $n = 100$*, covariance exponentielle de portée* $b = 0.1$*).*

vraies ZCAs indique que la discontinuité est plus souvent détectée lorsque a augmente, et est presque toujours détectée à partir de $a = 2.5$. Le taux de détection (de vraies et d'autres ZCAs) décroît avec l'augmentation du niveau $1 - \alpha$.

Les résultats de la table 11.1 montrent également que pour un échantillon de 100 points dans un carré de côté un, il n'y a pas de différence significative du taux de détection pour les niveaux $1 - \hat{\alpha}$ ou 0.999. Aussi pour une telle densité de points, nous pouvons utiliser comme première approximation le niveau 0.999 au lieu du niveau $1 - \hat{\alpha}$ qui nécessite plus de calculs.

Nous avons calculé les histogrammes (figure 11.1) de la coordonnée i des pixels $[i, j]$ correspondants à des ZCAs significatives au niveau $1 - \alpha = 0.995$ (en noir) et $1 - \hat{\alpha}$ (en blanc), pour 100 simulations de 100 points sur une grille 60×60, avec $a = 2.5$, et un modèle de covariance exponentiel de portée $b = 0.1$. La discontinuité a été introduite en $i = 24$ et est représentée d'une couleur différente que le reste de l'histogramme. Ces

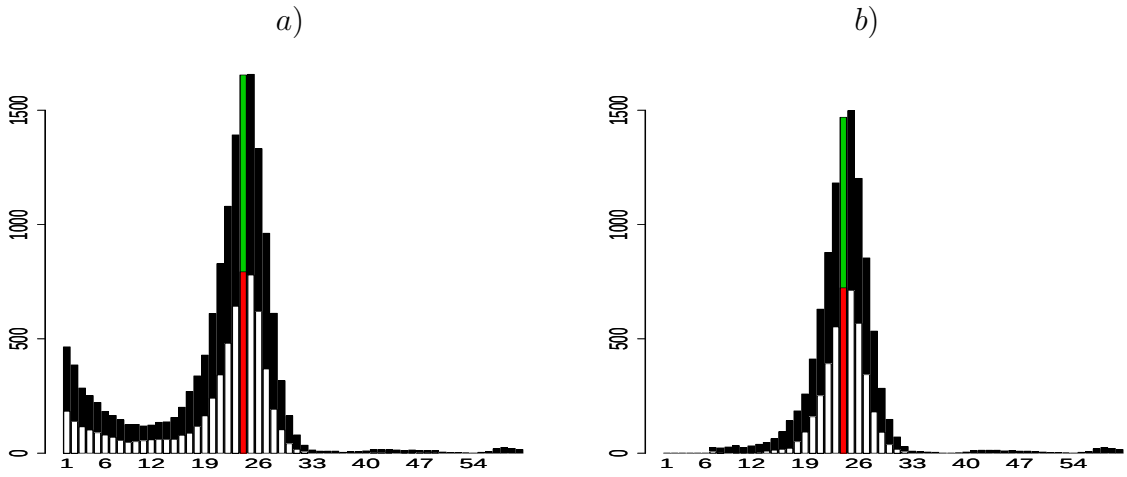


FIG. 11.1: *Histogrammes du cumul des coordonnées i des pixels $[i, j]$ correspondants à des ZCAs significatives pour différents niveaux $1 - \alpha$ ($1 - \hat{\alpha}$ en blanc et 0.995 en noir) pour 100 simulations. Le cumul des pixels correspondant à la discontinuité ($a = 2.5$) en $i = 24$ a été tracé de couleur différente. Les bords ont été conservés figure a) et éliminés figure b).*

histogrammes montrent que les pixels constituant les ZCAs significatives appartiennent le plus souvent à la discontinuité ou en sont très proche. La figure 11.1*a* illustre l'effet de bord. L'élimination des bords (figure 11.1*b*), n'a pas d'influence sur la détection des ZCAs le long de la discontinuité. L'effet de bord peut donc être éliminé sans que la détection de ZCAs ne soit affectée.

11.2 Effet de la densité de points

La méthode de détection de ZCAs étant basée sur le gradient local, la densité locale d'échantillonnage a une conséquence directe sur sa puissance. Nous avons déterminé le nombre de simulations ayant au moins une vraie ZCA et le nombre de simulations ayant au moins une autre ZCA au niveau $1 - \hat{\alpha}$ sur une grille 60×60 pour un ensemble de 100 simulations de n points issus d'un champ gaussien de fonction de covariance exponentielle de portée $b = 0.1$, en faisant varier la nombre de points, $n \in \{50, 100, 200\}$ et la valeur de la discontinuité $a \in \{1, 1.5, 2, 2.5, 3\}$. Les résultats obtenus sont donnés dans la table 11.2. Les tailles moyennes des ZCAs détectées sont également données dans cette table. Ces

	$n = 50$		$n = 100$		$n = 200$	
			Vraies ZCAs			
a	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$
1	9	*0.12*	5	*0.04*	10	*0.03*
1.5	21	*0.34*	17	*0.13*	23	*0.11*
2	43	*0.87*	45	*0.51*	58	*0.41*
2.5	75	*2.33*	73	*1.20*	86	*1.07*
3	88	*4.24*	90	*2.45*	100	*2.09*
			Autres ZCAs			
a	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$
1	6	*0.05*	3	*0.01*	8	*0.02*
2	15	*0.29*	3	*0.01*	9	*0.02*
3	37	*1.06*	8	*0.06*	10	*0.03*

TAB. 11.2: *Nombre de simulations, N, sur 100 présentant au moins une vraie ZCA et taille moyenne, $\bar{s}$, des ZCAs détectées, en fonction de la densité de points n et de la valeur de la discontinuité a.*

résultats montrent que le taux de détection de vraies ZCAs est d'autant plus important que le nombre de points est élevé. Cela ne semble pas évident lorsque le nombre de points est faible : il y a peu de différences, en terme de taux de détection de vraies ZCAs, entre les simulations avec $n = 50$ et $n = 100$. Nous avons calculé les histogrammes des coordonnées i des pixels $[i, j]$ appartenant aux ZCAs significatives pour une discontinuité $a = 2.5$. Ces histogrammes sont représentés sur la figure 11.2, pour $n = 50$ (figure 11.2*a*), $n = 100$ (figure 11.2*b*) et $n = 200$ (figure 11.2*c*). Cette figure illustre le fait que plus le nombre de points est important, plus les ZCAs s'alignent le long de la discontinuité. Ce phénomène est aussi observable dans la table 11.2. La taille moyenne des vraies ZCAs diminue lorsque le nombre de points augmente.

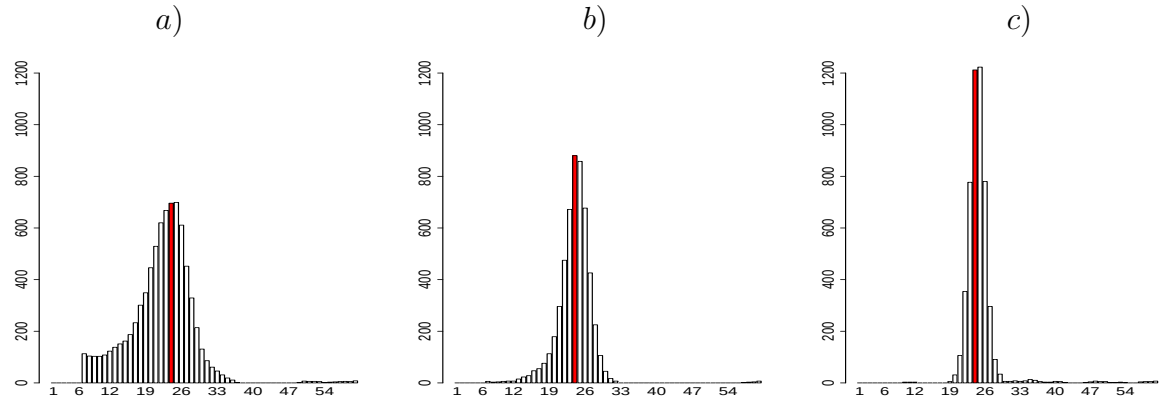


FIG. 11.2: *Histogrammes des coordonnées i des pixels $[i, j]$ constituant les ZCAs significatives au niveau $1 - \hat{\alpha}$ pour 100 simulations, en fonction de la densité d'échantillonnage : a) $n = 50$, b) $n = 100$ et c) $n = 200$. Le cumul des pixels correspondant à la discontinuité ($a = 2.5$) en $i = 24$ a été tracé de couleur différente.*

En ce qui concerne les autres ZCAs, le taux de détection est bien plus élevé pour $n = 50$ que dans les deux autres cas. Cela vient de l'élimination du bord, des ZCAs peuvent être subdivisées et leur maximum local ne se trouve plus dans la zone définissant les vraies ZCAs. Ce taux est faible pour $n = 100$ et a tendance à augmenter avec le nombre de points. Cela est dû au fait que plus le nombre de points est important, plus la probabilité que deux points proches aient des valeurs très différentes et présentent un gradient fort est élevée. Rappelons que la méthode peut être utlisée pour détecter des valeurs aberrantes.

Chapitre 12

Sensibilité

12.1 Par rapport à l'estimation de la covariance

Une fonction de covariance est caractérisée par sa famille paramétrique (exponentielle, sphérique, ...) et en général par deux paramètres : la variance, σ^2, et la portée, b. Dans la section 7.3, nous avons étudié l'estimation de la variance sous l'alternative et proposé une procédure itérative permettant d'estimer la fonction de covariance, le niveau $1-\alpha$ et les ZCAs et corrigeant le biais introduit dans l'estimation de la variance sous l'alternative. Nous nous intéressons dans cette section, à la sensibilité de la méthode par rapport à une erreur dans l'estimation de la portée d'une part, et par rapport à l'estimation de la famille paramétrique d'autre part.

Nous avons effectué 100 simulations en considérant comme modèle de covariance, un modèle exponentiel de portée $b = 0.1$. Nous avons ensuite appliqué la méthode avec cinq fonctions de covariance "erronées" : quatre d'entre elles avec la même forme paramétrique mais dont nous avons ajouté $\pm 30\%$ ou $\pm 50\%$ à la portée, et une de forme paramétrique différente. Dans le dernier cas nous avons utilisé un modèle de covariance sphérique. La portée du modèle sphérique a été choisie de sorte que les deux modèles de covariance aient des portées similaires, *i.e.* pour $\mathbf{h} = b$, $\exp(-\mathbf{h}/0.1) = 0.05 \Rightarrow b = 0.3$. La table 12.1 contient le nombre de simulations sur 100 ayant des vraies des ZCAs, le nombre de simulations sur 100 ayant d'autres ZCAs, ainsi que les tailles moyennes des ZCAs détectées, pour les différents modèles de covariance.

Il n'y a pas de différence importante du taux de détection et de la taille moyenne des ZCAs pour les modèles exponentiels de portée $b = 0.07$ et $b = 0.13$ et pour le modèle

	Exp(0.05)		Exp(0.07)		Exp(0.1)		Exp(0.13)		Exp(0.2)		Sph(0.3)	
						Vraies ZCAs						
a	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$
1	6	*0.04*	5	*0.03*	5	*0.04*	13	*0.10*	33	*0.39*	14	*0.11*
1.5	22	*0.18*	15	*0.11*	17	*0.13*	36	*0.35*	57	*0.91*	36	*0.30*
2	43	*0.50*	40	*0.42*	45	*0.51*	58	*0.86*	83	*1.91*	58	*0.75*
2.5	79	*1.32*	69	*1.09*	73	*1.20*	82	*1.93*	95	*3.44*	83	*1.63*
3	97	*2.56*	92	*2.23*	90	*2.45*	95	*3.40*	100	*5.23*	97	*2.89*
						Autres ZCAs						
a	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$
1	4	*<0.01*	2	*<0.01*	3	*0.01*	8	*0.04*	39	*0.32*	18	*0.07*
2	23	*0.12*	9	*0.04*	3	*0.01*	8	*0.04*	38	*0.31*	16	*0.06*
3	63	*0.69*	17	*0.11*	8	*0.06*	11	*0.08*	35	*0.30*	17	*0.09*

TAB. 12.1: *Nombre, N, de simulations (sur 100) ayant des ZCAs significatives et taille moyenne, $\bar{s}$, des ZCAs détectées, en considérant comme modèle vrai un modèle exponentiel de portée $b = 0.1$, pour une grille 60×60 et $n = 100$. Les ZCAs ont été calculées pour cinq autres modèles (caractérisant une erreur dans l'estimation du paramètre de portée ou de la forme paramétrique).*

sphérique. En revanche, pour les modèles exponentiels où la portée a été grossièrement mal estimée ($\pm 50\%$ de la portée vraie), il y a des différences non négligeables. Pour le modèle exponentiel de portée $b = 0.05$, le taux de détection de vraies ZCAs n'est pas significativement différent de celui du cas témoin, bien qu'un peu supérieur, mais le taux de détection de fausses ZCAs est bien supérieur à celui du cas exponentiel de portée $b = 0.1$. En regardant les simulations plus en détail, nous avons pu constater que lorsque la portée est courte les ensembles d'excursion sont moins lisses et sont souvent constitués de petites composantes connexes. Parmi elles, certaines ont leur centre hors des limites utilisées pour définir les vraies ZCAs, bien que très proche de ces limites. Elles contribuent par conséquent à augmenter le nombre d'autres ZCAs et leur taille moyenne. Pour le modèle exponentiel de portée $b = 0.2$, les taux de détection de vraies et de fausses ZCAs sont élevés. Cela vient du fait que le champ aléatoire sous-jacent est plus lisse que pour une portée $b = 0.1$, ce qui aboutit à plus de rejet de l'hypothèse nulle en présence d'une discontinuité. Par conséquent, lorsque la portée est sur-estimée, les ZCAs sont plus détectées. Pour les mêmes raisons de régularité, la taille des ZCAs détectées augmente avec le paramètre de portée. Notons que la taille des autres ZCAs est toujours très petite ($< 0.1\%$ de la surface du domaine d'étude pour $b \in \{0.07, 0.1, 0.13\}$) en accord avec le fait que les autres ZCAs sont reliées à l'erreur de type I.

La détection de ZCAs n'est donc pas très sensible à une erreur commise dans l'estimation de la fonction de covariance pourvu que l'erreur ne soit pas grossière, auquel cas le taux de détection de fausses ZCAs ne serait pas négligeable. Une sur-estimation du paramètre de portée est une erreur plus acceptable qu'une sous-estimation en terme de détection : elle détecte plus de vraies ZCAs sans augmenter le FDR. Mal-estimer la forme paramétrique de la fonction de covariance est d'une importance secondaire si la portée est bien estimée.

12.2 Par rapport à la discrétisation

Nous avons vu dans la section 7.1 que selon la discrétisation utilisée les composantes connexes de taille inférieure à un pixel peuvent ne pas être détectées. Ce phénomène est illustré dans la figure 12.1. Les ZCAs au niveau $1 - \hat{\alpha}$ ont été obtenues sur une simulation pour différentes valeurs de la discrétisation : 15×15, 30×30 et 60×60. La figure

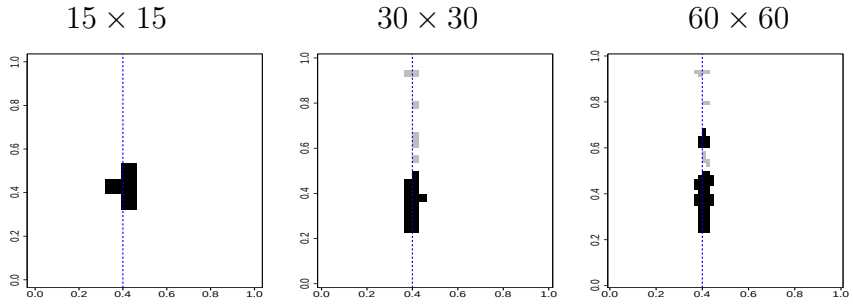


FIG. 12.1: *Effet de la discrétisation sur la détection des ZCAs au niveau* $1 - \hat{\alpha}$ *(significatives en noir, non significatives en gris).*

12.1 montre que lorsque la discrétisation est trop grossière (15×15), les ZCAs de petite taille (inférieure à un pixel) ne sont pas détectées. Au contraire lorsque la grille devient assez fine (60×60), des petites ZCAs apparaissent et s'alignent le long de la discontinuité.

Nous avons calculé sur 100 simulations, le nombre de simulations ayant au moins une vraie ZCA et le nombre de simulations ayant au moins une autre ZCA ainsi que la taille moyenne des ZCAs détectées pour des grilles de taille 15×15, 30×30 et 60×60. Les résultats sont reportés dans la table 12.2. Le taux de détection des vraies ZCAs et d'autres ZCAs reste constant quelle que soit la discrétisation utilisée et quelle que soit la valeur

	Grille 15×15		Grille 30×30		Grille 60×60	
a			*Vraies ZCAs*			
	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$
1	8	*0.07*	5	*0.05*	5	*0.04*
1.5	16	*0.16*	15	*0.14*	17	*0.13*
2	40	*0.52*	45	*0.52*	45	*0.51*
2.5	70	*1.28*	75	*1.40*	73	*1.20*
3	91	*2.54*	92	*2.66*	90	*2.45*
			Autres ZCAs			
	N	$\bar{s}$	N	$\bar{s}$	N	$\bar{s}$
1	2	*0.01*	1	*<0.01*	3	*0.01*
2	4	*0.02*	1	*<0.01*	3	*0.01*
3	6	*0.06*	6	*0.04*	8	*0.06*

TAB. 12.2: *Nombre, N, de simulations (sur 100) ayant des ZCAs significatives au niveau $1 - \hat{\alpha}$ et taille moyenne, $\bar{s}$, des ZCAs détectées, pour différentes discrétisation. La méthode utilise les paramètres d'une fonction de covariance exponentielle de portée $b = 0.1$ et des échantillons de $n = 100$ points.*

de la discontinuité a. La taille moyenne des ZCAs détectées ne varie pas non plus d'une discrétisation à l'autre.

Nous avons calculé les histogrammes des coordonnées i des pixels $[i, j]$ appartenant aux ZCAs significatives pour une discontinuité $a = 2.5$. Ces histogrammes sont représentés sur la figure 12.2, pour une grille 15×15 (figure 12.2*a*), 30×30 (figure 12.2*b*) et 60×60 (figure 12.2*c*). Ces histogrammes montrent que plus la grille est fine, plus nous pouvons gagner en précision de la délimitation de la discontinuité par la prise en compte des petites composantes connexes.

Par conséquent, nous pouvons conclure de ces simulations qu'il n'y a pas d'effet de discrétisation sur la méthode en terme de taux de détection de ZCAs. Cependant, considérer des discrétisations grossières ne permet pas d'obtenir une bonne visualisation des ZCAs détectées.

12.3 Par rapport au choix de l'interpolateur spatial

Dans la section 8.1 nous avons évoqué le problème du centrage de la variable lorsque nous ne connaissons pas la moyenne, et donc le problème de l'utilisation du krigeage

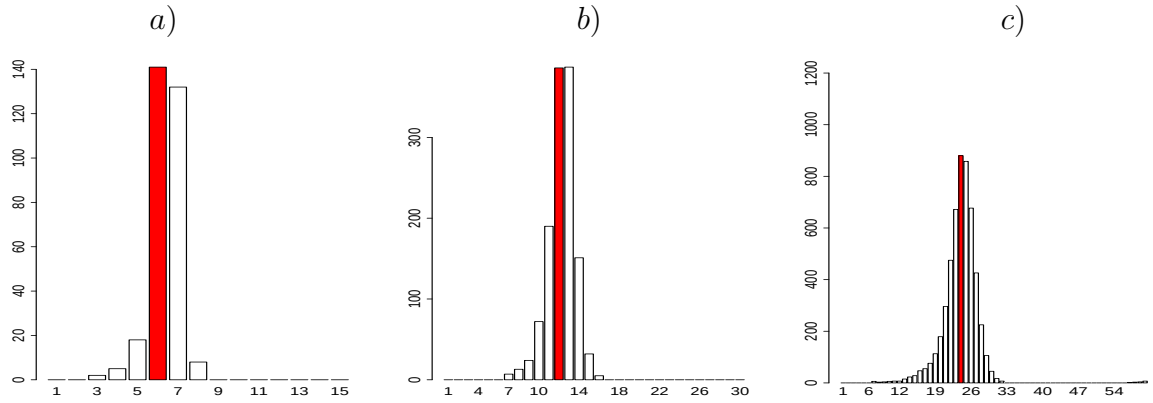


FIG. 12.2: *Histogrammes des coordonnées i des pixels $[i, j]$ constituant les ZCAs significatives au niveau $1 - \hat{\alpha}$ pour 100 simulations, en fonction de la discrétisation : a) 15×15, b) 30×30 et c) 60×60. Le cumul des pixels correspondant à la discontinuité ($a = 2.5$) a été tracé de couleur différente.*

simple dans de telles circonstances. Dans la table 12.3, nous avons reporté le nombre de simulations (sur 100) ayant au moins une vraie ZCA et le nombre de simulations ayant au moins une autre ZCA au niveau $1 - \hat{\alpha}$, dans le cas d'un modèle exponentiel de paramètre $b = 0.1$, pour $n = 100$. La méthode a été appliquée en prenant comme estimateur du champ aléatoire le krigeage simple d'une part et le krigeage ordinaire d'autre part. Le taux de rejet de l'hypothèse nulle est relativement identique que nous utilisions le krigeage simple ou le krigeage ordinaire.

Nous avons comparé le cumul des pixels correspondants à des ZCAs significatives selon que nous considérons le krigeage simple ou le krigeage ordinaire. Pour cela nous avons généré 100 simulations de 100 points sur une grille 60×60, de modèle exponentiel de portée $b = 0.1$ dans le cas $a = 2.5$. La figure 12.3 représente le cumul de la coordonnée i des pixels $[i, j]$ des ZCAs au niveau $1-\hat{\alpha}$. Elle illustre les résultats obtenus avec le krigeage simple (première ligne) et le krigeage ordinaire (deuxième ligne), selon que le bord soit conservé (figure 12.3a et 12.3c) ou non (12.3b et 12.3d). Quelque soit le krigeage utilisé, la discontinuité, qui correspond à $i = 24$, est détectée avec les mêmes proportions. Une différence notable apparaît entre les histogrammes 12.3a et 12.3c : le krigeage ordinaire élimine les effets de bord. Ayant en début de partie choisi d'éliminer les bords, cela n'a aucune influence sur nos résultats de simulations du fait que dans ce cas il n'y a plus de différence que nous considérions le krigeage simple ou le krigeage ordianire (figures 12.3b et 12.3d). Lorsqu'en pratique la méthode est appliquée et que nous utilisons le krigeage simple, la variable est centrée par rapport à la moyenne arithmétique de l'échantillon. Si l'échantillon est assez dense, la moyenne arithmétique fournit une bonne estimation

a	Krigeage simple		Krigeage ordinaire	
	Vraies ZCAs			
	N	$\bar{s}$	N	$\bar{s}$
1	5	*0.04*	5	*0.04*
1.5	17	*0.13*	18	*0.13*
2	45	*0.51*	46	*0.49*
2.5	73	*1.20*	72	*1.18*
3	90	*2.45*	90	*2.38*
	Autres ZCAs			
	N	$\bar{s}$	N	$\bar{s}$
1	3	*0.01*	2	*<0.01*
2	3	*0.01*	3	*0.02*
3	8	*0.06*	3	*0.04*

TAB. 12.3: *Nombre, N, de simulations (sur 100) ayant des ZCAs et taille moyenne, $\bar{s}$, des ZCAs détectées pour deux interpolateurs du champ : krigeage simple et krigeage ordinaire (grille* 60×60*, $n = 100$, covariance exponentielle de portée* $b = 0.1$*).*

de la moyenne et la méthode devient équivalente quelquesoit l'estimateur de la variable. Inversement, si l'échantillon est peu dense, la variable est mal représentée et la moyenne mal estimée. Dans de telles circonstances il est préférable d'utiliser le krigeage ordinaire.

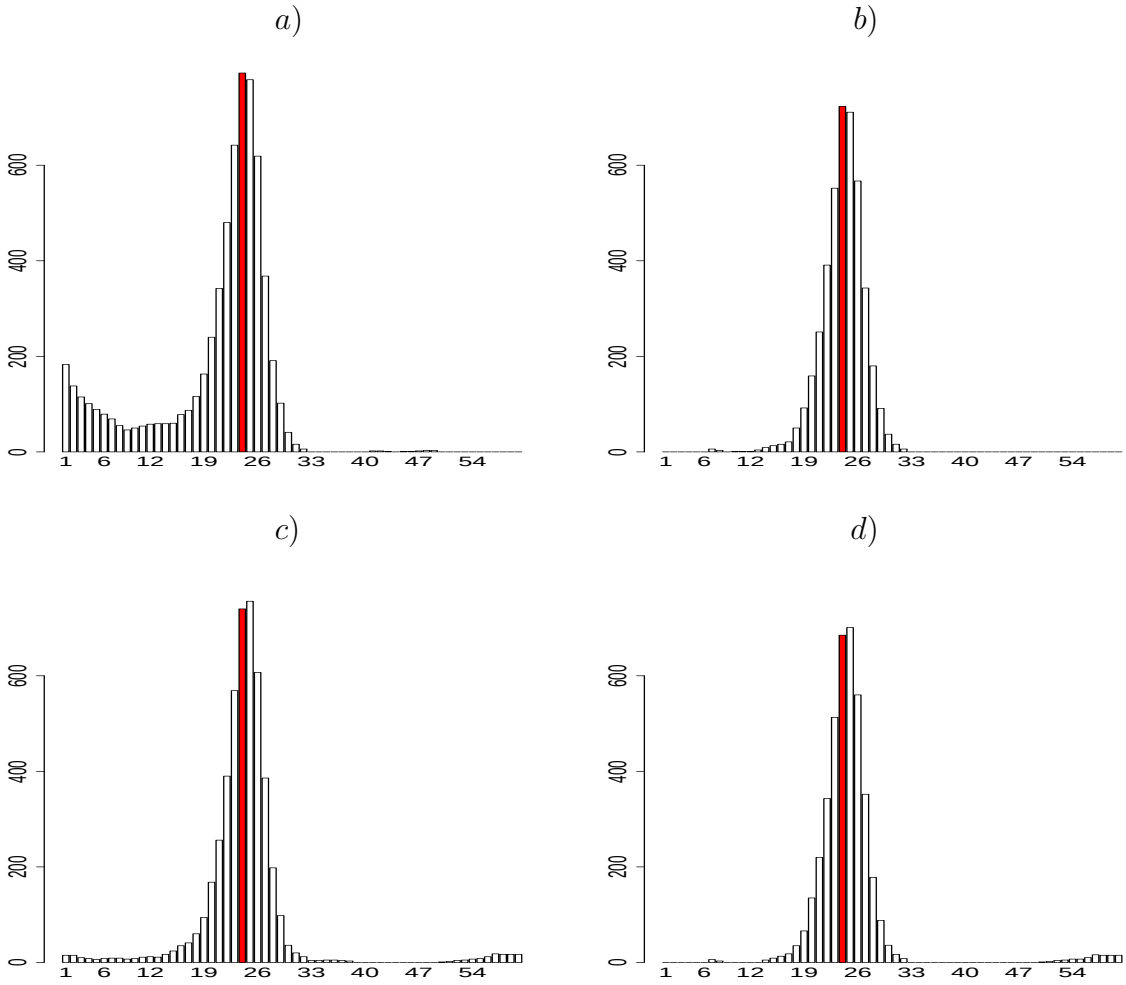


FIG. 12.3: *Histogrammes du cumul des coordonnées i des pixels* $[i, j]$ *correspondants à des ZCAs significatives au niveau* $1 - \hat{\alpha}$*, pour 100 simulations. Les interpolateurs utilisés sont le krigeage simple (première ligne) et le krigeage ordinaire (deuxième ligne). Le cumul des pixels correspondant à la discontinuité* $(i = 24)$ *a été tracé de couleur différente. Les bords ont été conservés figures a) et c) et éliminés figures b) et d).*

Discussion

La méthode de détection de ZCAs se révèle performante. Nous avons pu montrer que pour un échantillon de faible densité (100 points sur le domaine d'étude), une discontinuité d'une valeur de 2.5 écarts-type est presque toujours détectée et qu'elle apparaît, au moins partiellement, sur les représentations graphiques. Plus l'échantillon de points est dense, meilleure est la détection et plus nous gagnons en terme de précision de délimitation de la discontinuité.

L'étude sur simulations détaillée dans cette partie a permis de valider la robustesse de la méthode par rapport à différents critères. Le taux de détection est insensible à la discrétisation de la grille d'interpolation. Il est cependant préférable d'utiliser des grilles assez fines pour pouvoir capturer les ZCAs de petites tailles. La méthode est robuste par rapport à une erreur commise dans l'estimation de la fonction de covariance. Nous avons pu montrer que pour des erreurs d'estimation du paramètre de portée de l'ordre de $\pm 30\%$, le taux de détection et la surface des ZCAs évoluent peu. L'estimation du modèle paramétrique est quand à elle de second ordre si la portée est bien estimée. Notons qu'une sur-estimation de la portée est une erreur plus acceptable qu'une sous-estimation en terme de détection de ZCAs. Utiliser le krigeage simple ou le krigeage ordinaire comme interpolateur n'a pas d'effet sur la détection des ZCAs. Cependant en pratique, il vaut mieux utiliser le krigeage ordinaire, les densités de points considérées étant trop faibles pour estimer correctement la moyenne.

Cinquième partie

Application aux données d'agriculture de précision

Introduction

L'agriculture de précision vise à proposer des méthodes adaptées au traitement par modulation spatiale. Cela peut être le cas de la modulation de la fertilisation azotée afin d'optimiser la production en termes de quantité et de qualité et de minimiser la pollution des nappes phréatiques. Afin de parvenir à une modulation spatiale, il est nécessaire de considérer les caractéristiques du sol et leur variabilité spatiale et de définir des zones homogènes. La méthode de détection de Zones de Changement Abrupt décrite précédemment ne fournit pas un zonage de la parcelle, mais met en évidence les structures existantes, ce qui associé à d'autres informations (comme la pédologie et l'évolution temporelle) permettra de déterminer un zonage.

Nous allons considérer deux types de variables : des variables permanentes correspondant aux variables de constitution du sol (comme le limon, le calcaire ou l'argile), et des variables dynamiques (teneur en eau et en azote minéral) qui traduisent le fonctionnement hydrique et minéral du sol. Ces variables sont décrites dans le chapitre 13 et analysées dans le chapitre 14. Les variables dynamiques ont permis de mettre en évidence des structures permanentes superposables avec la carte pédologique de la parcelle et les structures obtenues avec les variables de constitution.

Chapitre 13

Présentation des données

Les données ont été collectées dans une parcelle de 10 hectares située à Chambry, près de Laon dans le nord de la France. Ce site est dédié au projet agriculture de précision conduit par l'INRA sur plusieurs années [Guérif *et al.* (2001)]. Le schéma d'échantillonnage utilisé est le même pour l'ensemble des variables. Dans ce chapitre, nous allons décrire successivement le plan d'échantillonnage adopté, les variables dynamiques et les variables permanentes.

13.1 Plan d'échantillonnage

Les variables ont été mesurées par carottage sur quatre horizons de 30 cm. Chaque carottage représente la moyenne de trois prélèvements pris à 50 cm de distance. Les variables permanentes sont mesurées sur le premier horizon, alors que les variables dynamiques représentent le cumul des quatre horizons. Chaque variable a été échantillonnée sur une grille pseudo-régulière. La distance entre nœuds est 36 m. Les points d'échantillonnage sont notés de A1 à G13. Des données similaires ont été collectées dans une parcelle voisine et présentées et analysées dans Mary *et al.* (2001).

L'INRA d'Orléans, en collaboration avec l'INRA de Laon, a réalisé la carte pédologique de cette parcelle. Les caractéristiques typologiques du sol ont été mesurées jusqu'à 150 cm de profondeur. Ces caractéristiques peuvent être de différente nature :

- quantitative, telle que la classe texturale qui intègre une notion de structure liée au matériau parental ou à l'évolution pédologique.
- qualitative, comme la profondeur, la teneur en calcaire, en cailloux et en graviers.
- synthétique, comme le substrat, le type de profil, le drainage interne.

Sur la parcelle d'étude, les différentes unités cartographiques de sol différenciées ont été regroupées en 12 types de sol (figure 13.1). La carte pédologique a été étalie en suivant un modèle pédogénétique. Les pédologues ont, à partir de la reconstruction de l'historique des sols, tracé manuellement les contours entre les différentes zones. Cette carte peut être

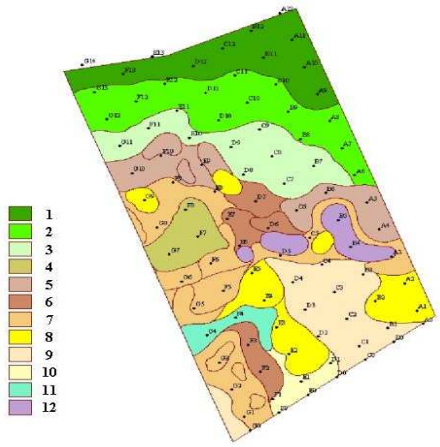


FIG. 13.1: *Carte pédologique [Nicoullaud* et al. *(2003)].*

1. *Luvisols limoneux sur matériau cryoturbé profond*
2. *Néoluvisols limoneux sur matériau cryoturbé profond*
3. *Complexe de calcisols et de néoluvisols érodés sur matériau cryoturbé*
4. *Complexe de calcisols et de calcosols sur craie remaniée sableuse*
5. *Complexe de calcosols sur craie, craie altérée et matériau cryoturbé*
6. *Calcosols sur craie remaniée*
7. *Calcosols sur craie sableuse et craie remaniée*
8. *Calcosols assez caillouteux sur craie*
9. *Calcosols peu caillouteux sur craie*
10. *Calcosols très peu caillouteux sur craie*
11. *Colluviosols*
12. *Calcomagnésisols sur sable grésifié*

synthétisée en regroupant différents types de sol, ce qui est illustré sur la figure 13.2*a*, avec :

- en rouge : les sols profonds, non calcaires et non caillouteux (1+2+3+4),
- en vert : les sols calcaires développés sur craie peu profonde (8+9+10) et le sol colluvial avec craie moyennement profonde (11),
- en orange : les sols calcaires développés sur des matériaux issus de la craie (6+7) et les sols calcaires sur craie altérée et matériau cryoturbé (5),
- en noir : les sols sur sable grésifié, trop particuliers pour être regroupés avec d'autres types de sols (12).

Les points d'échantillonnage (figure 13.2*b*) sont représentés par la couleur du type de sol auquel ils correspondent.

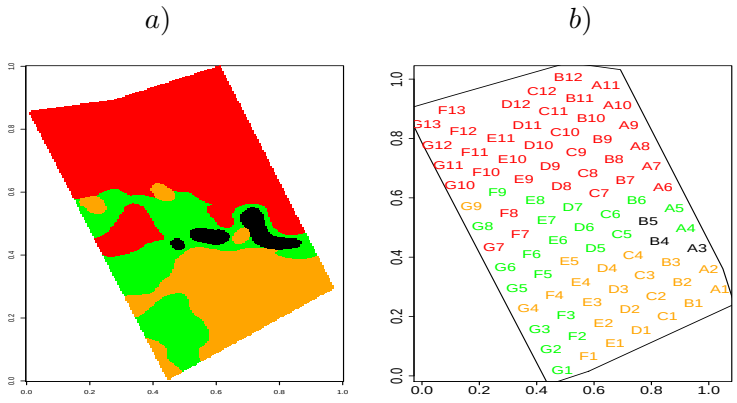


FIG. 13.2: *a) Carte pédologique simplifiée ; b) Schéma d'échantillonnage.*

- *Sols profonds non calcaires, non caillouteux,*
- *Sols calcaires développés sur des matériaux issus de la craie et sols calcaires sur craies altérées et matériaux cryoturbés,*
- *Sols calcaires développés sur craie peu profonde et sol colluvial avec craie moyennement profonde,*
- *Sols sur sable grésifié.*

13.2 Variables dynamiques

Les variables considérées sont la teneur en eau (en mm) et les reliquats en azote minéral (en kg/ha). La teneur en eau a été mesurée par pesée avant et après passage des échantillons du sol à l'étuve. Les reliquats d'azote minéral ont été extraits par une solution de chlorure de potassium molaire et déterminés par colorimétrie.
Les échantillons ont été prélevés à huit dates réparties sur trois années. Durant ces années diverses cultures se sont succédées. Les dates de prélèvement et l'évolution culturale sont données dans la table 13.1. La succession culturale est chronologiquement : betterave, blé, pois, blé. Les dates de prélèvement correspondent au semis et à la récolte des cultures, et à la période hivernale (février) correspondant à la période où la capacité de rétention d'eau est maximale. Le nombre de points des échantillons varie selon les années (table 13.1). En effet, les mesures étant faites sur les quatre horizons, selon les dates il est arrivé que le sol soit localement trop dur pour permettre certains prélèvements.

La figure 13.3 représente les échantillons de points de la teneur en eau aux différentes dates ainsi que leur interpolation (les valeurs fortes sont représentées en blanc et les valeurs faibles en noir). L'interpolation a été effectuée sur une grille 62×98 avec le krigeage ordinaire. Cette figure montre que la parcelle est plus humide dans sa partie sud

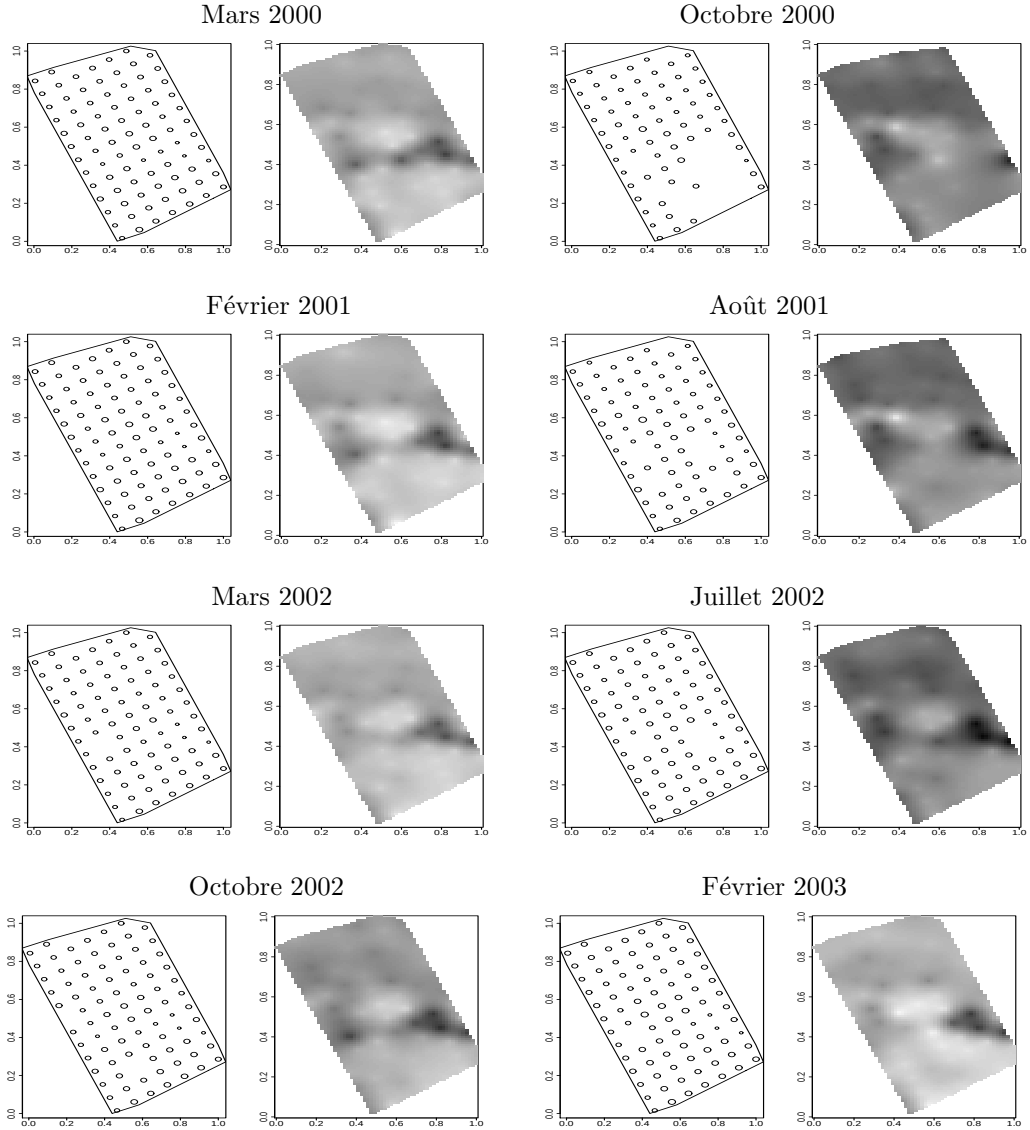


FIG. 13.3: *Points d'échantillonnage (proportionnels à leur valeur) et interpolation de la teneur en eau par krigeage ordinaire (valeurs fortes en blanc et faibles en noir). La même échelle est utilisée pour illustrer les interpolations.*

que dans sa partie nord. Une zone de très faible teneur en eau apparaît dans la partie centre-est de la parcelle pour toutes les dates, mais aussi dans la partie centre-ouest pour une partie des dates. Nous retrouvons sur ces interpolations la structure de la carte

pédologique (figure 13.2*a*). Les capacités de réserve en eau du sol sont très liées aux propriétés des matériaux constitutifs du sol.

La table 13.2 fournit le résumé statistique de la teneur en eau en fonction de la date. Les interpolations de la figure 13.3 sont représentées à la même échelle. Cette figure associée à la table 13.2 illustre la variabillité de la teneur en eau au cours du temps. Il apparaît

Date	Nombre de points	Etat de la culture
Mars 2000	85	semis de la betterave
Octobre 2000	67	récolte de la betterave et semis du blé
Février 2001	85	tallage du blé
Août 2001	79	récolte du blé
Mars 2002	84	semis du pois
Juillet 2002	82	récolte du pois
Octobre 2002	83	semis du blé
Février 2003	84	tallage du blé

TAB. 13.1: *Evolution culturale et nombre de points d'échantillonnage des variables dynamiques en fonction de la date de prélèvement.*

en particulier que la parcelle est de faible humidité aux mois d'octobre 2000, d'août 2001 et de juillet 2002. Ces dates correspondent aux périodes de récolte, les cultures ont donc eu récemment une forte consommation en eau. En revanche l'humidité est élevée en mars 2000, février 2001, mars 2002 et février 2003, *i.e.* aux périodes hivernales, correspondant à une faible consommation par les cultures par rapport au remplissage en eau des sols dû à la saison. Octobre 2000 et octobre 2002 constituent des dates intermédiaires (début de recharge de la réserve par pluies).

	Min.	1^{er} quartile	Médiane	Moyenne	$3^{ème}$ quartile	Max.	Ecart-type
Mars 2000	181.0	319.0	345.0	342.3	378.0	436.0	49.4
Octobre 2000	137.0	233.0	256.0	270.5	313.0	412.0	49.1
Février 2001	179.0	327.0	350.0	351.4	389.0	463.0	51.6
Août 2001	131.0	240.5	264.0	277.3	320.5	448.0	54.9
Mars 2002	201.0	327.3	355.5	351.6	384.3	433.0	45.7
Juillet 2002	98.0	238.8	268.5	267.9	307.8	352.0	52.7
Octobre 2002	157.0	296.5	322.0	321.6	360.0	405.0	51.5
Février 2003	168.0	337.8	364.0	360.5	402.3	452.0	55.2

TAB. 13.2: *Résumé statistique de la teneur en eau (en mm).*

Les points d'échantillonnage de la teneur en azote minéral et leur interpolation, sont représentés sur la figure 13.4. Cette figure montre que la teneur en azote minéral est un peu plus élevée dans le sud que dans le nord de la parcelle. Nous retrouvons moins nettement la structuration liée à la disposition des sols. La teneur en azote est liée aux propriétés physico-chimiques des sols, mais aussi à la culture (résidus de la culture précédente, prélèvements par la culture en place, ...). La table 13.3 fournit le résumé statistique de

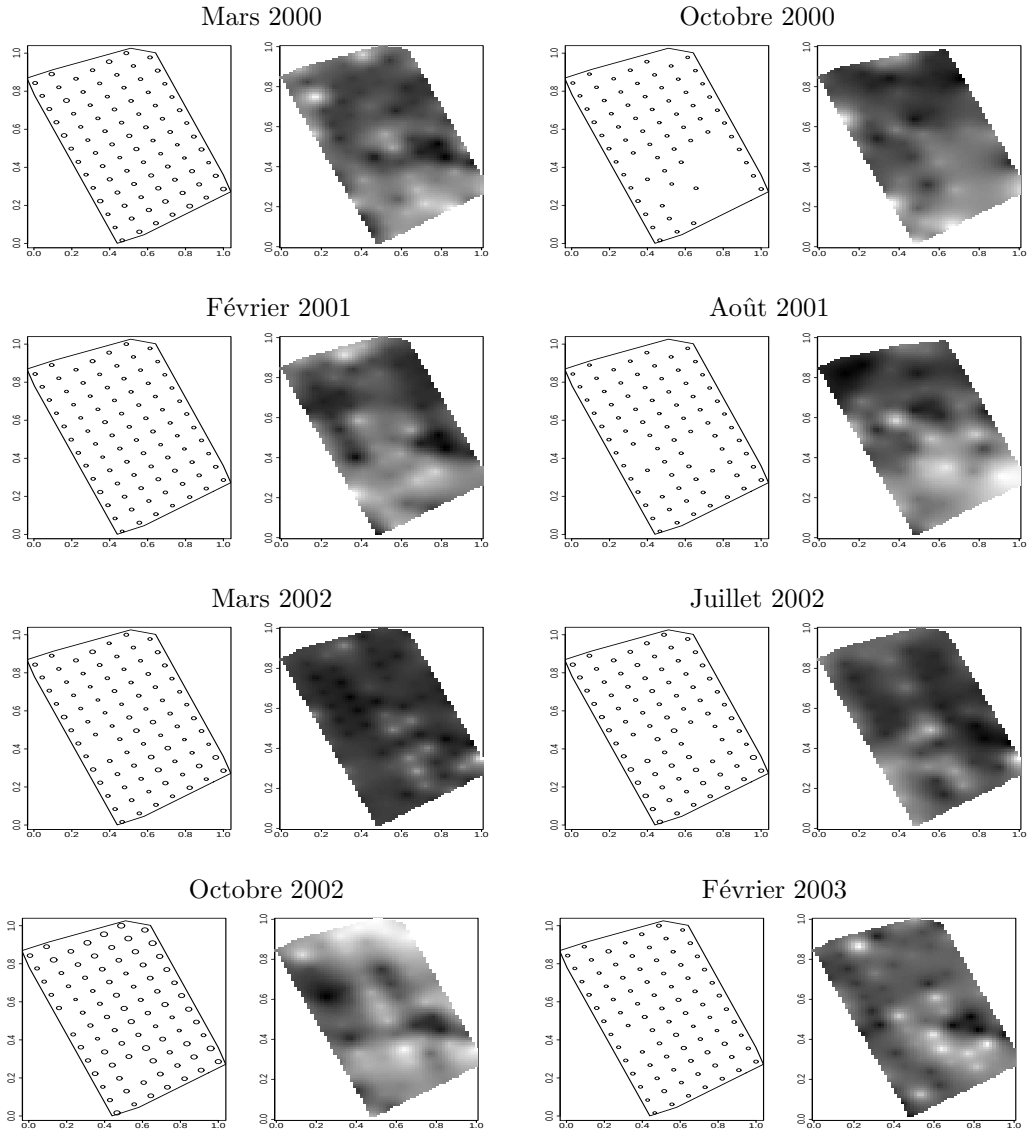


FIG. 13.4: *Points d'échantillonnage (proportionnels à leur valeur) et interpolation de la teneur en azote minéral (valeurs fortes en blanc, valeurs faibles en noir). Des échelles différentes sont utilisées pour illustrer les interpolations.*

cette variable. Nous pouvons remarquer que la teneur en azote minéral est très variable

	Min.	1er quartile	Médiane	Moyenne	3ème quartile	Max.	Ecart-type
Mars 2000	40.00	57.00	69.00	69.98	82.00	111.00	16.2
Octobre 2000	8.00	14.00	16.00	17.67	20.00	34.00	5.8
Février 2001	20.00	30.00	39.00	38.73	45.00	68.00	10.0
Août 2001	18.00	23.50	27.00	27.92	32.00	41.00	5.79
Mars 2002	36.00	50.75	61.00	63.81	73.25	111.00	17.8
Juillet 2002	19.00	35.00	45.00	47.73	56.00	101.00	15.1
Octobre 2002	47.00	91.00	110.00	108.70	128.00	159.00	25.9
Février 2003	18.00	31.75	38.00	38.50	44.00	69.00	11.4

TAB. 13.3: *Résumé statistique de la teneur en azote minéral (en kg/ha).*

d'une date à l'autre. Elle est très faible en octobre 2000 et août 2001, ces dates correspondent aux périodes de récolte, la culture a donc consommé une grande quantité d'azote (cela dépend de la culture en place, par exemple la betterave est plus consommatrice que le blé). La teneur en azote minéral est faible durant les périodes hivernales, février 2001 et février 2003. Durant ces périodes la réserve en eau est pleine et il y a un phénomène de lessivage. De plus l'engrais pour la culture n'a pas encore été apporté. Octobre 2002 présente la plus forte concentration en azote minéral, sans doute parce que l'apport d'engrais a récemment été effectué.

13.3 Variables permanentes

Les variables de constitution considérées sont la teneur en argile, en limon (fin et grossier), en sable (fin et grossier) et en calcaire. Ces variables sont sans unité, elles correspondent à des taux (en $^0/_{00}$). Ces variables ont été échantillonnées en les 84 points de la figure 13.2*b*. Le résumé statistique de ces variables est donné dans la table 13.4. La

	Min.	1er quartile	Médiane	Moyenne	3ème quartile	Max.	Ecart-type
Argile	98.0	158.5	182.5	191.2	225.3	294.0	43.8
Limon fin	157.0	216.0	224.0	223.9	232.0	286.0	16.7
Limon grossier	167.0	194.0	230.0	230.5	259.8	315.0	39.9
Sable fin	157.0	197.0	219.0	222.1	237.0	417.0	38.3
Sable grossier	19.0	32.00	43.0	52.71	71.25	125.0	26.0
Calcaire	11.0	32.75	174.5	176.0	291.3	439.0	134.9

TAB. 13.4: *Résumé statistique des variables permanentes.*

figure 13.5 représente les variables permanentes interpolées sur une grille 62×98 avec le krigeage ordinaire. Les valeurs fortes sont représentées en blanc et les valeurs faibles en noir. Cette figure montre que le nord de la parcelle est plus argileux que le sud. Le limon

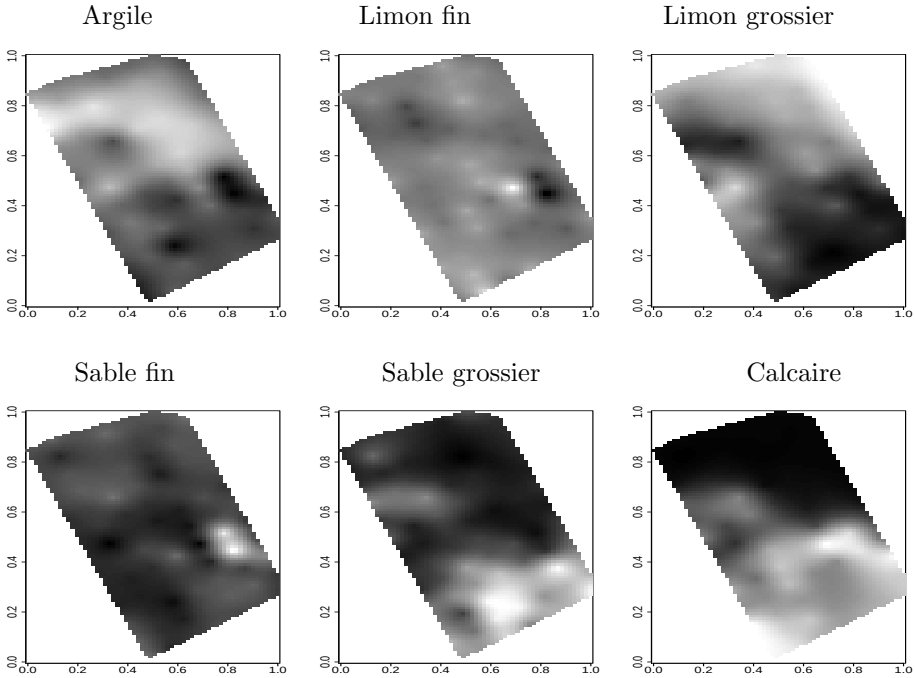


FIG. 13.5: *Interpolation des variables permanentes (valeurs élevées en blanc et valeurs faibles en noir).*

fin est réparti de façon assez homogène sur l'ensemble de la parcelle hormis près de la zone sur craie sableuse où il y a un contraste entre valeur forte et valeur faible. Le limon grossier s'avère plus élevé dans la partie nord que dans la partie sud. La zone sur craie sableuse révèle une forte concentration de sable fin. Le sable grossier est bien plus présent dans l'extrême sud de la parcelle. Enfin, comme nous l'avons vu précédemment, le nord est moins calcaire que le sud, et une forte concentration de calcaire apparaît en particulier sur la zone sur craie sableuse. Nous pouvons donc constater une nette opposition entre d'une part l'argile et d'autre part le calcaire et le sable grossier, qui sont les caractéristiques des zones du sol.

Chapitre 14

Résultats

Nous avons appliqué la méthode de détection de Zones de Changement Abrupt aux données décrites dans le chapitre 13 et utilisé l'amélioration de la procédure itérative correctrice du biais introduit dans l'estimation de la variance en présence d'une discontinuité (sections 7.3 et 8.4).

L'application de la procédure itérative décrite dans la section 7.3 nécessite l'estimation du modèle de covariance et la détermination du niveau $1-\alpha$ à chaque itération. Nous avons utilisé des modèles de covariance exponentiels : $\exp(-\mathbf{h}/b)$, $\mathbf{h} \geq 0$. Pour les raisons de rapidité de calcul évoquées dans la section 7.2, nous avons utilisé comme niveau de confiance $1-\alpha_G$ défini dans l'équation (7.3). En ce qui concerne le niveau $1-\alpha_0$ intervenant dans la procédure itérative, nous avons considéré $1-\alpha_0 = 0.99$

14.1 Variables dynamiques

Afin d'illustrer les différentes étapes de la procédure itérative, nous allons détailler les différentes itérations obtenues pour la teneur en eau pour une date, par exemple pour mars 2000.

La figure 14.3*a* représente les points d'échantillonnage. Chaque cercle est proportionnel à la valeur. La variable est interpolée par krigeage ordinaire sur une grille 62×98 (figure 14.3*b*). Un pixel représente un carré de $5m \times 5m$. Les valeurs faibles sont représentées en noir et les valeurs fortes en blanc. Les paramètres, portée b et variance σ^2, de chaque modèle de covariance sont estimés à chaque itération de la façon suivante. Compte tenu de la relation entre fonction de covariance et variogramme (équation (2.3)), nous avons dans un premier temps calculé le variogramme expérimental, puis nous avons estimé le

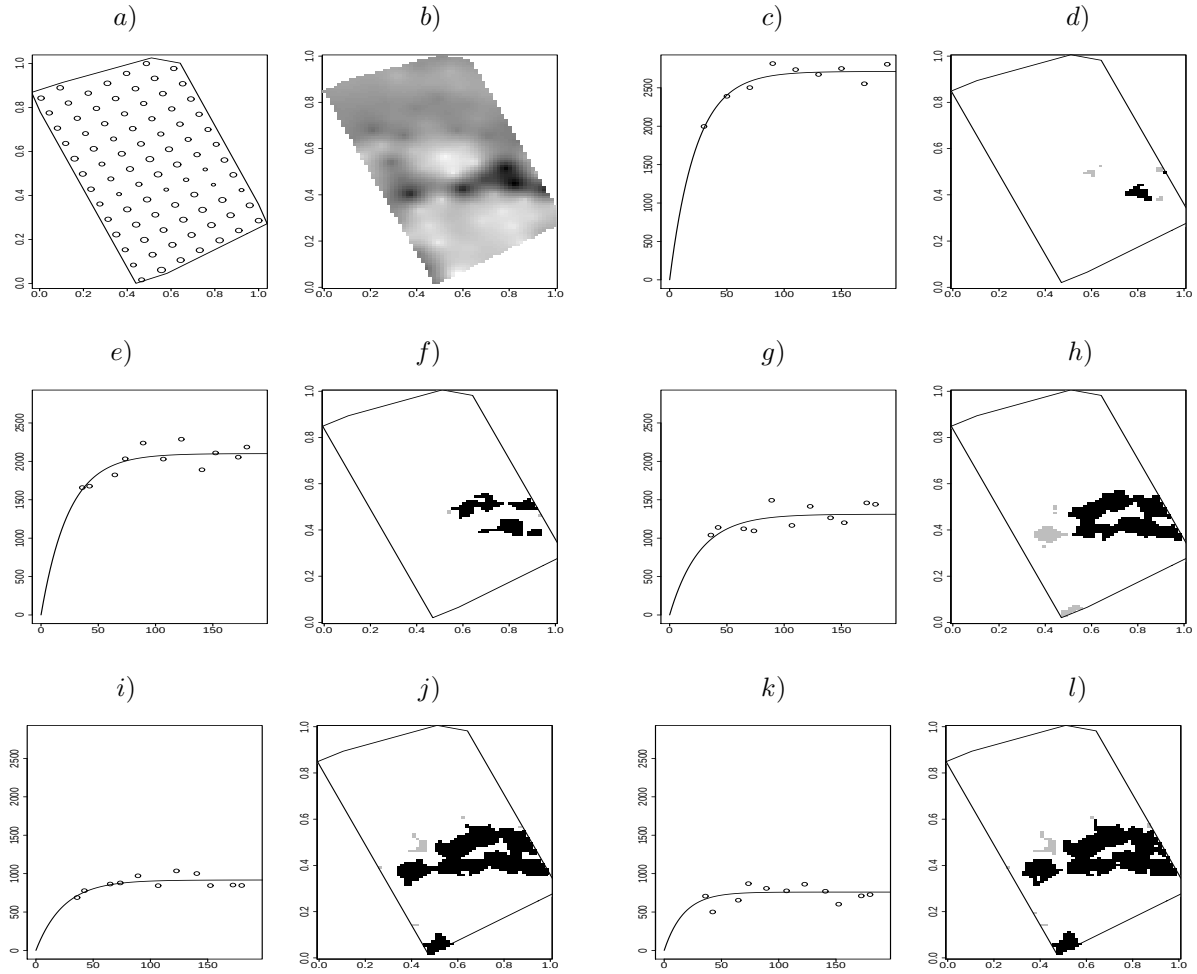


FIG. 14.1: *a) Points d'échantillonnage (proportionnels à leur valeur) de la teneur en eau en mars 2000. b) Interpolation sur une grille* 62×98 *par krigeage ordinaire. c) et d) (resp. e) et f), g) et h), i) et j), k) et l)) Ajustement du variogramme et ZCAs détectées à la première (resp. deuxième, troisième, quatrième, cinquième) itération. Les ZCAs significatives sont représentées en noir et les non significatives en gris.*

variogramme théorique à l'aide d'un processus de lissage utilisant les moindres carrés pondérés [Cressie (1993)]. Pour la première itération nous avons utilisé tous les couples de points pour calculer le variogramme expérimental. La figure 14.3*c* illustre le variogramme ajusté ($b = 23\ m$, $\sigma^2 = 2713\ mm^2$). Les ZCAs obtenues au niveau $1 - \alpha$ correspondant ($1 - \alpha_G = 0.9984$) sont représentées sur la figure 14.3*d*. Les ZCAs significatives apparaissent en noir et les ZCAs non significatives en gris. Afin d'estimer le variogramme à la deuxième itération, nous avons éliminé dans le calcul du variogramme expérimental les couples $\{Z(\mathbf{x}_i), Z(\mathbf{x}_j)\}$ pour lesquels le segment $[\mathbf{x}_i, \mathbf{x}_j]$ intersecte une ZCA au niveau $1 - \alpha_0$ contenant une ZCA significative au niveau $1 - \alpha_G$ (figure 8.3). Les figures 14.3*e* et 14.3*f* illustrent le variogramme estimé et les ZCAs obtenues au niveau $1 - \alpha_G = 0.9983$

lors de la deuxième itération. La même procédure est appliquée pour les itérations suivantes. Les figures 14.3*g* et 14.3*h*, 14.3*i* et 14.3*j*, 14.3*k* et 14.3*l*. montrent les résultats obtenus pour, respectivement, les troisième, quatrième et cinquième itérations. La table 14.1 fournit les paramètres b, σ^2 et $1-\alpha_G$ évalués pour l'ensemble des itérations. Les

Itération	b	σ^2	$1-\alpha_G$
1	23	2713	0.9984
2	24	2101	0.9983
3	26	1312	0.9980
4	24	917	0.9983
5	17	760	0.9991

TAB. 14.1: *Portée b (en m), variance (en mm^2) et niveau $1-\alpha_G$ évalués lors de la procédure itérative pour la teneur en eau en mars 2000.*

figures 14.3*j* et 14.3*l* illustrent la convergence : les ZCAs varient très peu d'une itération à l'autre (le pattern est le même, seule la surface des ZCAs varie légèrement). La convergence a donc lieu en quatre itérations. Plus nous avançons dans le processus itératif, plus le palier diminue (table 14.1), ce qui caractérise l'erreur introduite dans l'estimation de la variance par la présence de discontinuités. Dans cet exemple la variance a été divisée par trois entre la première itération et la convergence.

Utiliser le niveau $1-\alpha_G$ au lieu du niveau $1-\hat{\alpha}$ n'a pas d'effet sur les résultats. En effet nous avons également appliqué la procédure itérative en estimant le niveau $1-\hat{\alpha}$ à chaque itération. Les variogrammes estimés restent les mêmes pour l'ensemble des itérations et nous obtenons la convergence à la même itération. Seuls les niveaux $1-\alpha$ diffèrent comme le montre la table 14.2. Les ZCAs obtenues à la convergence sont représentées sur la figure

$1-\alpha$	Itération	1	2	3	4
$1-\alpha_G$		0.9984	0.9983	0.9980	0.9983
$1-\hat{\alpha}$		0.9992	0.9989	0.9985	0.9989

TAB. 14.2: *Niveaux de confiance associés au test local de détection obtenus aux différentes itérations pour la teneur en eau en mars 2000.*

14.2. Cette figure illustre les ZCAs obtenues en utilisant $1-\alpha = 1-\alpha_G$ (figure 14.2*a*) et $1-\alpha = 1-\hat{\alpha}$ (figure 14.2*b*). Nous pouvons constater que les ZCAs détectées sont

identiques à quelques pixels près.

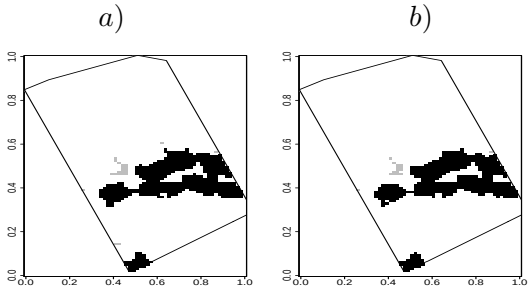


FIG. 14.2: *ZCAs obtenues à la convergence pour la teneur en eau en mars 2000, avec : a)* $1-\alpha = 1-\alpha_G$*, b)* $1-\hat{\alpha}$ *dans la procédure itérative.*

La table 14.3 contient les valeurs de la portée b, de la variance σ^2 et du niveau $1-\alpha_G$ obtenus à la première itération et à la convergence pour la teneur en eau en considérant l'ensemble des dates. Cette table montre que lorsque la procédure converge toujours en

	Itération 1			Convergence			
Date	b	σ^2	$1-\alpha_G$	Itération	b	σ^2	$1-\alpha_G$
Mars 2000	23	2713	0.9984	4	24	917	0.9983
Octobre 2000	26	2509	0.9980	1	-	-	-
Février 2001	27	2929	0.9982	2	26	1645	0.9980
Août 2001	27	3129	0.9978	4	23	688	0.9984
Mars 2002	23	2209	0.9984	1	-	-	-
Juillet 2002	32	2952	0.9970	1	-	-	-
Octobre 2002	27	2906	0.9978	4	43	892	0.9946
Février 2003	25	3388	0.9982	3	46	1169	0.9938

TAB. 14.3: *Portée b (en m), variance* σ^2 *(en* mm^2*) et niveau local de confiance* $1-\alpha_G$ *obtenus à la première itération et à la convergence pour la teneur en eau.*

moins de cinq itérations (de une à quatre) et lorsqu'il n'y a pas convergence à la première itération, la variance estimée à la convergence est de 2 à 4 fois plus petite que celle estimée à la première itération. La figure 14.3 représente les ZCAs obtenues à la convergence pour les différentes dates, avec en noir les ZCAs significatives et en gris les non significatives. Des ZCAs sont détectées à toutes les dates dans la partie centre-est de la parcelle, sauf pour octobre 2000 probablement à cause de l'absence de points d'échantillonnage dans la partie sud-est de la parcelle. Les ZCAs détectées sont de surface variable en fonction de la

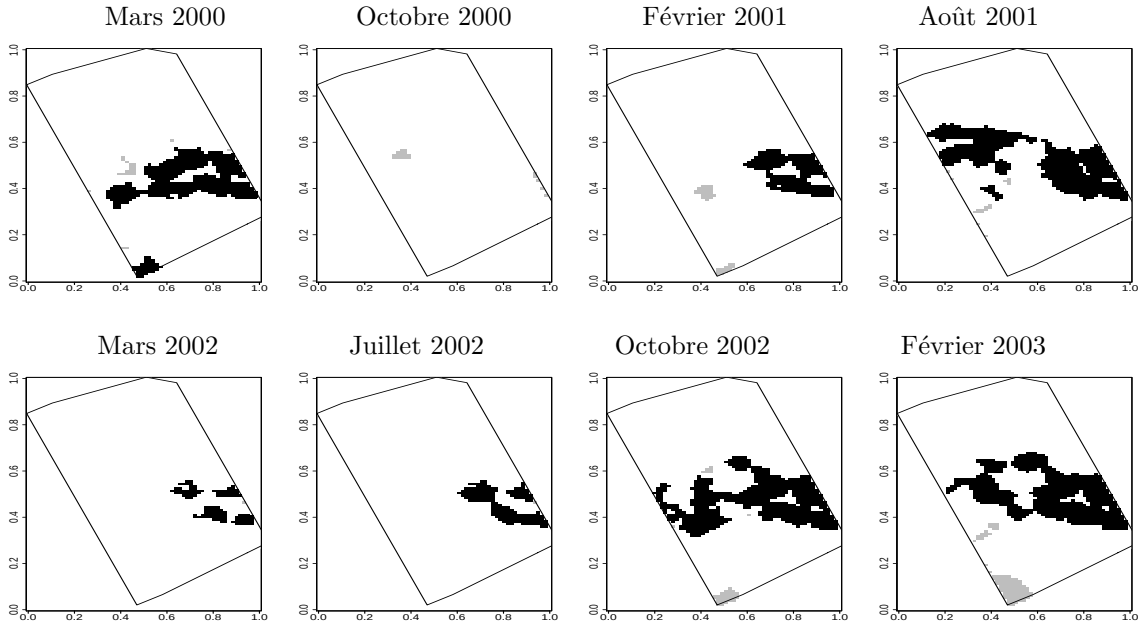


FIG. 14.3: *ZCAs (significatives en noir, non significatives en gris) détectées pour la teneur en eau à la convergence.*

date. Le manque de recul (en terme d'années) ne permet pas de dire si la surface des ZCAs semble plus liée à la succession culturale qu'à l'état hydrique de la parcelle. Nous avons reporté dans la table 14.4 les surfaces et les valeurs critiques des ZCAs obtenues pour la teneur en eau lors de la convergence. Ces résultats nous rappellent que la valeur critique associée à une ZCA est liée à sa surface : plus la surface est grande, plus la valeur critique est petite. Lorsque les ZCAs sont très grandes la valeur critique associée peut prendre des valeurs proches de zéro. Remarquons cependant que la valeur critique, p, ne dépend pas uniquement de la surface S de la ZCA. Elle est aussi fonction du déterminant de la matrice de courbure, $\mathbf{\Lambda}$, du champ de χ^2 en le maximum $\mathbf{x}_{max}$ de la ZCA et du niveau $1-\alpha$: $p = \exp\left(-t_\alpha \det\left(\mathbf{\Lambda}(\mathbf{x}_{max})\right)^{1/2} S/2\pi\right)$. Aussi il peut arriver qu'en comparant les valeurs critiques à une date donnée (et donc à niveau $1-\alpha$ fixé) nous ayons des valeurs plus faibles pour des surfaces plus grandes. C'est par exemple le cas en août 2001, où nous avons deux grandes ZCAs : une de 8925 m^2 et une autre de 7400 m^2. Les valeurs critiques associées à ces ZCAs sont respectivement 5.10^{-13} et 4.10^{-21} car la valeur du déterminant de la matrice $\mathbf{\Lambda}$ est quatre fois plus petite pour la plus grande ZCA que pour la plus petite.

Pour la teneur en azote minéral, la méthode ne détecte aucune ZCA (sauf dans la

Date									
Mars 2000	S	*950*	*50*	*11650*	*25*	*350*	*25*	*50*	*25*
	p	2.10^{-3}	0.55	7.10^{-39}	0.81	0.052	0.55	0.59	0.59
Octobre 2000	S	*25*	*25*	*25*	*50*	*375*			
	p	0.80	0.82	0.77	0.25	0.32			
Février 2001	S	*350*	*625*	*6100*					
	p	0.24	0.22	10^{-29}					
Août 2001	S	*25*	*25*	*225*	*450*	*8925*	*100*	*25*	*7400*
	p	0.93	0.75	0.21	9.10^{-3}	5.10^{-13}	0.26	0.83	4.10^{-21}
Mars 2002	S	*600*	*950*	*25*	*850*	*575*			
	p	10^{-4}	6.10^{-4}	0.87	6.10^{-3}	6.10^{-33}			
Juillet 2002	S	*3725*	*25*	*750*					
	p	2.10^{-26}	0.93	7.10^{-33}					
Octobre 2002	S	*825*	*16650*	*25*	*25*	*150*			
	p	0.16	0	0.18	0.90	0.67			
Février 2003	S	*2050*	*25*	*25*	*25*	*225*	*200*	*16825*	
	p	0.25	0.94	0.95	0.93	0.49	0.73	10^{-14}	

TAB. 14.4: *Surfaces, S, (en m^2) et valeurs critiques, p, des ZCAs obtenues lors de la convergence pour la teneur en eau.*

partie centre-ouest de la parcelle en août 2001). Ce résultat est associé aux commentaires de la figure 13.4.

14.2 Variables permanentes

La même procédure a été appliquée aux variables permanentes. La convergence a été atteinte entre la première et la troisième itération selon la variable. Les ZCAs obtenues à la convergence sont représentées sur la figure 14.4, avec en noir les ZCAs significatives et en gris les ZCAs non significatives. Pour l'argile, le limon fin, le sable fin et le calcaire une ZCA est détectée dans la partie centre-est de la parcelle. Une ZCA est également détectée dans la partie centre-ouest de la parcelle pour l'argile. En revanche, aucune ZCA significative n'est détectée pour le limon grossier et le sable grossier.

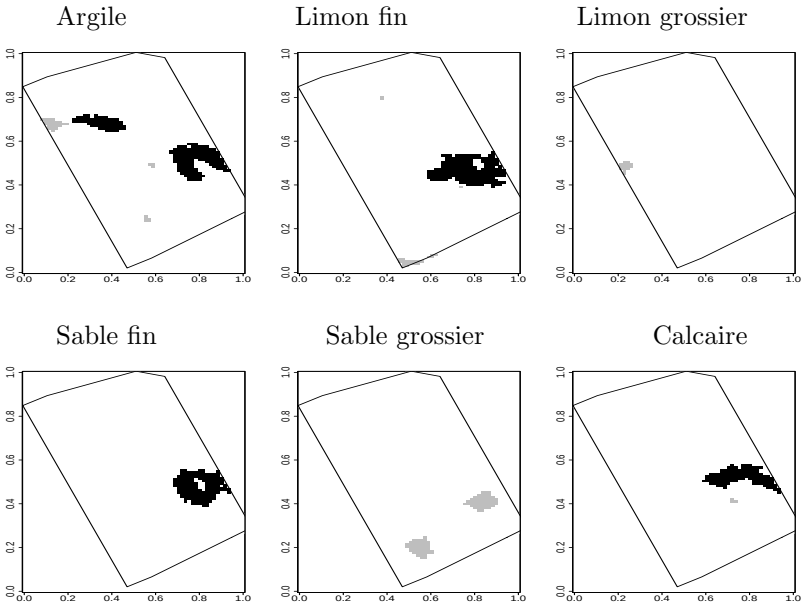


FIG. 14.4: *ZCAs (significatives en noir, non significatives en gris) détectées pour les variables permanentes à la convergence.*

14.3 Validation de la méthode sur les données

Dans cette section, nous testons la sensibilité de la méthode, non plus sur des simulations (comme dans le chapitre 12), mais sur des données.

Sensibilité par rapport à une variation des paramètres de la fonction de covariance

Variation de la portée

Nous avons appliqué la procédure itérative à la teneur en eau en mars 2000, en considérant d'une part le modèle estimé (modèle exponentiel de portée $b = 23\ m$ et de variance $\sigma^2 = 2713\ mm^2$) et d'autre part deux autres modèles de covariance, de même forme paramétrique que le modèle estimé, mais de portée $b \pm 30\%$. Nous avons au cours de la procédure utilisé les niveaux $1-\alpha = 1-\hat{\alpha}$ et $1-\alpha_0 = 0.99$. La table 14.5 fournit les paramètres b et σ^2 estimés aux différentes itérations selon le modèle de covariance initial utilisé. Cette table montre que quelquesoit le modèle de covariance initial, les paramètres estimés de la fonction de covariance pour les modèles initiaux "erronés" correspondent à ceux obtenus pour le modèle estimé. En effet, les paramètres de la troisième itération du modèle $b = 16\ m$ correspondent à ceux de la troisième itération du modèle estimé

	Itération 1		Itération 2		Itération 3		Itération 4		Itération 5	
Modèle	b	σ^2	b	σ^2	b	σ^2	b	σ^2	b	σ^2
$b = 16$	16	2713	25	2168	26	1312	24	917	17	760
$b = 23$	23	2713	24	2101	26	1312	24	917	17	760
$b = 30$	30	2713	27	1322	24	917	17	760	-	-

TAB. 14.5: *Portée b (en m) et variance σ^2 (en mm^2) des fonctions de covariance estimées aux différentes itérations, en fonction du modèle de covariance initial.*

($b = 23\ m$) et les paramètres la troisième itération du modèle $b = 30\ m$ correspondent à ceux de la quatrième itération du modèle estimé. Par conséquent, la procédure convergence vers, exactement, le même ensemble de ZCAs que pour pour le modèle $b = 23\ m$.

Variation du modèle paramétrique

Lors de l'étude de simulations nous avons constaté (table 12.1) que si la portée était bien estimée il était tout à fait raisonnable de se servir des modèles de covariance sphérique. La procédure itérative a donc été appliquée à la teneur en eau en juillet 2002, en considérant d'une part le modèle estimé (modèle exponentiel de portée $b = 23\ m$ et de variance $\sigma^2 = 2713\ mm^2$) et d'autre part un modèle sphérique de paramètres estimés : $b = 68\ m$ et $\sigma^2 = 2860\ mm^2$. Nous avons au cours de la procédure itérative utilisé les niveaux $1-\alpha = 1-\hat{\alpha}$ et $1-\alpha_0 = 0.99$. La convergence a lieu à la première itération pour

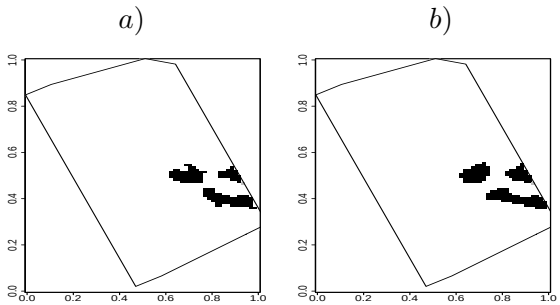


FIG. 14.5: *ZCAs (significatives en noir, non significatives en gris) obtenues pour la teneur en eau en juillet 2002, avec un modèle de covariance a) exponentiel, au niveau $1-\hat{\alpha} = 0.9985$; b) sphérique, au niveau $1-\hat{\alpha} = 0.9964$.*

les deux modèles. La figure 14.5 représente les ZCAs (significatives en noir) obtenues à la convergence pour le modèle exponentiel (figure 14.5*a*) et le modèle sphérique (14.5*b*). Nous pouvons constater que les ZCAs détectées sont les mêmes à quelques pixels près.

Effet de la discrétisation

Nous avons également comparé les résultats de la teneur en eau pour juillet 2002 obtenus pour différentes discrétisations. La figure 14.6 illsutre les ZCAs (significatives en noir, non significatives en gris) obtenues au niveau $1-\alpha_G =$ (première ligne) et au niveau $1-\hat{\alpha}$ (deuxième ligne) avec une grille 31×49, 1 pixel $= 100\,m^2$ (figures 14.6*a* et 14.6*d*), une grille 62×98, 1 pixel $= 25\,m^2$ (figures 14.6*b* et 14.6*e*) et avec une grille 155×245, 1 pixel $= 4\,m^2$ (figures 14.6*c* et 14.6*f*). La discrétisation n'a pas d'effet sur la détection des ZCAs.

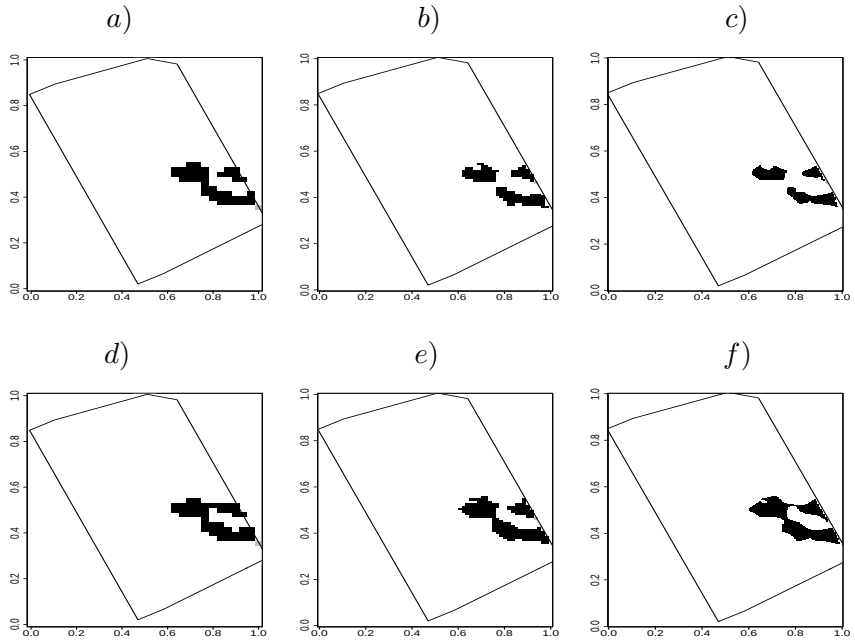


FIG. 14.6: *ZCAs (significatives en noir, non significatives en gris) obtenues pour la teneur en eau en juillet 2002, au niveau* $1-\alpha_G$ *(première ligne) ou* $1-\hat{\alpha}$ *(deuxième ligne), avec différentes discrétisations :* 31×49 *(a) et d)),* 62×98 *(b) et e)) et* 155×245 *(c) et f)).*

Effet du choix de l'interpolateur

Nous avons appliqué la méthode à la teneur en eau en juillet 2002 en utilisant différents interpolateurs. Pour les deux interpolateurs la convergence a lieu à la première itération. La figure 14.7 représente les ZCAs, au niveau $1 - \alpha_G$, obtenues à la convergence. Cette figure montre que les ZCAs détectées sont identiques que nous utilisions le krigeage ordinaire (figure 14.7*a*) ou le krigeage simple (figure 14.7*b*).

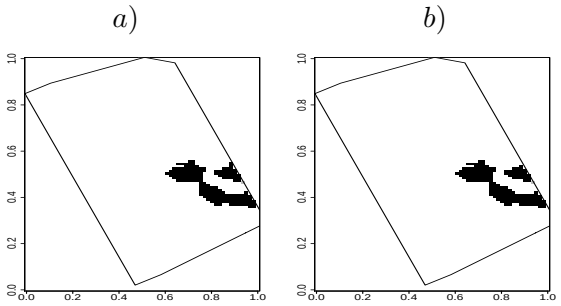


FIG. 14.7: *ZCAs (significatives en noir) obtenues pour la teneur en eau en juillet 2002 au niveau* $1-\alpha_G$*, en utilisant comme interpolateur a) le krigeage ordinaire, b) le krigeage simple.*

14.4 Effet de la densité locale de l'échantillon

Nous illustrons l'influence de la densité de points de l'échantillon sur la détection de ZCAs. Considérons la teneur en eau en juillet 2002. A cette date l'échantillon comporte 82 points. Nous avons retiré aléatoirement $m\%$, $m \in \{10, 20, 30, 40, 50, 60, 70\}$, des points de l'échantillon initial. Les échantillons obtenus sont représentés sur la figure 14.8. Les

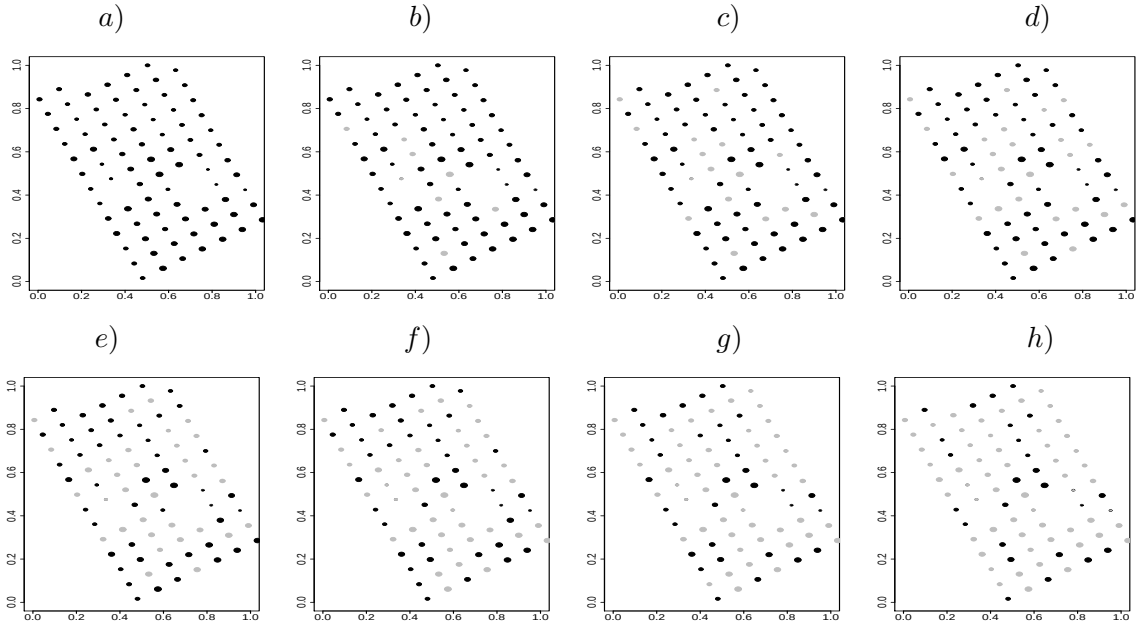


FIG. 14.8: *Echantillon de points de la teneur en eau en juillet 2002 obtenus en éliminant* m *% des points de l'échantillon initial. a)* $m = 0$*, b)* $m = 10$*, c)* $m = 20$*, d)* $m = 30$*, e)* $m = 40$*, f)* $m = 50$*, g)* $m = 60$*, h)* $m = 70$*. Les points gris correspondent aux points éliminés et les points noirs aux points conservés.*

points gris représentent les points qui ont été éliminés de l'échantillon initial et les points noirs correspondent aux points restants. Pour chaque échantillon de points, les ZCAs

ont été estimées sur une grille de dimension 62×98. Nous avons appliqué la procédure itérative en utilisant les niveaux $1-\alpha = 1-\alpha_G$ et $1-\alpha_0 = 0.99$ et en considérant deux cas.

Cas I : Les variogrammes de la première itération ont été estimés pour chaque échantillon. Les résultats de la procédure itérative sont donnés dans la table 14.6. La procédure

		Première itération			Convergence			
Nombre de points	m	b	σ^2	$1-\alpha_G$	Itération	b	σ^2	$1-\alpha_G$
82	0	32	2952	0.9970	1	-	-	-
74	10	31	2946	0.9972	1	-	-	-
66	20	31	3221	0.9972	1	-	-	-
57	30	30	3485	0.9974	2	46	1695	0.9938
49	40	34	3850	0.9966	2	58	1860	0.9902
41	50	52	4816	0.9921	1	-	-	-
33	60	49	5566	0.9929	2	55	3974	0.9912
25	70	41	4361	0.9951	1	-	-	-

TAB. 14.6: *Portée, b (en m), variance, σ^2 (en mm^2) et niveau $1-\alpha_G$ obtenus lors de la procédure itérative appliquée à la teneur en eau en juillet 2002, pour des échantillons obtenus en éliminant m % des points de l'échantillon initial.*

converge en maximum deux itérations. Les résultats de la table 14.6 montrent que lorsque environ la moitié des points a été éliminée, la portée et la variance estimées à la première itération augmentent. Lorsque la convergence est atteinte à la deuxième itération, la variance obtenue diminue de 30 à 50 % par rapport à la variance estimée lors de la première itération.

Les ZCAs au niveau $1-\alpha_G$ obtenues à la convergence sont illustrées sur la figure 14.9. Les ZCAs significatives sont représentées en noir et les ZCAs non significatives en gris. Cette figure montre que les ZCAs détectées pour l'échantillon initial ($m = 0$) sont encore détectées malgré le retrait de 60 % des points de l'échantillon initial. Une ZCA non significative est détectée lorsque 70 % des points ont été retirés. Nous pouvons constater que les ZCAs détectées lorsque $30 \leq m \leq 60$ sont plus larges que dans les cas où $m \geq 20$. Cela vient du fait que d'une part pour ces échantillons la fonction de covariance estimée à la convergence a une variance plus petite et une portée plus grande (et donc un niveau $1-\alpha_G$ plus faible) et que d'autre part lorsque l'échantillon de points est moins dense les ZCAs ont tendance à être plus étalées et plus larges (section 11.2).

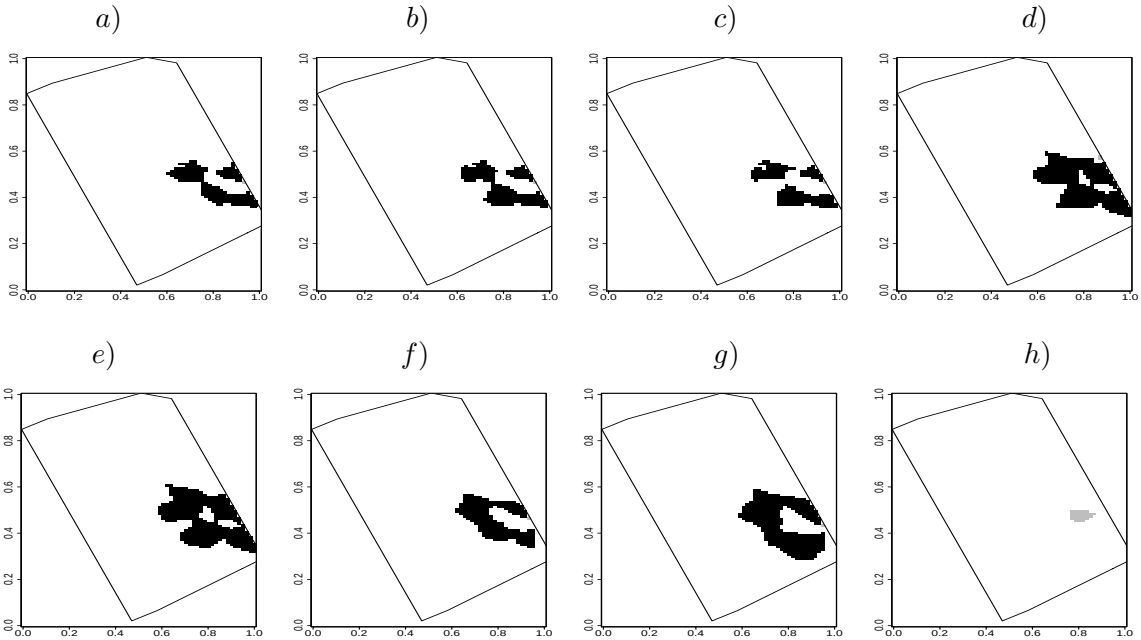


FIG. 14.9: *ZCAs (significatives en noir, non significatives en gris) obtenues à la convergence du cas I pour la teneur en eau en juillet 2002 en éliminant m % des points de l'échantillon initial. a) $m = 0$, b) $m = 10$, c) $m = 20$, d) $m = 30$, e) $m = 40$, f) $m = 50$, g) $m = 60$, h) $m = 70$.*

Cas II : Les variogrammes de la première itération sont les mêmes pour chaque échantillon. Le modèle considéré est celui estimé avec les 82 points.

Dans ce cas, la procédure itérative converge toujours à la première itération. Les ZCAs au niveau $1 - \alpha_G$ sont représentées en noir sur la figure 14.10. Cette figure montre que, comme pour le cas I, les ZCAs détectées pour l'échantillon initial sont encore détectées malgré le retrait de 60 % des points de l'échantillon initial. En revanche, dans le cas II la ZCA est toujours détectée lorsque 70 % des points ont été éliminés, mais elle n'a plus la même forme (puisque dépendante de la configuration des points d'échantillonnage). Les figures 14.10*a* à 14.10*g* illustrent les résultats de la section 11.2 : moins l'échantillon de points est dense, moins nous ne pouvons estimer les ZCAs avec précision et donc plus les ZCAs détectées sont larges.

Le cas I est utilisé lorsque nous n'avons aucune connaissance de la variable sur la parcelle. En revanche, le cas II permet, lorsque nous connaissons les variations spatiales de la variable, par exemple à une saison donnée, de n'échantillonner que peu de points l'année suivante, en privilégiant les zones où il existe des ZCAs et de façon à ce que l'échantillonnage soit homogène sur le reste de la parcelle, afin d'en détecter les (éventuelles) fortes discontinuités.

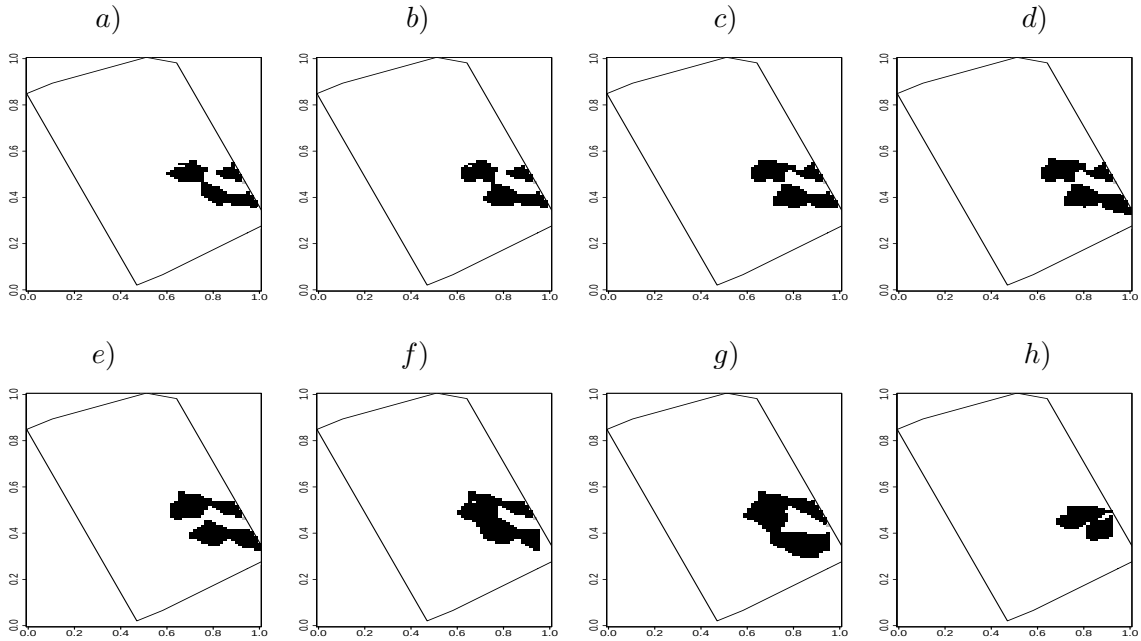


FIG. 14.10: *ZCAs (significatives en noir) obtenues à la convergence du cas II pour la teneur en eau en juillet 2002 en éliminant m % des points de l'échantillon initial. a)* $m = 0$*, b)* $m = 10$*, c)* $m = 20$*, d)* $m = 30$*, e)* $m = 40$*, f)* $m = 50$*, g)* $m = 60$*, h)* $m = 70$*.*

Nous avons donc pu illustrer que seuls 33 points sur une parcelle de 10 hectares suffisent pour pouvoir détecter des ZCAs. Cela tient au fait que d'une part les points échantillonnés contiennent la discontinuité et que d'autre part nous avons une couverture assez homogène du domaine et un échantillonnage de toutes les distances. Or ces deux conditions sont absentes pour octobre 2000, ce qui explique qu'aucune ZCA n'est été détectée.

Nous avons calculé la puissance conditionnelle locale (section 10.4) en utilisant une fenêtre de taille 7×7. Sur la figure 14.11, nous avons représenté la puissance pour une discontinuité de 3 écarts-type (ordre de grandeur des ZCAs détectées pour la teneur en eau en juillet 2002). Les points d'échantillonnage sont superposés sur ces images. Les bords blancs correspondent aux pixels en lesquels la puissance n'a pas pu être calculée, car ces pixels sont le centre d'une fenêtre contenant des données manquantes (extérieures au domaine). La figure 14.11 montre que la puissance dépend de la densité locale de l'échantillon. Moins l'échantillon est localement dense, plus la puissance est faible.

L'étude de la sensibilité de la détection de ZCAs par rapport à la densité de points d'échantillonnage permet d'expliquer pourquoi aucune ZCA n'est détectée en octobre

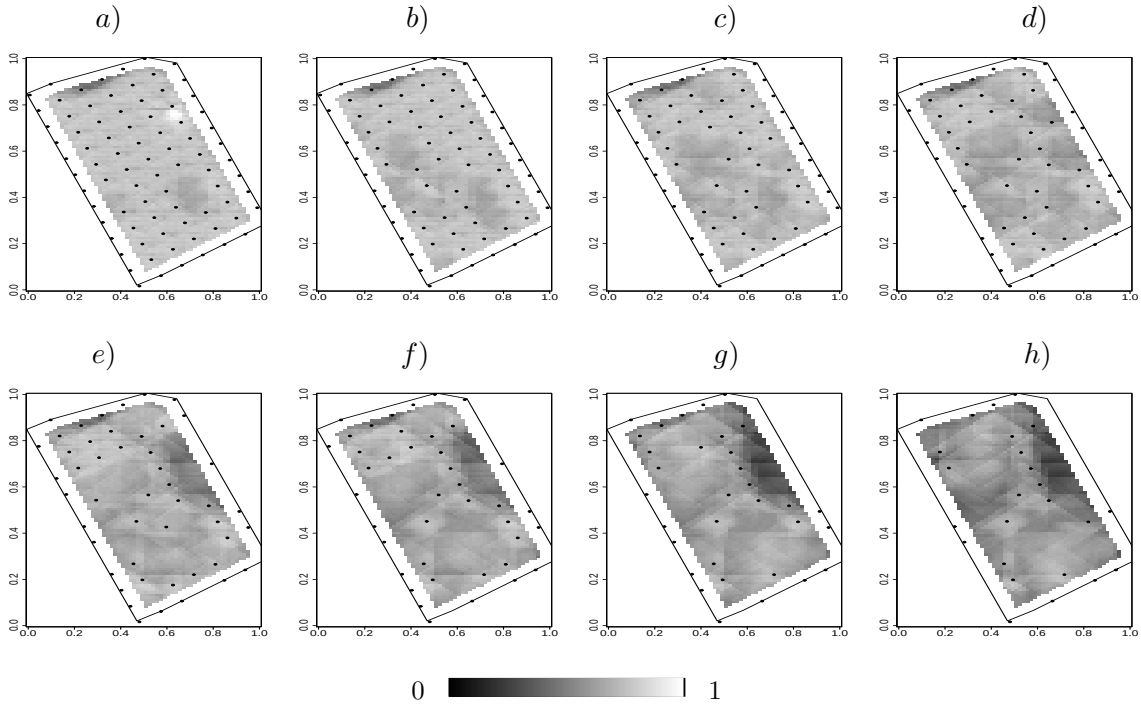


FIG. 14.11: *Puissance calculée en le centre d'une fenêtre de dimension* 7×7*, pour une discontinuité de 3 écarts-type pour l'échantillon de la teneur en eau en juillet 2002 en éliminant* m *% des points de l'échantillon initial. a)* $m = 0$*, b)* $m = 10$*, c)* $m = 20$*, d)* $m = 30$*, e)* $m = 40$*, f)* $m = 50$*, g)* $m = 60$*, h)* $m = 70$*.*

2000. A cette date, l'échantillon de points n'est pas homogène sur la parcelle : la partie sud-est de la parcelle contient une grande zone sans point d'échantillonnage. Nous avons calculé, pour mars 2000 et octobre 2000, la puissance en utilisant une fenêtre de dimension $(2k+1) \times (2k+1)$, avec $k \in \{0, 1, 3, 5\}$, pour une discontinuité de 3 écarts-type. Les cartographies de ces puissances sont représentées sur la figure 14.12. Ces figures montrent que lorsque l'échantillon de points est assez dense, comme pour mars 2000 (figure 14.12 première ligne), la puissance ne dépend que de la taille de la fenêtre. Par contre, lorsque l'échantillon présente des zones faiblement échantillonnés, comme pour octobre 2000 (figure 14.12 deuxième ligne), la puissance dépend aussi de la densité locale de l'échantillon.

Comme les ZCAs détectées à la première itération de chacune des sept autres dates se situent précisément dans la zone de faible puissance, si de telles ZCAs existaient pour octobre 2000, elles n'auraient pas pu être détectées. Aussi, les zones de faible puissance définissent les lieux où l'échantillonnage est localement trop faible pour détecter les éventuelles ZCAs.

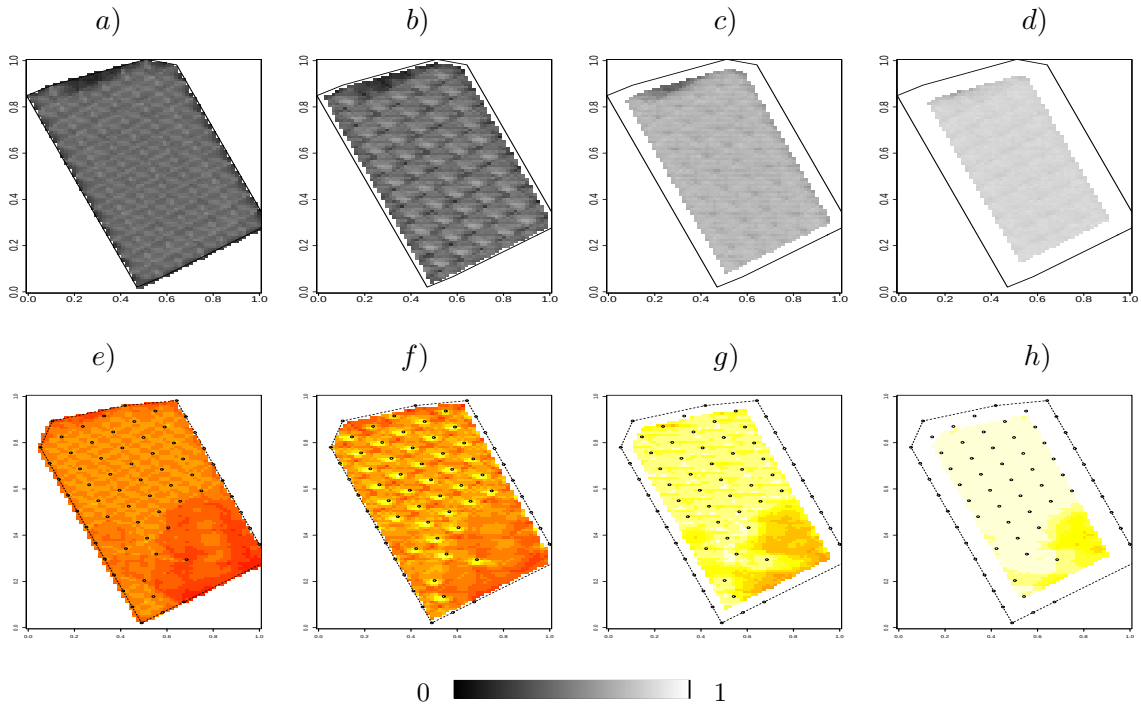


FIG. 14.12: *Cartographie de la puissance calculée pour mars 2000 (première ligne) et pour octobre 2000 (deuxième ligne) pour une discontinuité de 3 écarts-type et pour différentes tailles de la fenêtre $dim(\mathcal{F})$: a) et e) puissance ponctuelle, b) et f) $dim(\mathcal{F}) = 3\times3$, c) et g) $dim(\mathcal{F}) = 7\times7$, d) et h) $dim(\mathcal{F}) = 11\times11$.*

Discussion

La méthode de détection de ZCAs se révèle facilement applicable et interprétable en pratique. Elle présente un réel avantage par rapport aux méthodes telles que le wombling et ses dérivées (section 1.1) ou la méthode proposée par Hall *et al.* (2001) (section 1.2), en le sens que ses résultats dépendent des données : ces méthodes imposent l'existence de discontinuités. Nous avons par exemple pu voir pour la teneur en azote minéral qu'aucune ZCA n'est détectée, car les données ne présentent pas de gradient suffisamment élevé pour que ce soit le cas.

La méthode de détection de Zones de Changement Abrupt a permi de faire ressortir un caractère permanent de la parcelle : elle possède au centre-est une zone relativement moins humide que le reste. Une comparaison visuelle entre ces résultats (figure 14.3) et la carte pédologique (figure 13.2*a*) permet de constater que la méthode met en avant les sols sur sables grésifiés. Pour certaines dates ces ZCAs sont prolongées jusqu'au centre-ouest de la parcelle, et marquent une transition entre des sols calcaires et non calcaires. La zone sur sable grésifiée est aussi détectée avec des variables de constitution, telles que l'argile, le limon fin, le sable fin et la calcaire.
Cette analyse indique que la parcelle peut être divisée en deux zones de gestion : une au nord et une au sud. Une troisième zone est définie, mais est bien plus petite que les deux autres, ce qui d'un point de vue de gestion de la parcelle n'est pas nécessairement intéressant.

La figure 14.3 montre que les ZCAs peuvent avoir une surface assez importante par rapport à la surface de la parcelle (par exemple pour mars 2000). Mais ces ZCAs sont tout à fait justifiées compte tenu de la complexité pédologique de la parcelle (figure 13.1) et de la relation entre la teneur en eau et les propriétés des matériaux constitutifs du sol. De plus, l'échantillon de points n'étant pas très dense, cela implique un manque de précision sur la délimitation des ZCAs et l'obtention de ZCAs "épaisses".

Conclusion

Ce travail s'inscrit dans le cadre de la détection de ruptures dans le plan, le problème étant de détecter des Zones de Changement Abrupt dans des données faiblement et irrégulièrement échantillonnées dans un domaine $\mathcal{D}$ de $\mathbb{R}^2$. Les ZCAs sont définies comme des régions contigües où l'espérance du champ aléatoire modélisant la variable présente des discontinuités.

14.5 Bilan

La première partie de ce travail a consisté à étudier les méthodes existantes ayant une problématique proche de la nôtre. Le bilan de cette étude est qu'il existe une vaste littérature pour la détection de ruptures en traitement d'images, mais assez peu de méthodes statistiques dans le cas de données spatiales. De plus ces méthodes présentent certains inconvénients, elles supposent implicitement l'existence de discontinuités [Hall *et al.* (2001)], ne testent pas la significativité des ZCAs détectées (wombling et ses dérivées) ou nécessitent de lourds calculs et une représentation graphique difficile à interpréter [Godtliebsen *et al.* (2002)].
La méthode que nous proposons dans la deuxième partie répond à tous ces problèmes. Elle est basée sur un test statistique, dont l'hypothèse nulle est la stationnarité de l'espérance du champ aléatoire contre l'alternative d'existence de discontinuités de l'espérance. Une étape préliminaire consiste à interpoler la variable sur $\mathcal{D}$ via un outil géostatistique, le krigeage. La méthode se décompose en deux étapes. La première estime les ZCAs potentielles, *i.e.* en chaque point $\mathbf{x}$ de $\mathcal{D}$ (ou, en pratique, en chaque pixel de la grille d'interpolation), nous testons localement s'il y a un changement abrupt. Sous l'hypothèse de gaussianité du champ aléatoire et sous certaines conditions de régularité de sa fonction de covariance, la statistique de test, $T(\cdot)$, est une forme quadratique de l'interpolation du gradient du champ normalisé par sa variance. Le champ aléatoire $T(\cdot)$ a pour distribution une χ^2 à deux degrés de liberté et est non stationnaire (à cause de la non stationnarité du champ

interpolé). L'ensemble des points où l'hypothèse locale nulle est rejetée, *i.e.* pour lesquels $T(\mathbf{x})$ est supérieure au $(1-\alpha)$-quantile d'une $\chi^2(2)$, t_α, définit les ZCAs potentielles. Autrement dit, les ZCAs potentielles correspondent à l'ensemble d'excursion du champ $T(\cdot)$ au dessus du seuil t_α. La deuxième étape consiste à tester la significativité des ZCAs détectées. Pour cela, nous avons dû démontrer de nouveaux résultats sur le champ de χ^2 non stationnaire considéré. En particulier nous avons pu montrer que la loi des excès en un maximum est exponentielle et nous avons établi une distribution asymptotique pour la taille d'une composante connexe de l'ensemble d'excursion lorsque le seuil est élevé. La significativité des ZCAs potentielles détectées est testée en utilisant comme statistique de test la taille d'une ZCA potentielle. La valeur critique associée à ce test est comparée à un niveau global η. Une ZCA potentielle significative est appelée ZCA.

Le niveau local α correspondant à un niveau global η fixé, par exemple $\eta = 5\%$, a pu être déterminé. Nous avons retenu deux possibilités pour ce niveau. La première est obtenue par des simulations de type Monte-Carlo et dépend de la discrétisation de la grille d'interpolation, du schéma d'échantillonnage et de la fonction de covariance du champ aléatoire. La seconde est liée à la portée intégrale du champ, elle dépend du schéma d'échantillonnage et de la fonction de covariance, mais pas de la discrétisation.

La statistique de test, $T(\mathbf{x})$, étant proportionelle à la variance du gradient estimé, une sur-estimation de la variance entraîne de faibles valeurs de $T(\mathbf{x})$ et une sous-estimation des ZCAs. Comme nous nous plaçons sous l'hypothèse d'absence de ZCAs, il faut corriger le biais introduit dans le cas de l'existence d'une ZCA. Pour ce faire, nous avons proposé d'estimer itérativement la fonction de covariance et les ZCAs. A chaque itération les couples $\{Z(x_i), Z(x_j)\}$ tels que le segment formé par les points x_i et x_j intersecte une ZCA sont éliminés de la procédure d'estimation des paramètres de la fonction de covariance. Nous avons pu montrer que cette procédure itérative converge rapidement (souvent en moins de cinq itérations) et qu'elle corrige parfaitement le biais introduit dans l'estimation de la variance en présence d'une discontinuité.

Dans la troisième partie de ce travail, nous avons proposé un calcul de la puissance locale du test de détection de ZCAs, *i.e.* la probabilité de détecter en $\mathbf{x}$ une courbe de discontinuité passant par $\mathbf{x}$. Ce travail fait suite au problème de l'impact de la densité locale de l'échantillon sur la puissance de la méthode. La puissance conditionnelle locale a été calculée en surmontant le problème de la non indépendance des tests locaux et en conditionnant par rapport à la valeur de la statistique de test, T, aux pixels définissant une fenêtre centrée en $\mathbf{x}$. La cartographie de cette puissance permet de visualiser les zones où la densité locale d'échantillonnage n'est pas adaptée à la détection de ZCAs.

La méthode de détection de ZCAs a été validée par une étude sur simulations. Elle se révèle puissante, une discontinuité de 2.5 écarts-type est presque toujours détectée pour un échantillon de 100 points. Elle est robuste par rapport au choix de la discrétisation. Le taux de détection est indépendant de la discrétisation. Cependant, plus la discrétisation est fine, mieux les ZCAs sont délimitées car les petites ZCAs sont alors prises en compte. La méthode est peu sensible à une erreur dans l'estimation des paramètres de la fonction de covariance. En effet, le taux de détection varie peu lorsque la portée s'écarte de $\pm 30\%$ de la portée vraie. La forme paramétrique est d'importance secondaire si la portée est bien estimée.

L'application à des données d'agriculture de précision (variables permanentes : limon, calcaire, argile, entre autres, et variables dynamiques : teneur en eau et en azote minéral dans une parcelle agricole du nord de la France) s'est révélée intéressante. Les donnéees à différentes dates des variables dynamiques ont permis de mettre en évidence des structures permanentes superposables avec les structures détectées d'une part dans les variables permanentes et d'autre part avec la carte pédologique de la parcelle (obtenue de façon indépendante). La représentation cartographique de la puissance permet de visualiser les zones où l'échantillon est localement trop peu dense pour pouvoir détecter des ZCAs et par conséquent de prévoir une stratégie d'échantillonnage différente lors de prélèvements ultérieurs.

14.6 Perspectives

La méthode de détection de ZCAs a été développée sous l'hypothèse de gaussianité des données. Cette hypothèse est assez forte dans le sens que peu de phénomènes biologiques sont modélisés par une distribution gaussienne. Cependant si les données ne sont pas gaussiennes, elles peuvent être transformées soit en utilisant une transformation paramétrique (log transformation ou Cox transformation), soit en utilisant une transformation non paramétrique. Ces transformations n'assurent pas la gaussianité du champ aléatoire, mais pourraient être suffisantes pour que la méthode soit utilisable en pratique.

Les résultats de l'application aux données d'agriculture de précision soulèvent deux questions. En effet, les ZCAs détectées pour la teneur en eau de la parcelle apparaissent permanentes dans le temps, aussi la première question que nous sommes amenés à nous poser est : comment intégrer l'aspect temporel dans la méthode de détection de ZCAs ? De façon plus générale, l'application a montré que les structures détectées sont les mêmes

pour les différentes variables. La seconde question n'est pas indépendante de la première : comment étendre la méthode au multivariable ? Nous avons traité le cas où l'échantillon $\mathbf{Z}$ est unidimensionnel, est-ce que les résultats mathématiques resteraient valables pour $\mathbf{Z}$ de dimension $p > 1$? L'extension au multivariable engendre de nouveaux problèmes tels que l'estimation du modèle de covariance croisée, ou encore la prise en compte des différents schémas d'échantillonnage des variables.

En introduction de ce travail nous avons affirmé que la méthode de détection de ZCAs est complémentaire aux méthodes de classification, car les frontières des zones issues de la classification ne définissent pas toujours des ZCAs et inversément les ZCAs ne définissent pas nécessairement des groupes spatialement bien séparés. Il serait donc intéressant de pouvoir comparer, par simulations ou sur des données, les résultats des deux types de méthodes. Ce travail est en cours. En effet, nous avons appliqué des méthodes de classification aux données d'agriculture de précision. Ces méthodes reposent sur l'*algorithme EM* (Estimation Maximisation). L'algorithme EM est un outil pour le calcul de l'estimateur du maximum de vraisemblance des paramètres d'un modèle en présence de données manquantes (les classes). Différents modèles existent selon qu'il y ait indépendance ou corrélation spatiale entre les classes et indépendance ou corrélation entre les observations sachant les classes. Par exemple le modèle de champ de Markov caché correspond au cas où la classe d'une observation est corrélée avec la classe des autres observations et est modélisée par un champ de Markov, et où les observations sachant les classes sont indépendantes. Nous avons également considéré un autre modèle. Dans ce dernier les classes sont indépendantes et les observations sachant les classes sont corrélées, appelons ce modèle *modèle géostatistique*. Le modèle de champ de Markov caché a été étudié dans Peyrard (2001) dans un contexte de segmentation d'images et le modèle géostatistique dans Allemand (2004). L'algorithme EM a été adapté à ces modèles pour des données spatiales. Les résultats de classification et les résultats de la méthode de détection de ZCAs sont apparus complémentaires. En effet, en ce qui concerne la teneur en eau de la parcelle, la méthode de détection de ZCAs détecte pour l'ensemble des dates des ZCAs dans la partie centre-est de la parcelle et pour quelques dates seulement une ZCA permettant de scinder la parcelle en deux zones, une au nord, l'autre au sud. En revanche, avec les méthodes de classification, les zones au nord et au sud sont obtenues pour l'ensemble des dates, alors que la zone de la partie centre-est est obtenue pour seulement la moitié des dates. Ce résultat est probablement dû au fait qu'une classe n'est jamais constitutée d'un ou deux points, alors que la méthode détection est adaptée à ce genre de situation.

Bien qu'encourageants ces résultats sont à considérer avec prudence, car ces travaux sont récents et en cours de validation.

Nous avons eu l'occasion d'appliquer la méthode de détection de ZCAs à des données dans un contexte d'épidémiologie, où il ne s'agissait pas de détecter des changements abrupts entre différentes zones, mais de détecter des clusters dûs à l'émergence d'une maladie. La méthode s'est montrée efficace sur ce type de données car elle est parfaitement adaptée à la détection de valeurs aberrantes.
La méthode de détection de ZCAs devrait être appliquée (et adaptée) à d'autres types de données.

Sixième partie

Annexes

Annexe A

Décomposition du champ $T(\cdot)$

La décomposition du champ $T(\cdot)$ fait appel au théorème de Cholesky.

Théorème 12 (Factorisation de Cholesky d'une matrice)
Si $\mathbf{A}$ *est une matrice symétrique, définie-positive, il existe (au moins) une matrice triangulaire inférieure* $\mathbf{B}$ *telle que :*

$$\mathbf{A} = \mathbf{B}\mathbf{B}'.$$

De plus, on peut supposer que les éléments diagonaux de $\mathbf{B}$ *soient tous* > 0*, la factorisation est alors unique.*

Preuve de la proposition 3

Nous allons montrer, en utilisant le changement de variables $U(\mathbf{x}) = \mathbf{\Sigma}^{-1/2}(\mathbf{x})W(\mathbf{x})$, que le champ $T(\mathbf{x})$ se décompose en la somme du carré des champs $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ définis par :

$$U_1(\mathbf{x}) = \frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})} \text{ et } U_2(\mathbf{x}) = \frac{1}{\sqrt{1-\rho^2(\mathbf{x})}} \left\{ \frac{W_2(\mathbf{x})}{\sigma_2(\mathbf{x})} - \rho(\mathbf{x}) \frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})} \right\}.$$

$\mathbf{\Sigma}(\mathbf{x})$ est une matrice symétrique, définie positive :

$$\mathbf{\Sigma}(\mathbf{x}) = \begin{pmatrix} \sigma_1^2(\mathbf{x}) & \sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\rho(\mathbf{x}) \\ \sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\rho(\mathbf{x}) & \sigma_2^2(\mathbf{x}) \end{pmatrix}. \tag{A.1}$$

D'après le théorème 12 elle se décompose en le produit d'une matrice triangulaire inférieure et de sa transposée : $\mathbf{\Sigma}(\mathbf{x}) = \mathbf{B}(\mathbf{x})\mathbf{B}'(\mathbf{x})$.

$$\mathbf{\Sigma}(\mathbf{x}) = \begin{pmatrix} b_1(\mathbf{x}) & 0 \\ b_2(\mathbf{x}) & b_3(\mathbf{x}) \end{pmatrix} \begin{pmatrix} b_1(\mathbf{x}) & b_2(\mathbf{x}) \\ 0 & b_3(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} b_1^2(\mathbf{x}) & b_1(\mathbf{x})b_2(\mathbf{x}) \\ b_1(\mathbf{x})b_2(\mathbf{x}) & b_2^2(\mathbf{x}) + b_3^2(\mathbf{x}) \end{pmatrix}. \tag{A.2}$$

La résolution du système (A.1) = (A.2), fournit les résultats suivants :
$b_1(\mathbf{x}) = \pm\sigma_1(\mathbf{x})$,
$b_2(\mathbf{x}) = \sigma_2(\mathbf{x})\rho(\mathbf{x})$ si $b_1(\mathbf{x})$ est positif, $b_2(\mathbf{x}) = -\sigma_2(\mathbf{x})\rho(\mathbf{x})$ si $b_1(\mathbf{x})$ est négatif,
$b_3(\mathbf{x}) = \pm\sigma_2(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})}$.

Ainsi, il y a quatre possibilités pour $\mathbf{\Sigma}^{1/2}(\mathbf{x})$:
$$\mathbf{B}_1 = \begin{pmatrix} \sigma_1(\mathbf{x}) & 0 \\ \sigma_2(\mathbf{x})\rho(\mathbf{x}) & \sigma_2(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})} \end{pmatrix}, \mathbf{B}_2 = \begin{pmatrix} \sigma_1(\mathbf{x}) & 0 \\ \sigma_2(\mathbf{x})\rho(\mathbf{x}) & -\sigma_2(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})} \end{pmatrix},$$
$$\mathbf{B}_3 = \begin{pmatrix} -\sigma_1(\mathbf{x}) & 0 \\ -\sigma_2(\mathbf{x})\rho(\mathbf{x}) & \sigma_2(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})} \end{pmatrix} \text{ et } \mathbf{B}_4 = \begin{pmatrix} -\sigma_1(\mathbf{x}) & 0 \\ -\sigma_2(\mathbf{x})\rho(\mathbf{x}) & -\sigma_2(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})} \end{pmatrix}.$$

Nous choisissons de prendre $\mathbf{\Sigma}^{1/2}(\mathbf{x}) = \mathbf{B}_1$. Alors,

$$\begin{aligned} \mathbf{\Sigma}^{-1/2}(\mathbf{x}) &= \mathbf{B}_1^{-1} \\ &= \frac{1}{\sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})}} \begin{pmatrix} \sigma_2(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})} & 0 \\ -\sigma_2(\mathbf{x})\rho(\mathbf{x}) & \sigma_1(\mathbf{x}) \end{pmatrix} \\ &= \begin{pmatrix} 1/\sigma_1(\mathbf{x}) & 0 \\ -\rho(\mathbf{x})/\sigma_1(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})} & 1/\sigma_2(\mathbf{x})\sqrt{1-\rho^2(\mathbf{x})} \end{pmatrix}. \end{aligned}$$

Nous en déduisons facilement $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ à partir de $U(\mathbf{x}) = \mathbf{\Sigma}^{-1/2}(\mathbf{x})W(\mathbf{x})$. □

Si nous avions considéré une autre solution pour la matrice $\mathbf{\Sigma}^{-1/2}(\mathbf{x})$, les champs $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ seraient les mêmes au signe près, *i.e.* les champs $U(\mathbf{x})$ obtenus avec les autres solutions sont : $(U_1(\mathbf{x}), -U_2(\mathbf{x}))'$, $(-U_1(\mathbf{x}), U_2(\mathbf{x}))'$ et $(-U_1(\mathbf{x}), -U_2(\mathbf{x}))'$. Toutes les propriétés des champs $U_i(\mathbf{x})$, $i = 1, 2$ restent valablent pour ces autres possibilités (à quelques équations au signe près) et surtout le champ $T(\cdot)$ reste le même puisqu'il s'agit de la somme du carré des champs $U_i(\mathbf{x})$, $i = 1, 2$.

Annexe B

Propriétés du champ $U(\cdot)$

Preuve de la proposition 4

(i) Nous allons montrer que les champs $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ sont indépendants.

Comme $U(\mathbf{x}) = \mathbf{\Sigma}^{-1/2}(\mathbf{x})W(\mathbf{x}) = (U_1(\mathbf{x}), U_2(\mathbf{x}))'$, nous avons (quelle que soit la racine carrée de la matrice $\mathbf{\Sigma}^{-1}(\mathbf{x})$) :

$$\begin{aligned}
\text{Var}(\mathbf{x}) &= \mathbb{E}\left[\mathbf{\Sigma}^{-1/2}(\mathbf{x})W(\mathbf{x})W'(\mathbf{x})\mathbf{\Sigma}^{-1/2}(\mathbf{x})\right] - \mathbb{E}\left[\mathbf{\Sigma}^{-1/2}(\mathbf{x})W(\mathbf{x})\right]\mathbb{E}'\left[\mathbf{\Sigma}^{-1/2}(\mathbf{x})W(\mathbf{x})\right] \\
&= \mathbf{\Sigma}^{-1/2}(\mathbf{x})\mathbf{\Sigma}(\mathbf{x})\mathbf{\Sigma}^{-1/2}(\mathbf{x}) \\
&= \mathbf{Id}_2,
\end{aligned}$$

où $\mathbf{Id}_2$ est la 2×2-matrice identité. Puisque $U(\mathbf{x})$ est un vecteur gaussien, il en ressort que $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ sont indépendants.

(ii) Nous utilisons les relations (6.1) et (6.2) pour donner l'expression analytique des corrélations entre $U_i(\mathbf{x})$ et $U_j(\mathbf{y})$, $i, j = 1, 2$.

$$\begin{aligned}
\mathbb{E}\left[U_1(\mathbf{x})U_1(\mathbf{y})\right] &= \mathbb{E}\left[\frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})}\frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})}\right] = \mathbb{E}\left[\frac{\partial_1 C'(\mathbf{x})\mathbf{C}^{-1}\mathbf{Z}}{\sigma_1(\mathbf{x})}\frac{\mathbf{Z}'\mathbf{C}^{-1}\partial_1 C(\mathbf{y})}{\sigma_1(\mathbf{y})}\right] \\
&= \frac{\partial_1 C'(\mathbf{x})\mathbf{C}^{-1}}{\sigma_1(\mathbf{x})}\mathbb{E}\left[\mathbf{Z}\mathbf{Z}'\right]\frac{\mathbf{C}^{-1}\partial_1 C(\mathbf{y})}{\sigma_1(\mathbf{y})} = \frac{\partial_1 C'(\mathbf{x})\mathbf{C}^{-1}\partial_1 C(\mathbf{y})}{\sigma_1(\mathbf{x})\sigma_1(\mathbf{y})} \\
&= \rho_{11}(\mathbf{x}, \mathbf{y}).
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[U_1(\mathbf{x})U_2(\mathbf{y})\right] &= \mathbb{E}\left[\frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})}\frac{1}{\sqrt{1-\rho^2(\mathbf{y})}}\left\{\frac{W_2(\mathbf{y})}{\sigma_2(\mathbf{y})}-\rho(\mathbf{y})\frac{W_1(\mathbf{y})}{\sigma_1(\mathbf{y})}\right\}\right] \\
&= \frac{1}{\sqrt{1-\rho^2(\mathbf{y})}}\left\{\frac{\partial_1 C'(\mathbf{x})\mathbf{C}^{-1}\partial_2 C(\mathbf{y})}{\sigma_1(\mathbf{x})\sigma_2(\mathbf{y})}-\rho(\mathbf{y})\frac{\partial_1 C'(\mathbf{x})\mathbf{C}^{-1}\partial_1 C(\mathbf{y})}{\sigma_1(\mathbf{x})\sigma_1(\mathbf{y})}\right\} \\
&= \frac{1}{\sqrt{1-\rho^2(\mathbf{y})}}\left\{\rho_{12}(\mathbf{x},\mathbf{y})-\rho(\mathbf{y})\rho_{11}(\mathbf{x},\mathbf{y})\right\}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[U_2(\mathbf{x})U_2(\mathbf{y})\right] &= \mathbb{E}\left[\frac{1}{\sqrt{1-\rho^2(\mathbf{x})}}\left\{\frac{W_2(\mathbf{x})}{\sigma_2(\mathbf{x})}-\rho(\mathbf{x})\frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})}\right\}\right. \\
&\quad \left.\frac{1}{\sqrt{1-\rho^2(\mathbf{y})}}\left\{\frac{W_2(\mathbf{y})}{\sigma_2(\mathbf{y})}-\rho(\mathbf{y})\frac{W_1(\mathbf{y})}{\sigma_1(\mathbf{y})}\right\}\right] \\
&= \frac{1}{\sqrt{1-\rho^2(\mathbf{x})}\sqrt{1-\rho^2(\mathbf{y})}}\mathbb{E}\left[\frac{W_2(\mathbf{x})}{\sigma_2(\mathbf{x})}\frac{W_2'(\mathbf{y})}{\sigma_2(\mathbf{y})}-\frac{W_2(\mathbf{x})}{\sigma_2(\mathbf{x})}\rho(\mathbf{y})\frac{W_1'(\mathbf{y})}{\sigma_1(\mathbf{y})}\right. \\
&\quad \left.-\rho(\mathbf{x})\frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})}\frac{W_2'(\mathbf{y})}{\sigma_2(\mathbf{y})}+\rho(\mathbf{x})\frac{W_1(\mathbf{x})}{\sigma_1(\mathbf{x})}\rho(\mathbf{y})\frac{W_1'(\mathbf{y})}{\sigma_1(\mathbf{y})}\right] \\
&= \frac{1}{\sqrt{1-\rho^2(\mathbf{x})}\sqrt{1-\rho^2(\mathbf{y})}}\left\{\frac{\partial_2 C'(\mathbf{x})\mathbf{C}^{-1}\partial_2 C(\mathbf{y})}{\sigma_2(\mathbf{x})\sigma_2(\mathbf{y})}\right. \\
&\quad -\rho(\mathbf{y})\frac{\partial_2 C'(\mathbf{x})\mathbf{C}^{-1}\partial_1 C(\mathbf{y})}{\sigma_2(\mathbf{x})\sigma_1(\mathbf{y})}-\rho(\mathbf{x})\frac{\partial_1 C'(\mathbf{x})\mathbf{C}^{-1}\partial_2 C(\mathbf{y})}{\sigma_1(\mathbf{x})\sigma_2(\mathbf{y})} \\
&\quad \left.+\rho(\mathbf{x})\rho(\mathbf{y})\frac{\partial_1 C'(\mathbf{x})\mathbf{C}^{-1}\partial_1 C(\mathbf{y})}{\sigma_1(\mathbf{x})\sigma_1(\mathbf{y})}\right\} \\
&= \frac{1}{\sqrt{1-\rho^2(\mathbf{x})}\sqrt{1-\rho^2(\mathbf{y})}}\left\{\rho_{22}(\mathbf{x},\mathbf{y})-\rho(\mathbf{x})\rho_{12}(\mathbf{x},\mathbf{y})-\rho(\mathbf{y})\rho_{21}(\mathbf{x},\mathbf{y})\right. \\
&\quad \left.+\rho(\mathbf{x})\rho(\mathbf{y})\rho_{11}(\mathbf{x},\mathbf{y})\right\}
\end{aligned}$$

□

Preuve du lemme 4

Dans ce qui suit, $i, j \in \{1, 2\}$ avec $i \neq j$. Pour une question d'allégement des équations à venir, pour $\mathbf{x} = (x^1, x^2)'$, nous désignerons :

$$\frac{\partial}{\partial x^k}U_i(\mathbf{x}) = \frac{\partial}{\partial x^k}U_i(x^1, x^2) = \begin{cases} \lim_{h\to 0}\left(U_i(x^1+h, x^2) - U_i(x^1, x^2)\right)/h, & \text{si } k = 1 \\ \lim_{h\to 0}\left(U_i(x^1, x^2+h) - U_i(x^1, x^2)\right)/h, & \text{si } k = 2 \end{cases}$$

par $\frac{\partial}{\partial x^k}U_i(\mathbf{x}) = \lim_{h\to 0}\left(U_i(\mathbf{x}+h) - U_i(\mathbf{x})\right)/h$.

(i) Nous avons $\frac{\partial}{\partial x^k}\mathbb{E}\left[U_i(\mathbf{x}) \mid U_j(\mathbf{x})\right] = \frac{\partial}{\partial x^k}\int_{\mathbb{R}} u_i f_{U_i(\mathbf{x})|U_j(\mathbf{x})}(u_i)\,\mathrm{d}u_i$, où $f_{U_i(\mathbf{x})|U_j(\mathbf{x})}(u_i)$ est la densité marginale de $U_i(\mathbf{x}) \mid U_j(\mathbf{x})$. Comme $U_i(\mathbf{x})$ et $U_j(\mathbf{x})$ sont indépendants, nous avons $f_{U_i(\mathbf{x})|U_j(\mathbf{x})}(u_i) = f_{U_i(\mathbf{x})}(u_i)$. Alors,

$$\begin{aligned}\frac{\partial}{\partial x^k}\mathbb{E}\left[U_i(\mathbf{x}) \mid U_j(\mathbf{x})\right] &= \frac{\partial}{\partial x^k}\int_{\mathbb{R}} u_i f_{U_i(\mathbf{x})}(u_i)\,\mathrm{d}u_i \\ &= \int_{\mathbb{R}} u_i \frac{\partial}{\partial x^k} f_{U_i(\mathbf{x})}(u_i)\,\mathrm{d}u_i, \text{ d'après le théorème de Lebesgue.}\end{aligned}$$

L'indépendance entre $U_i(\mathbf{x})$ et $U_j(\mathbf{x})$ permet également d'écrire :

$$f_{U_i(\mathbf{x}),U_j(\mathbf{x})}(u_i,u_j) = f_{U_i(\mathbf{x})}(u_i) f_{U_j(\mathbf{x})}(u_j),$$

et nous avons :

$$\frac{\partial}{\partial x^k} f_{U_i(\mathbf{x}),U_j(\mathbf{x})}(u_i,u_j) = \frac{\partial}{\partial x^k} f_{U_i(\mathbf{x})}(u_i) f_{U_j(\mathbf{x})}(u_j).$$

Or, $\frac{\partial}{\partial x^k} f_{U_i(\mathbf{x}),U_j(\mathbf{x})}(u_i,u_j) = \frac{\partial}{\partial y} f_{U_i(\mathbf{y}),U_j(\mathbf{x})}(u_i,u_j)\big|_{y=x^k} + \frac{\partial}{\partial y} f_{U_i(\mathbf{x}),U_j(\mathbf{y})}(u_i,u_j)\big|_{y=x^k}$
et $\frac{\partial}{\partial x^k} f_{U_i(\mathbf{x})}(u_i) f_{U_j(\mathbf{x})}(u_j) = \frac{\partial}{\partial y} f_{U_i(\mathbf{y})}(u_i)\big|_{y=x^k} f_{U_j(\mathbf{x})}(u_j) + f_{U_i(\mathbf{x})}(u_i)\frac{\partial}{\partial y} f_{U_j(\mathbf{x})}(u_j)\big|_{y=x^k}$
impliquent $\frac{\partial}{\partial y} f_{U_i(\mathbf{y}),U_j(\mathbf{x})}(u_i,u_j)\big|_{y=x^k} = \frac{\partial}{\partial y} f_{U_i(\mathbf{y})}(u_i)\big|_{y=x^k} f_{U_j(\mathbf{x})}(u_j)$.
Alors,

$$\begin{aligned}\mathbb{E}\left[\frac{\partial U_i(\mathbf{x})}{\partial x^k}\Big| U_j(\mathbf{x})\right] &= \mathbb{E}\left[\lim_{h\to 0}\frac{U_i(\mathbf{x}+h)-U_i(\mathbf{x})}{h}\Big| U_j(\mathbf{x})\right] \\ &= \lim_{h\to 0}\frac{1}{h}\left(\mathbb{E}\left[U_i(\mathbf{x}+h) \mid U_j(\mathbf{x})\right] - \mathbb{E}\left[U_i(\mathbf{x}) \mid U_j(\mathbf{x})\right]\right) \\ &= \lim_{h\to 0}\frac{1}{h}\left(\int_{\mathbb{R}} u_i f_{U_i(\mathbf{x}+h)|U_j(\mathbf{x})}(u_i)\,\mathrm{d}u_i - \int_{\mathbb{R}} u_i f_{U_i(\mathbf{x})|U_j(\mathbf{x})}(u_i)\right)\mathrm{d}u_i \\ &= \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\frac{u_i}{f_{U_j(\mathbf{x})}(u_j)}\left(f_{U_i(\mathbf{x}+h),U_j(\mathbf{x})}(u_i,u_j) - f_{U_i(\mathbf{x}),U_j(\mathbf{x})}(u_i,u_j)\right)\mathrm{d}u_i \\ &= \int_{\mathbb{R}}\frac{u_i}{f_{U_j(\mathbf{x})}(u_j)}\lim_{h\to 0}\frac{1}{h}\left(f_{U_i(\mathbf{x}+h),U_j(\mathbf{x})}(u_i,u_j) - f_{U_i(\mathbf{x}),U_j(\mathbf{x})}(u_i,u_j)\right)\mathrm{d}u_i \\ &= \int_{\mathbb{R}}\frac{u_i}{f_{U_j(\mathbf{x})}(u_j)}\frac{\partial}{\partial y} f_{U_i(\mathbf{y}),U_j(\mathbf{x})}(u_i,u_j)\big|_{y=x^k}\,\mathrm{d}u_i \\ &= \int_{\mathbb{R}}\frac{u_i}{f_{U_j(\mathbf{x})}(u_j)}\frac{\partial}{\partial y} f_{U_i(\mathbf{y})}(u_i)\big|_{y=x^k} f_{U_j(\mathbf{x})}(u_j) \\ &= \int_{\mathbb{R}} u_i \frac{\partial}{\partial y} f_{U_i(\mathbf{y})}(u_i)\big|_{y=x^k}\,\mathrm{d}u_i \\ &= \frac{\partial}{\partial x^k}\mathbb{E}\left[U_i(\mathbf{x}) \mid U_j(\mathbf{x})\right].\end{aligned}$$

(ii) Cette propriété s'obtient en remplaçant $U_j(\mathbf{x})$ par $\frac{\partial U_j(\mathbf{x})}{\partial x^l}$ dans la démonstration de (i). □

Preuve de la proposition 5

Les dérivées k-ième de $U_1(\mathbf{x})$, $\partial^{(k)}U_1(\mathbf{x})$, et l-ième de $U_2(\mathbf{x})$, $\partial^{(l)}U_2(\mathbf{x})$, sont indépendantes. En effet, nous allons dans un premier temps montrer que $\mathbb{E}\left[\partial^{(k)}U_1(\mathbf{x})U_2(\mathbf{x})\right] = \mathbf{0}$, puis que $\mathbb{E}\left[\partial^{(k)}U_1(\mathbf{x})\partial^{(l)}U_2'(\mathbf{x})\right] = \mathbf{0}$ à k fixé.

Nous avons d'une part,

$$\mathbb{E}\left[U_1(\mathbf{x}) \mid U_2(\mathbf{x})\right] = 0 \tag{B.1}$$

car par hypothèse $U_1(\mathbf{x})$ et $U_2(\mathbf{x})$ sont indépendants et d'espérance nulle. D'autre part, d'après la proposition 1, $\partial U_1(\mathbf{x})$ est un champ gaussien, aussi l'espérance conditionnelle de $\partial U_1(\mathbf{x})$ sachant $U_2(\mathbf{x})$ équivaut à la régression linéaire :

$$\begin{aligned}\mathbb{E}\left[\partial U_1(\mathbf{x}) \mid U_2(\mathbf{x})\right] &= \mathbb{E}\left[\partial U_1(\mathbf{x})\right] + \mathrm{Cov}\left(\partial U_1(\mathbf{x}), U_2(\mathbf{x})\right)\mathrm{Var}^{-1}\left(U_2(\mathbf{x})\right)U_2(\mathbf{x}) &\text{(B.2)}\\ &= \mathbb{E}\left[\partial U_1(\mathbf{x})U_2'(\mathbf{x})\right]U_2(\mathbf{x}). &\text{(B.3)}\end{aligned}$$

De plus, d'après la relation (i) du lemme 4 :

$$\partial\mathbb{E}\left[U_1(\mathbf{x}) \mid U_2(\mathbf{x})\right] = \mathbb{E}\left[\partial U_1(\mathbf{x}) \mid U_2(\mathbf{x})\right]. \tag{B.4}$$

D'après les équations (B.1), (B.3) et (B.4), il s'en suit que $\mathbb{E}\left[\partial U_1(\mathbf{x})U_2(\mathbf{x})\right] = \mathbf{0}$ car le champ $U_2(\cdot)$ n'est pas identiquement nul.

Montrons par récurrence sur k que $\mathbb{E}\left[\partial^{(k)}U_1(\mathbf{x})U_2(\mathbf{x})\right] = \mathbf{0}$. Nous venons de voir que l'égalité est vraie au rang $k = 1$ (et elle est vraie au rang $k = 0$ par hypothèse). Supposons que l'égalité soit vraie au rang k et montrons que c'est encore le cas au rang $k+1$. Comme,

$$\begin{aligned}\mathbb{E}\left[\partial^{(k)}U_1(\mathbf{x}) \mid U_2(\mathbf{x})\right] &= \mathbf{0} \text{ car } \partial^{(k)}U_1(\mathbf{x}) \text{ et } U_2(\mathbf{x}) \text{ sont indépendants et centrés,}\\ \partial\mathbb{E}\left[\partial^{(k)}U_1(\mathbf{x}) \mid U_2(\mathbf{x})\right] &= \mathbb{E}\left[\partial^{(k+1)}U_1(\mathbf{x}) \mid U_2(\mathbf{x})\right], \text{ d'après le lemme 4 } (i)\\ \mathbb{E}\left[\partial^{(k+1)}U_1(\mathbf{x}) \mid U_2(\mathbf{x})\right] &= \mathbb{E}\left[\partial^{(k+1)}U_1(\mathbf{x})U_2(\mathbf{x})\right]U_2(\mathbf{x})\end{aligned}$$

nous en déduisons que $\mathbb{E}\left[\partial^{(k+1)}U_1(\mathbf{x})U_2(\mathbf{x})\right]U_2(\mathbf{x}) = \mathbf{0}$.

Montrons par récurrence sur l que $\mathbb{E}\left[\partial^{(k)}U_1(\mathbf{x})\partial^{(l)}U_2'(\mathbf{x})\right] = \mathbf{0}$ à k fixé. Comme,

$$\begin{aligned}
\mathbb{E}\left[U_2(\mathbf{x}) \mid \partial^{(k)}U_1(\mathbf{x})\right] &= 0 \text{ car } \partial^{(k)}U_1(\mathbf{x}) \text{ et } U_2(\mathbf{x}) \text{ sont indépendants et centrés,} \\
\partial\mathbb{E}\left[U_2(\mathbf{x}) \mid \partial^{(k)}U_1(\mathbf{x})\right] &= \mathbb{E}\left[\partial U_2(\mathbf{x}) \mid \partial^{(k)}U_1(\mathbf{x})\right], \text{ d'après le lemme 4 } (ii) \\
\mathbb{E}\left[\partial U_2(\mathbf{x}) \mid \partial^{(k)}U_1(\mathbf{x})\right] &= \mathbb{E}\left[\partial U_2(\mathbf{x})\partial^{(k)}U_1'(\mathbf{x})\right] \mathrm{Var}^{-1}\left(\partial^{(k)}U_1(\mathbf{x})\right)\partial^{(k)}U_1(\mathbf{x})
\end{aligned}$$

nous en déduisons que $\mathbb{E}\left[\partial U_2(\mathbf{x})\partial^{(k)}U_1'(\mathbf{x})\right] = \mathbf{0}$. L'égalité est vraie au rang $l = 1$, supposons qu'elle soit vraie au rang l et montrons qu'elle est vérifiée au rang $l+1$. Comme,

$$\begin{aligned}
\mathbb{E}\left[\partial^{(l)}U_2(\mathbf{x}) \mid \partial^{(k)}U_1(\mathbf{x})\right] &= \mathbf{0} \text{ car } \partial^{(k)}U_1(\mathbf{x}) \text{ et } \partial^{(l)}U_2(\mathbf{x}) \text{ sont indépendants et centrés,} \\
\partial\mathbb{E}\left[\partial^{(l)}U_2(\mathbf{x}) \mid \partial^{(k)}U_1(\mathbf{x})\right] &= \mathbb{E}\left[\partial^{(l+1)}U_2(\mathbf{x}) \mid \partial^{(k)}U_1(\mathbf{x})\right], \text{ d'après le lemme 4 } (ii) \\
\mathbb{E}\left[\partial^{(l+1)}U_2(\mathbf{x}) \mid \partial^{(k)}U_1(\mathbf{x})\right] &= \mathbb{E}\left[\partial^{(l+1)}U_2(\mathbf{x})\partial^{(k)}U_1'(\mathbf{x})\right] \mathrm{Var}^{-1}\left(\partial^{(k)}U_1(\mathbf{x})\right)\partial^{(k)}U_1(\mathbf{x})
\end{aligned}$$

nous en déduisons que $\mathbb{E}\left[\partial^{(l+1)}U_2(\mathbf{x})\partial^{(k)}U_1'(\mathbf{x})\right] = \mathbf{0}$. □

Preuve du résultat 1

Pour $k, l = 1, 2$, l'élément $[k, l]$ de la matrice de covariance de $\dot{U}_2$ est, en enlevant la dépendance en $\mathbf{x}$ pour alléger les notations :

$$\mathbf{\Lambda}_{2\ [kl]} = \mathbb{E}\left[\partial_k U_2 \partial_l U_2\right],$$

avec

$$\partial_k U_2 = \frac{\rho\partial_k\rho}{1-\rho^2}U_2 + \frac{1}{\sqrt{1-\rho^2}}\left\{-\frac{\partial_k\sigma_2}{\sigma_2}\frac{W_2}{\sigma_2} + \frac{\partial_k W_2}{\sigma_2} - \partial_k\rho\frac{W_1}{\sigma_1} + \rho\frac{\partial_k\sigma_1}{\sigma_1}\frac{W_1}{\sigma_1} - \rho\frac{\partial_k W_1}{\sigma_1}\right\}$$

Dans la suite nous noterons $D_1 = D_1(\mathbf{x}) = \partial_1 C(\mathbf{x})$ et $D_2 = D_2(\mathbf{x}) = \partial_2 C(\mathbf{x})$.

$$\begin{aligned}
\mathbf{\Lambda}_{2\ [kl]} &= \mathbb{E}\left[\partial_k U_2 \partial_l U_2\right] \\
&= \frac{\rho^2\partial_k\rho\partial_l\rho}{(1-\rho^2)^2} \\
&\quad + \frac{\rho\partial_k\rho}{(1-\rho^2)^2}\left\{-\frac{\partial_l\sigma_2}{\sigma_2} + \frac{\partial_l\sigma_2}{\sigma_2} - \rho\partial_l\rho + \rho^2\frac{\partial_l\sigma_1}{\sigma_1} - \rho\frac{D_2'\mathbf{C}^{-1}\partial_l D_1}{\sigma_1\sigma_2}\right. \\
&\qquad \left. + \rho^2\frac{\partial_l\sigma_2}{\sigma_2} - \rho\frac{\partial_l D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2} + \rho\partial_l\rho - \rho^2\frac{\partial_l\sigma_1}{\sigma_1} + \rho^2\frac{\partial_l\sigma_1}{\sigma_1}\right\} \\
&\quad + \frac{\rho\partial_l\rho}{(1-\rho^2)^2}\left\{-\frac{\partial_k\sigma_2}{\sigma_2} + \frac{\partial_k\sigma_2}{\sigma_2} - \rho\partial_k\rho + \rho^2\frac{\partial_k\sigma_1}{\sigma_1} - \rho\frac{D_2'\mathbf{C}^{-1}\partial_k D_1}{\sigma_1\sigma_2}\right. \\
&\qquad \left. + \rho^2\frac{\partial_k\sigma_2}{\sigma_2} - \rho\frac{\partial_k D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2} + \rho\partial_k\rho - \rho^2\frac{\partial_k\sigma_1}{\sigma_1} + \rho^2\frac{\partial_k\sigma_1}{\sigma_1}\right\}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{1-\rho^2}\left\{\frac{\partial_k\sigma_2}{\sigma_2}\frac{\partial_l\sigma_2}{\sigma_2}-\frac{\partial_k\sigma_2}{\sigma_2}\frac{\partial_l\sigma_2}{\sigma_2}+\rho\partial_l\rho\frac{\partial_k\sigma_2}{\sigma_2}-\rho^2\frac{\partial_k\sigma_2}{\sigma_2}\frac{\partial_l\sigma_1}{\sigma_1}+\rho\frac{\partial_k\sigma_2}{\sigma_2}\frac{D_2'\mathbf{C}^{-1}\partial_l D_1}{\sigma_1\sigma_2}\right.\\
&-\frac{\partial_k\sigma_2}{\sigma_2}\frac{\partial_l\sigma_2}{\sigma_2}+\frac{\partial_k D_2'\mathbf{C}^{-1}\partial_l D_2}{\sigma_2^2}-\partial_l\rho\frac{\partial_k D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2}+\rho\frac{\partial_l\sigma_1}{\sigma_1}\frac{\partial_k D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2}-\rho\frac{\partial_k D_2'\mathbf{C}^{-1}\partial_l D_1}{\sigma_1\sigma_2}\\
&+\rho\partial_k\rho\frac{\partial_l\sigma_2}{\sigma_2}-\partial_k\rho\frac{\partial_k D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2}+\partial_k\rho\partial_l\rho-\rho\partial_k\rho\frac{\partial_l\sigma_1}{\sigma_1}+\rho\partial_k\rho\frac{\partial_l\sigma_1}{\sigma_1}\\
&-\rho^2\frac{\partial_k\sigma_1}{\sigma_1}\frac{\partial_l\sigma_2}{\sigma_2}+\rho\frac{\partial_k\sigma_1}{\sigma_1}\frac{\partial_l D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2}-\rho\partial_l\rho\frac{\partial_k\sigma_1}{\sigma_1}+\rho^2\frac{\partial_k\sigma_1}{\sigma_1}\frac{\partial_l\sigma_1}{\sigma_1}-\rho^2\frac{\partial_k\sigma_1}{\sigma_1}\frac{\partial_l\sigma_1}{\sigma_1}\\
&\left.+\rho\frac{\partial_l\sigma_2}{\sigma_2}\frac{D_2'\mathbf{C}^{-1}\partial_k D_1}{\sigma_1\sigma_2}-\rho\frac{\partial_l D_2'\mathbf{C}^{-1}\partial_k D_1}{\sigma_1\sigma_2}+\rho\partial_l\rho\frac{\partial_k\sigma_1}{\sigma_1}-\rho^2\frac{\partial_k\sigma_1}{\sigma_1}\frac{\partial_l\sigma_1}{\sigma_1}+\rho^2\frac{\partial_k D_1'\mathbf{C}^{-1}\partial_l D_1}{\sigma_1^2}\right\}\\
=\;&-\frac{\rho^2\partial_k\rho\partial_l\rho}{\left(1-\rho^2\right)^2}+\frac{1}{1-\rho^2}\left\{\partial_k\rho\partial_l\rho-\frac{\partial_k\sigma_2}{\sigma_2}\frac{\partial_l\sigma_2}{\sigma_2}+\frac{\partial_k D_2'\mathbf{C}^{-1}\partial_l D_2}{\sigma_2^2}\right.\\
&+\rho\left[\frac{\partial_k\sigma_2}{\sigma_2}\left(\partial_l\rho+\frac{D_2'\mathbf{C}^{-1}\partial_l D_1}{\sigma_1\sigma_2}\right)+\frac{\partial_l\sigma_2}{\sigma_2}\left(\partial_k\rho+\frac{D_2'\mathbf{C}^{-1}\partial_k D_1}{\sigma_1\sigma_2}\right)\frac{\partial_l\sigma_1}{\sigma_1}\frac{\partial_k D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2}\right.\\
&\left.-\frac{\partial_k D_2'\mathbf{C}^{-1}\partial_l D_1}{\sigma_1\sigma_2}+\frac{\partial_k\sigma_1}{\sigma_1}\frac{\partial_l D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2}-\frac{\partial_l D_2'\mathbf{C}^{-1}\partial_k D_1}{\sigma_1\sigma_2}\right]\\
&+\rho^2\left[\frac{\partial_k D_1'\mathbf{C}^{-1}\partial_l D_1}{\sigma_1^2}-\frac{\partial_k\sigma_1}{\sigma_1}\frac{\partial_l\sigma_1}{\sigma_1}+\frac{\partial_k\sigma_1}{\sigma_1}\frac{\partial_l\sigma_2}{\sigma_2}-\frac{\partial_k\sigma_2}{\sigma_2}\frac{\partial_l\sigma_1}{\sigma_1}\right]-\partial_l\rho\frac{\partial_k D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2}\\
&\left.-\partial_k\rho\frac{\partial_l D_2'\mathbf{C}^{-1}D_1}{\sigma_1\sigma_2}\right\}
\end{aligned}
$$

□

Preuve du lemme 5

La démonstration du lemme 5 utilise le résultat classique suivant [Miller (1964)] :

Lemme 7 *Soit* $\mathbf{Z}=(Z_1,\ldots,Z_4)'$ *un vecteur gaussien de moyenne* $\mu=(\mu_1,\ldots,\mu_4)'$ *et de covariance* $\Gamma_{ij},\ 1\leq i,j,\leq 4$. *Alors,*

$$\mathbb{E}\left[Z_1Z_2Z_3\right]=\mu_1\mu_2\mu_3+\sum_i\mu_i\Gamma_{jk}\ \textit{et}\ \mathbb{E}\left[Z_1Z_2Z_3Z_4\right]=\mu_1\mu_2\mu_3\mu_4+\sum_{j\neq i}\mu_i\mu_j\Gamma_{kl}+\sum_{j>i}\Gamma_{ij}\Gamma_{kl}$$

où $j\neq i$, $k\neq i,j$ *et* $l\neq i,j,k$.

Preuve du lemme 5

(i) Il s'agit d'une conséquence directe du fait que la variance de $U_i(\mathbf{x})$ est constante. En effet, puisque $\mathbb{E}\left[U_i(\mathbf{x})U_i(\mathbf{x})\right]=1$, nous avons que $\partial_k\mathbb{E}\left[U_i(\mathbf{x})U_i(\mathbf{x})\right]=0$, *i.e.* :

$$2\mathbb{E}\left[U_i(\mathbf{x})\partial_k U_i(\mathbf{x})\right]=0.$$

S'agissant de champs gaussiens, l'absence de corrélation est équivalente à l'indépendance.

Afin d'alléger les notations, nous écrirons U_i au lieu de $U_i(\mathbf{x})$.
(ii) En utilisant le lemme 7 et (i),

$$\begin{aligned}\mathbb{E}\left[\partial_k U_i \partial_l U_i \partial^2_{k'l'} U_i \mid U\right] &= \mathbb{E}\left[\partial_k U_i \mid U\right] \mathbb{E}\left[\partial_l U_i \partial^2_{k'l'} U_i \mid U\right] + \mathbb{E}\left[\partial_l U_i \mid U\right] \mathbb{E}\left[\partial_k U_i \partial^2_{k'l'} U_i \mid U\right] \\ &\quad + \mathbb{E}\left[\partial^2_{k'l'} U_i \mid U\right] \mathbb{E}\left[\partial_k U_i \partial_l U_i \mid U\right] \\ &= \mathbb{E}\left[\partial^2_{k'l'} U_i \mid U\right] \mathbb{E}\left[\partial_k U_i \partial_l U_i \mid U\right], \text{ puisque } \mathbb{E}\left[\partial_k U_i \mid U\right] = 0\end{aligned}$$

Par conséquent, $\mathrm{Cov}\left(\partial^2_{k'l'} U_i, \partial_k U_i . \partial_l U_i \mid U\right) = 0$. Notons que ce résultat peut aussi être démontré inconditionnellement à U.

(iii) D'après la proposition 5, $\ddot{\mathbf{U}}_i$ est indépendant de U_j pour $i \neq j$ et $i, j = 1, 2$. Aussi conditionner $\ddot{\mathbf{U}}_i$ par U revient à conditionner par U_i.
Montrons d'abord que :

$$\mathbb{E}\left[\ddot{\mathbf{U}}_i \mid U_i\right] = -U_i \mathbf{\Lambda}_i. \tag{B.5}$$

D'après (i), $\mathbb{E}\left[U_i \partial_k U_i\right] = 0$, alors $\partial_l \mathbb{E}\left[U_i \partial_k U_i\right] = 0$, *i.e.* :

$$\mathbb{E}\left[\partial_k U_i \partial_l U_i\right] = -\mathbb{E}\left[U_i \partial^2_{kl} U_i\right]. \tag{B.6}$$

D'autre part la régression de $\partial^2_{kl} U_i$ sachant U_i est :

$$\mathbb{E}\left[\partial^2_{kl} U_i \mid U_i\right] = \mathbb{E}\left[\partial^2_{kl} U_i\right] + \mathrm{Cov}\left(U_i, \partial^2_{kl} U_i\right) \mathrm{Var}^{-1}(U_i) \left(U_i - \mathbb{E}\left[U_i\right]\right) = \mathbb{E}\left[U_i \partial^2_{kl} U_i\right] U_i \tag{B.7}$$

car $\mathbb{E}\left[U_i\right] = 0$, $\mathbb{E}\left[\partial^2_{kl} U_i\right] = 0$ et $\mathrm{Var}(U_i) = 1$. Le résultat découle de (B.6), (B.7) et $\mathbf{\Lambda}_i = \mathbb{E}\left[\dot{U}_i \dot{U}_i'\right]$. Regardons maintenant la covariance conditionnelle.

$$\begin{aligned}\mathrm{Cov}\left(\partial^2_{kl} U_i, \partial^2_{k'l'} U_i \mid U_i\right) &= \mathbb{E}\left[\partial^2_{kl} U_i \partial^2_{k'l'} U_i \mid U_i\right] - \mathbb{E}\left[\partial^2_{kl} U_i \mid U_i\right] \mathbb{E}\left[\partial^2_{k'l'} U_i \mid U_i\right] \\ &= \mathbb{E}\left[\partial^2_{kl} U_i \partial^2_{k'l'} U_i \mid U_i\right] - U_i^2 \mathbf{\Lambda}_{i\ [k,l]} \mathbf{\Lambda}_{i\ [k'l']}.\end{aligned}$$

Or d'après (B.7),

$$\begin{aligned}\partial^2_{kl} U_i &\overset{\mathcal{L}}{=} \mathrm{Cov}\left(\partial^2_{kl} U_i, U_i\right) U_i + \sqrt{\mathrm{Var}\left(\partial^2_{kl} U_i\right) - \mathrm{Cov}\left(\partial^2_{kl} U_i, U_i\right)^2}\, \varepsilon_{kl} \\ &\overset{\mathcal{L}}{=} -\mathbf{\Lambda}_{i\ [k,l]} U_i + \sqrt{\mathrm{Var}(\partial^2_{kl} U_i) - \mathbf{\Lambda}^2_{i\ [k,l]}}\, \varepsilon_{kl},\end{aligned} \tag{B.8}$$

où les égalités sont des égalités en loi et ε_{kl} est une variable aléatoire résiduelle $\mathcal{N}(0,1)$. Ainsi,

$$\begin{aligned}\mathrm{Cov}\left(\partial^2_{kl}U_i, \partial^2_{k'l'}U_i \mid U_i\right) &= \mathbb{E}\left[\left(-\mathbf{\Lambda}_{i\ [k,l]}U_i + \sqrt{\mathrm{Var}\left(\partial^2_{kl}U_i\right) - \mathbf{\Lambda}^2_{i\ [k,l]}}\,\varepsilon_{kl}\right)\right.\\ &\quad\left.\times\left(-\mathbf{\Lambda}_i[k',l']U_i + \sqrt{\mathrm{Var}\left(\partial^2_{k'l'}U_i\right) - \mathbf{\Lambda}^2_{i\ [k',l']}}\,\varepsilon_{k'l'}\right)\right]\\ &= \sqrt{\mathrm{Var}\left(\partial^2_{kl}U_i\right) - \mathbf{\Lambda}^2_{i\ [k,l]}}\sqrt{\mathrm{Var}\left(\partial^2_{k'l'}U_i\right) - \mathbf{\Lambda}^2_{i\ [k',l']}}\mathbb{E}\left[\varepsilon_{kl}\varepsilon_{k'l'}\right].\end{aligned}$$

(iv) D'après la proposition 5, $\ddot{\mathbf{U}}_i$ est indépendante de U_j et de $\dot{U}_j$ pour $i \neq j$ et $i, j = 1, 2$, aussi le conditionnement de $\ddot{\mathbf{U}}_i$ par $U, \dot{U}$ est équivalent au conditionnement par $U_i, \dot{U}_i$. Pour $\mathbf{V}_i = \left(U_i, \dot{U}_i\right)'$, nous pouvons écrire :

$$\mathbb{E}\left[\partial^2_{kl}U_i \mid \mathbf{V}_i\right] = \mathbb{E}\left[\partial^2_{kl}U_i\right] + \mathrm{Cov}\left(\partial^2_{kl}U_i, \mathbf{V}_i\right)\mathrm{Var}^{-1}\left(\mathbf{V}_i\right)\mathbf{V}_i,$$

avec $\mathrm{Cov}\left(\partial^2_{kl}U_i, \mathbf{V}_i\right) = \left(\mathbb{E}\left[\partial^2_{kl}U_iU_i\right] \quad \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]\right)$ et

$$\begin{aligned}\mathrm{Var}^{-1}\left(\mathbf{V}_i\right) &= \begin{pmatrix} \mathbb{E}\left[U_iU_i\right] & \mathbb{E}\left[U_i\partial_mU_i\right] & \mathbb{E}\left[U_i\partial_nU_i\right] \\ \mathbb{E}\left[U_i\partial_mU_i\right] & \mathbb{E}\left[\partial_mU_i\partial_mU_i\right] & \mathbb{E}\left[\partial_mU_i\partial_nU_i\right] \\ \mathbb{E}\left[U_i\partial_nU_i\right] & \mathbb{E}\left[\partial_mU_i\partial_nU_i\right] & \mathbb{E}\left[\partial_nU_i\partial_nU_i\right] \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mathbf{\Lambda}_{i[mm]} & \mathbf{\Lambda}_{i[mn]} \\ 0 & \mathbf{\Lambda}_{i[mn]} & \mathbf{\Lambda}_{i[nn]} \end{pmatrix}^{-1}\\ &= \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Lambda}_i \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Lambda}_i^{-1} \end{pmatrix}.\end{aligned}$$

Ainsi,

$$\mathbb{E}\left[\partial^2_{kl}U_i \mid \mathbf{V}_i\right] = -\mathbf{\Lambda}_{i[kl]}U_i + \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]\mathbf{\Lambda}_i^{-1}\dot{U}_i. \tag{B.9}$$

D'après (2.10), $\mathrm{Var}\left(\partial^2_{kl}U_i \mid \mathbf{V}_i\right) = \mathrm{Var}\left(\partial^2_{kl}U_i\right) - \underbrace{\mathrm{Cov}\left(\partial^2_{kl}U_i, \mathbf{V}_i\right)\mathrm{Var}^{-1}\left(\mathbf{V}_i\right)\mathrm{Cov}\left(\mathbf{V}_i, \partial^2_{kl}U_i\right)}_{(*)}$.

Or,

$$\begin{aligned}\mathrm{Var}\left(\partial^2_{kl}U_i\right) &= \mathrm{Var}\left(\partial^2_{kl}U_i \mid U_i\right) + \mathrm{Cov}\left(\partial^2_{kl}U_i, U_i\right)\mathrm{Var}^{-1}\left(U_i\right)\mathrm{Cov}\left(U_i, \partial^2_{kl}U_i\right)\\ &= \mathbf{M}_i(\mathbf{\Lambda}_i)_{[kl]} + \mathbf{\Lambda}^2_{i\ [kl]}\end{aligned}$$

et

$$\begin{aligned}(*) &= \left(\mathbb{E}\left[\partial^2_{kl}U_iU_i\right] \quad \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]\right)\begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Lambda}_i^{-1} \end{pmatrix}\begin{pmatrix} \mathbb{E}\left[\partial^2_{kl}U_iU_i\right] \\ \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right] \end{pmatrix}\\ &= \mathbf{\Lambda}^2_{i\ [kl]} + \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]\mathbf{\Lambda}_i^{-1}\mathbb{E}'\left[\partial^2_{kl}U_i\dot{U}_i\right].\end{aligned}$$

Ainsi, $\mathrm{Var}\left(\partial^2_{kl}U_i \mid \mathbf{V}_i\right) = \mathbf{M}_i(\mathbf{\Lambda}_i)_{[kl]} + \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]\mathbf{\Lambda}_i^{-1}{}_{[kl]}\mathbb{E}'\left[\partial^2_{kl}U_i\dot{U}_i\right]$. □

Annexe C

Propriétés du champ $T(\cdot)$

Afin d'alléger les notations, nous éliminerons souvent la dépendance en $\mathbf{x}$, *i.e.* nous noterons par exemple T à la place de $T(\mathbf{x})$.

Preuve de la proposition 7

La proposition 7 établit des propriétés sur les dérivées première et seconde du champ $T(\mathbf{x})$ conditionnelles à $U(\mathbf{x})$.
(i) D'après l'équation (6.4), $\dot{T} = 2(U_1\dot{U}_1 + U_2\dot{U}_2)$, ainsi conditionnellement à U, $\dot{T}$ est la somme de deux champs gaussiens, donc c'est un champ gaussien.
$\mathbb{E}\left[\dot{T} \mid U\right] = 0$ car $\dot{U}_i$ est d'espérance nulle et $\dot{U}_i$ est indépendant de U d'après le lemme 5 et la proposition 5.

$$\begin{aligned}
\mathrm{Var}(\dot{T} \mid U) &= 4\left(\mathrm{Var}(U_1\dot{U}_1 \mid U) + \mathrm{Var}(U_2\dot{U}_2 \mid U) + 2\mathrm{Cov}(U_1\dot{U}_1, U_2\dot{U}_2 \mid U)\right) \\
&= 4\left(U_1^2\mathrm{Var}(\dot{U}_1 \mid U_1) + U_2^2\mathrm{Var}(\dot{U}_2 \mid U_2) + 2U_1U_2\mathrm{Cov}(\dot{U}_1, \dot{U}_2 \mid U)\right), \\
&\quad \text{car } \dot{U}_i \text{ et } U \text{ sont indépendants} \\
&= 4U_1^2\mathbf{\Lambda}_1 + 4U_2^2\mathbf{\Lambda}_2, \text{ car } \dot{U}_1 \text{ et } \dot{U}_2 \text{ sont indépendants.}
\end{aligned}$$

(ii) L'équation (6.5) donne :

$$\ddot{\mathbf{T}} = 2\sum_{i=1}^{2}\left(\dot{U}_i\dot{U}_i' + U_i\ddot{\mathbf{U}}_i\right).$$

Comme $\mathbb{E}\left[\dot{U}_i\dot{U}_i'\right] = \mathbf{\Lambda}_i$ et $\mathbb{E}\left[U_i\ddot{\mathbf{U}}_i \mid U\right] = -U_i^2\mathbf{\Lambda}_i$ (équation B.5), nous avons :

$$\mathbb{E}\left[\partial^2_{kl}T \mid U\right] = 2\sum_{i=1}^{2}\left(1 - U_i^2\right)\mathbf{\Lambda}_{i[kl]}.$$

En ce qui concerne la variance de $\ddot{\mathbf{T}}_{|U}$, elle est obtenue en développant $\mathrm{Var}(\partial^2_{kl}T \mid U)$ à partir de (6.5) et en appliquant les résultats du lemme 5 :

$$\begin{aligned}
\mathrm{Var}\left(\partial^2_{kl}T \mid U\right) &= 4\mathrm{Var}\left(\sum_{i=1}^{2}\left(\partial_k U_i \partial_l U_i + U_i \partial^2_{kl}U_i\right) \mid U\right) \\
&= 4\left\{\sum_{i=1}^{2}\left[\mathrm{Var}\left(\partial_k U_i \partial_l U_i\right) + U_i^2 \mathrm{Var}\left(\partial^2_{kl}U_i \mid U_i\right)\right]\right. \\
&\quad +2\mathrm{Cov}\left(\partial_k U_1 \partial_l U_1, \partial_k U_2 \partial_l U_2\right) + 2U_1U_2\mathrm{Cov}\left(\partial^2_{kl}U_1, \partial^2_{kl}U_2 \mid U\right) \\
&\quad \left.+2\sum_{i,j=1}^{2} U_j \mathrm{Cov}\left(\partial_k U_i \partial_l U_i, \partial^2_{kl}U_j \mid U\right)\right\}
\end{aligned}$$

Le lemme 7 permet de montrer que :

$$\mathrm{Cov}\left(\partial_k U_i \partial_l U_i, \partial_k U_j \partial_l U_j\right) = 0 \text{ et } \mathrm{Cov}\left(\partial_k U_i \partial_l U_i, \partial^2_{kl}U_j \mid U\right) = 0, \ \forall i,j = 1,2.$$

D'après l'équation (B.8), $\partial^2_{kl}U_i \stackrel{\mathcal{L}}{=} -\mathbf{\Lambda}_{i\ [k,l]}U_i + \sqrt{\mathrm{Var}(\partial^2_{kl}U_i) - \mathbf{\Lambda}^2_{i\ [k,l]}}\,\varepsilon^i_{kl}$, ainsi :

$$\begin{aligned}
\mathrm{Cov}\left(\partial^2_{kl}U_1, \partial^2_{kl}U_2 \mid U\right) &= \mathbb{E}\left[\partial^2_{kl}U_1 \partial^2_{kl}U_2 \mid U\right] - \mathbb{E}\left[\partial^2_{kl}U_1 \mid U\right]\mathbb{E}\left[\partial^2_{kl}U_2 \mid U\right] \\
&= \sqrt{\mathrm{Var}(\partial^2_{kl}U_1) - \mathbf{\Lambda}^2_{1\ [k,l]}}\sqrt{\mathrm{Var}(\partial^2_{kl}U_1) - \mathbf{\Lambda}^2_{i\ [k,l]}}\,\mathbb{E}\left[\varepsilon^1_{kl}\varepsilon^2_{kl}\right] \\
&= 0 \text{ comme conséquence de la proposition 5.}
\end{aligned}$$

Finalement, $\mathrm{Var}\left(\partial^2_{kl}T \mid U\right) = 4\sum_{i=1}^{2} U_i^2 \mathrm{Var}\left(\partial^2_{kl}U_i \mid U_i\right) = 4\sum_{i=1}^{2} U_i^2 \mathbf{M}_i(\mathbf{\Lambda}_i)_{[kl]}$. □

Preuve de la proposition 8

Nous allons montrer que :

$$\partial^2_{kl}T(\mathbf{x})_{|U(\mathbf{x}),\dot{U}(\mathbf{x})} \sim \mathcal{N}\left(E_{\ddot{\mathbf{T}}_{kl}}, V_{\ddot{\mathbf{T}}_{kl}}\right),$$

avec :
$E_{\ddot{\mathbf{T}}_{kl}} = 2\sum_{i=1}^{2}\left(\partial_k U_i(\mathbf{x})\partial_l U_i(\mathbf{x}) - U_i^2(\mathbf{x})\mathbf{\Lambda}_i(\mathbf{x})_{[kl]} + U_i(\mathbf{x})\mathbb{E}\left[\partial^2_{kl}U_i(\mathbf{x})\dot{U}_i(\mathbf{x})\right]\mathbf{\Lambda}_i^{-1}(\mathbf{x})\dot{U}_i(\mathbf{x})\right)$
et $V_{\ddot{\mathbf{T}}_{kl}} = 4\sum_{i=1}^{2} U_i^2(\mathbf{x})\left(\mathbf{M}_i(\mathbf{\Lambda}_i(\mathbf{x}))_{[kl]} + \mathbb{E}\left[\partial^2_{kl}U_i(\mathbf{x})\dot{U}_i(\mathbf{x})\right]\mathbf{\Lambda}_i^{-1}(\mathbf{x})\mathbb{E}'\left[\partial^2_{kl}U_i(\mathbf{x})\dot{U}_i(\mathbf{x})\right]\right)$.

$\partial^2_{kl}T$ s'écrit $\partial^2_{kl}T = 2\sum_{i=1}^{2}\left(\partial_k U_i \partial_l U_i + U_i \partial^2_{kl}U_i\right)$. Alors,

$$\begin{aligned}\mathbb{E}\left[\partial^2_{kl}T \mid U,\dot{U}\right] &= 2\sum_{i=1}^{2}\mathbb{E}\left[\partial_k U_i\partial_l U_i + U_i\partial^2_{kl}U_i\right]\\ &= 2\sum_{i=1}^{2}\left(\partial_k U_i\partial_l U_i + U_i\mathbb{E}\left[\partial^2_{kl}U_i \mid U,\dot{U}\right]\right)\\ &= 2\sum_{i=1}^{2}\left(\partial_k U_i\partial_l U_i + U_i\left(-\mathbf{\Lambda}_{i\,[kl]}U_i + \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]\mathbf{\Lambda}_i^{-1}\dot{U}_i\right)\right)\\ &\quad \text{d'après le lemme 5}\end{aligned}$$

et

$$\begin{aligned}\mathrm{Var}\left(\partial^2_{kl}T \mid U,\dot{U}\right) &= 4\mathrm{Var}\left(\partial_k U_1\partial_l U_1 + \partial_k U_2\partial_l U_2 + U_1\partial^2_{kl}U_1 + U_2\partial^2_{kl}U_2 \mid U,\dot{U}\right)\\ &= 4\Big\{\mathrm{Var}\left(\partial_k U_1\partial_l U_1 \mid U,\dot{U}\right) + \mathrm{Var}\left(\partial_k U_2\partial_l U_2 \mid U,\dot{U}\right)\\ &\quad + \mathrm{Var}\left(U_1\partial^2_{kl}U_1 \mid U,\dot{U}\right) + \mathrm{Var}\left(U_2\partial^2_{kl}U_2 \mid U,\dot{U}\right)\\ &\quad + 2\mathrm{Cov}\left(\partial_k U_1\partial_l U_1, \partial_k U_2\partial_l U_2 \mid U,\dot{U}\right) + 2\mathrm{Cov}\left(\partial_k U_1\partial_l U_1, U_1\partial^2_{kl}U_1 \mid U,\dot{U}\right)\\ &\quad + 2\mathrm{Cov}\left(\partial_k U_1\partial_l U_1, U_2\partial^2_{kl}U_2 \mid U,\dot{U}\right) + 2\mathrm{Cov}\left(\partial_k U_2\partial_l U_2, U_1\partial^2_{kl}U_1 \mid U,\dot{U}\right)\\ &\quad + 2\mathrm{Cov}\left(\partial_k U_2\partial_l U_2, U_2\partial^2_{kl}U_2 \mid U,\dot{U}\right) + 2\mathrm{Cov}\left(U_1\partial^2_{kl}U_1, U_2\partial^2_{kl}U_2 \mid U,\dot{U}\right)\Big\}.\end{aligned}$$

Or,

• $\mathrm{Var}\left(\partial_k U_i\partial_l U_i \mid U,\dot{U}\right) = \mathrm{Var}\left(\partial_k U_i\partial_l U_i \mid U_1,U_2,\partial_k U_1,\partial_k U_2,\partial_l U_1,\partial_l U_2\right) = 0,\ \forall i = 1,2$ car si $\mathbf{X}$ est une variable aléatoire et si $h(\cdot)$ est de carré intégrable, $\mathbb{E}\left[h(\mathbf{X}) \mid \mathbf{X}\right] = h(\mathbf{X})$ et

$$\begin{aligned}\mathrm{Var}\left(h(\mathbf{X}) \mid \mathbf{X}\right) &= \mathbb{E}\left[h(\mathbf{X})h'(\mathbf{X}) \mid \mathbf{X}\right] - \mathbb{E}\left[h(\mathbf{X}) \mid \mathbf{X}\right]\mathbb{E}'\left[h(\mathbf{X}) \mid \mathbf{X}\right]\\ &= h(\mathbf{X})h'(\mathbf{X}) - h(\mathbf{X})h'(\mathbf{X}).\end{aligned}$$

• $\mathrm{Cov}\left(\partial_k U_i\partial_l U_i, U_j\partial^2_{kl}U_j \mid U,\dot{U}\right)$

$$\begin{aligned}&= \mathbb{E}\left[\partial_k U_i\partial_l U_i U_j\partial^2_{kl}U_j \mid U,\dot{U}\right] - \mathbb{E}\left[\partial_k U_i\partial_l U_i \mid U,\dot{U}\right]\mathbb{E}\left[U_j\partial^2_{kl}U_j \mid U,\dot{U}\right]\\ &= \partial_k U_i\partial_l U_i U_j\mathbb{E}\left[\partial^2_{kl}U_j \mid U,\dot{U}\right] - \partial_k U_i\partial_l U_i U_j\mathbb{E}\left[\partial^2_{kl}U_j \mid U,\dot{U}\right]\\ &= 0,\ \forall i,j = 1,2.\end{aligned}$$

• $\mathrm{Cov}\left(\partial_k U_1\partial_l U_1, \partial_k U_2\partial_l U_2 \mid U,\dot{U}\right)$

$$\begin{aligned}&= \mathbb{E}\left[\partial_k U_1\partial_l U_1\partial_k U_2\partial_l U_2 \mid U,\dot{U}\right] - \mathbb{E}\left[\partial_k U_1\partial_l U_1 \mid U,\dot{U}\right]\mathbb{E}\left[\partial_k U_2\partial_l U_2 \mid U,\dot{U}\right]\\ &= \partial_k U_1\partial_l U_1\partial_k U_2\partial_l U_2 - \partial_k U_1\partial_l U_1\partial_k U_2\partial_l U_2\\ &= 0.\end{aligned}$$

• D'après l'équation (B.9) nous avons :

$$\begin{aligned}\partial^2_{kl}U_i &\overset{\mathcal{L}}{=} \mathbb{E}\left[\partial^2_{kl}U_i \mid U,\dot{U}\right] + \sqrt{\mathrm{Var}\left(\partial^2_{kl}U_i\right) - \left(\mathbf{\Lambda}^2_{i\,[kl]} + \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]\mathbf{\Lambda}_i^{-1}\mathbb{E}'\left[\partial^2_{kl}U_i\dot{U}_i\right]\right)}\varepsilon^i_{kl}\\ &= \mathbb{E}\left[\partial^2_{kl}U_i \mid U,\dot{U}\right] + \sqrt{\mathbf{M}_i(\mathbf{\Lambda}_{i,[kl]}) - \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]\mathbf{\Lambda}_i^{-1}{}_{[kl]}\mathbb{E}'\left[\partial^2_{kl}U_i\dot{U}_i\right]}\varepsilon^i_{kl}\\ &= \mathbb{E}\left[\partial^2_{kl}U_i \mid U,\dot{U}\right] + s^i_{kl}\varepsilon^i_{kl}.\end{aligned}$$

Alors,

$$\begin{aligned}
\mathrm{Cov}\left(\partial^2_{kl}U_1, \partial^2_{mn}U_2 \mid U, \dot{U}\right) &= s^1_{kl}s^2_{mn}\mathbb{E}\left[\varepsilon^1_{kl}\varepsilon^2_{mn} \mid U, \dot{U}\right] \\
&= s^1_{kl}s^2_{mn}\mathbb{E}\left[\varepsilon^1_{kl}\varepsilon^2_{mn}\right] \text{ car } \varepsilon^1_{kl} \text{ et } \varepsilon^2_{mn} \text{ sont indépendants} \\
&\quad \text{de } U \text{ et } \dot{U} \\
&= \frac{1}{s^1_{kl}s^2_{mn}}\left\{\mathbb{E}\left[\partial^2_{kl}U_1\partial^2_{mn}U_2\right] - \mathbb{E}\left[\partial^2_{mn}U_2 \mid U, \dot{U}\right]\mathbb{E}\left[\partial^2_{kl}U_1\right]\right. \\
&\quad -\mathbb{E}\left[\partial^2_{kl}U_1 \mid U, \dot{U}\right]\mathbb{E}\left[\partial^2_{mn}U_2\right] \\
&\quad \left.+\mathbb{E}\left[\mathbb{E}\left[\partial^2_{kl}U_1 \mid U, \dot{U}\right]\mathbb{E}\left[\partial^2_{mn}U_2 \mid U, \dot{U}\right]\right]\right\} \\
&= 0 \text{ d'après la proposition 5.}
\end{aligned}$$

Finalement,

$$\begin{aligned}
\mathrm{Var}\left(\partial^2_{kl}T \mid U, \dot{U}\right) &= 4\sum_{i=1}^{2}\mathrm{Var}\left(\partial^2_{kl}U_i \mid U, \dot{U}\right) \\
&= 4\sum_{i=1}^{2}\left\{U_i^2\left(\mathbf{M}_i(\mathbf{\Lambda}_{i,[kl]}) + \mathbb{E}\left[\partial^2_{kl}U_i\dot{U}_i\right]{\mathbf{\Lambda}_i^{-1}}_{[kl]}\mathbb{E}'\left[\partial^2_{kl}U_i\dot{U}_i\right]\right)\right\}
\end{aligned}$$

□

Bibliographie

[1] Adams R. Contour mapping and differential systematics of geographic variation. *Systematic Zoology*, 19 :385–390, 1970.

[2] Adler R. On excursion sets, tube formulas and maxima of random fields. *The Annals of Applied Probability*, 10(1) :1–74, 2000.

[3] Adler R. *Geometry of Random Fields*. John Wiley and Sons, 1981.

[4] Aldous D. *Probability approximations via the poisson clumping heuristic*. Springer, New-York, 1989.

[5] Allemand B. Comparaison de deux modélisations probabilistes pour la classification de données géostatistiques : construction et analyse des algorithmes d'estimation. Technical report, Institut National de la Recherche Agronomique et Institut Universitaire Professionnalisé d'Avignon, 2004.

[6] Aronowich M. and Adler R. Extrema and level crossings of χ^2 processes. *Advances in Applied Probability*, 18 :901–920, 1986.

[7] Aronowich M. and Adler R. Sample path behaviour of χ^2 surfaces at extrema. *Advances in Applied Probability*, 20 :719–738, 1988.

[8] Azaïs J-M. and Delmas C. Asymptotic expansions for the distribution of the maximum of gaussian random fields. *Extremes*, 5(2) :183–214, 2002.

[9] Barbujani G., Oden N., and Sokal R. Detecting regions of abrupt change in maps of biological variables. *Systematic Zoology*, 28 :376–389, 1989.

[10] Benjamini Y. and Hochberg Y. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1) :289–300, 1995.

[11] Besag J. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, serie B*, 48 :259–302, 1986.

[12] Besag J., Green P., Higdon D., and Mengersen K. Bayesian computation and stochastic systems (with discussion). *Statistical Science*, 10 :3–66, 1995.

[13] Bocquet-Appel J.P. and Bacro J.N. Generalized wombling. *Systematic Biology*, 43(3) :442–448, 1994.

[14] Cao J. *Excursion sets of random fields with application to human brain mapping*. PhD thesis, McGill University, 1997.

[15] Cao J. The size of the connected components of excursion sets of χ^2, t and F fields. *Advances in Applied Probability*, 31 :579–595, 1999.

[16] Castrignanò A., Maiorana M, Pisante M, Vonella A.V., and Fecondo G. A methodology to define soil units in support to precision agriculture. In Werner A. and Jarfe A., editors, *Program book of the joint conference of ECPA-ECPLF*, 2003.

[17] Chaudhuri P. and Marron J. SiZer for exploration of structures in curves. *Journal of the American Statistical Association, Theory and Method*, 1999.

[18] Chaudhuri P. and Marron J. Scale space view of curve estimation. *The Annals of Statistics*, 28(2) :408–428, 2000.

[19] Chilès J-P. and Delfiner P. *Geostatistics : modeling spatial uncertainty*. Wiley, New-York, 1999.

[20] *Precision agriculture : spatial and temporal variability of environmental quality*. Ciba Foundation Symposium 210. Wiley, 1997.

[21] Cressie N. *Statistics for spatial data*. Wiley, New-York, revised edition, 1993.

[22] Delmas C. *Distribution du maximum d'un champ aléatoire et applications statistiques*. PhD thesis, Université Paul Sabatier, Toulouse, 2001.

[23] Friston K., Frith C., Liddle P., and Frackowiak R. Comparing functional (PET) images : the assessment of significance change. *Journal of Cerebral Blood Flow and Metabolism*, 11 :690–699, 1991.

[24] Friston K., Worsley K., Frackowiak R., Mazziotta J., and Evans A. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1 :214–220, 1994.

[25] Geman S. and Geman D. Stochastic relaxation, Gibbs distribution and the bayesian restoration of images. *IEEE Trans. Pattern Analysis end Machine Intelligence*, (6) :721–741, 1984.

[26] Gilliot J-M., Jullien A., Michelin J., and Huet P. *Essai de zonage d'un champ à partir de cartes de rendement*. Science update edition, 2004. in press.

[27] Gleyze J.F., Bacro J.N., and Allard D. Detecting regions of abrupt change : wombling procedure and statistical significance. In *geoENV III : Geostatistics for Environmental Applications*, 2001.

[28] Godtliebsen F., Marron J., and Chaudhuri P. Significance in scale space. December 2001.

[29] Godtliebsen F., Marron J., and Pizer S. *Spatial Clustering Modelling*, chapter Significance in scale-space for clustering. Chapman and Hall/CRC, 2002.

[30] Guérif M., Beaudoin N., Durr C., Houlès V., Machet J-M., Mary B., Moulin S., and Richard G. Designing a field experiment for assessing soil and crop spatial variability and defining site-specific management strategies. pages 677–682. Proceedings of the Third European Conference on Precision Agriculture, 2001.

[31] Hall P. and Rau C. Local likelihood tracking of fault lines and boundaries. *Journal of the Royal Statistical Society B*, 63(3) :569–582, 2001.

[32] Hart J. Central tendency in areal distributions. *Economic Geographyy*, (30) :48–59, 1954.

[33] Houlès V. *Mise au point d'un outil de modulation intra-parcellaire de la fertilisation azotée du blé d'hiver basé sur la télédétection et un modèle de culture.* PhD thesis, Institut National Agronomique Paris-Grignon, 2004.

[34] Lantuéjoul C. Ergodicity and integral range. *Journal of Microscopy*, 161 :387–403, 1991.

[35] Lawson A. and Denison D. *Spatial clustering modelling.* Chapman and Hall/CRC, 2002.

[36] Mary B., Beaudoin N., Machet J.M., Bruchou C., and Ariès F. Characterization and analysis of soil variability within two agricultural fields : the case of water and mineral N profiles. In Grenier and Blackmore, editors, *Proceedings of the Third European Conference on Precision Agriculture*, pages 431–436, 2001.

[37] Matheron G. *Traité de géostatistique appliquée, Tome I.* Number 14. Technip Paris edition, 1962.

[38] Matheron G. Principles of geostatistics. *Economic Geology*, (58) :1246–1266, 1963.

[39] Miller K. *Multidimensional Gaussian distribution.* SIAM series in Applied Mathematics, 1964.

[40] Nicoullaud B., Beaudoin N., Roque J., Couturier A., Maucorps J., and King D. Relevés pédologiques et agriculture de précision. In *Journée de restitution de l'AIP "Agriculture de précision"*, 2003.

[41] Peyrard N. *Approximations de type champ moyen de modèles de champ de Markov pour la segmentation de données spatiales*. PhD thesis, Université Joseph Fourier - Grenoble I, 2001.

[42] Qiu P. Discontinuous regression surfaces fitting. *The Annals of Statistics*, 26 :2218–2245, 1998.

[43] Shatar T. and McBratney A. Subdividing a field into contiguous management zones using k-zones algorithm. In *Third European Conference on Precision Agriculture*, volume 1, pages 115–120. agro Montpellier, 2001.

[44] Shen X., Huang H-C., and Cressie N. Nonparametric hypothesis testing for a spatial signal. *Journal of the American Statistical Association, Theory and Method*, 97(460) :1122–1140, 2002.

[45] Stein M. *Interpolation for spatial data : some theory for kriging*. Springer-Verlag, 1999.

[46] Wang Y. Change curve estimation via wavelets. *Journal of the American Statistical Association*, 93 :163–172, 1998.

[47] Whelan B. and McBratney A. Definition and interpretation of potential management zones in australia. In *Proceedings of the 11th Australian Agronomy Conference*, 2003.

[48] Womble W. Differential systematics. *Science*, 114(2961) :315–322, 1951.

[49] Worsley K., Evans A., Marrett S., and Neelin P. A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12 :900–918, 1992.

[50] Worsley K. Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Advances in Applied Probability*, 26 :13–42, 1994.

[51] Worsley K. Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *The Annals of Statistics*, 23(2) :640–669, 1995.

[52] Worsley K. Testing for signals with unknown location and scale in a χ^2 random field, with an application to fMRI. *Advances in Applied Probability*, 33(4) :773–793, 2001.

[53] Yaglom A. *Correlation theory of stationary and related random functions*. Springer-Verlag, Berlin, 1986.

[54] Ye J. Measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, 98 :120–131, 1998.

Table des figures

1.1 *Représentation d'un pixel de la grille d'interpolation.* 19
1.2 *Décomposition de* W. 21
1.3 *a) Notations, b) Décomposition de* $\widetilde{W}$. 22

2.1 *Illustration d'un minimum a), d'un maximum b) et d'un point col c).* . . . 41

3.1 *Représentation d'un champ aléatoire (en noir) et de sa prédiction (en orange) passant par les points d'échantillonnage (cercles rouges). Les carrés verts correspondent aux points en lesquels nous voulons prédire la dérivée du champ.* 53
3.2 *Modèles de covariance gaussien, exponentiel et sphérique, pour une portée fixée à* $b = 0.25$. 55
3.3 *Exemples de modèles de Matérn en fonction des paramètres de portée,* α^{-1}, *et de lissage,* ν. 56

4.1 *Illustration des propriétés de la définition axiomatique d'une composante connexe (définition 6).* 61
4.2 *a) Image artificielle dans* $\mathbb{R}^2$. *Ensemble d'excursion* $\mathcal{A}_z$ *au dessus de : b)* $z = 6$, *c)* $z = 10.5$ *et d)* $z = 13.4$. 63

5.1 *Echantillon de 100 points répartis aléatoirement, issus d'un processus : a) gaussien stationnaire, centré-réduit, de fonction de covariance exponentielle de portée* $b = 0.1$, *b) avec une discontinuité le long des pointillés (*$a = 2.5$*). c) ZCAs potentielles (en gris) obtenues sous l'alternative.* 78

6.1 *Représentation de la section horizontale,* S_0, *du paraboloïde elliptique,* $T(\mathbf{x})$ *dans un voisinage du maximum local* $\mathbf{0}$, *à la distance* $T_0 - t_\alpha$ *à partir du maximum.* 89

6.2 *Echantillon de 100 points répartis aléatoirement, issus d'un processus : a) stationnaire, b) avec une discontinuité le long des pointillés. Les ZCAs significatives sont colorées en noir et les non significatives en gris : c). La grille utilisée est de dimension* 60×60. 93

6.3 *ZCAs (significatives en noir et non significatives en gris) obtenues pour différentes valeurs de* $1-\alpha$ *: a)* $1-\alpha = 0.995$*, b)* $1-\alpha = 0.9994$ *et c)* $1-\alpha = 0.99995$. 94

7.1 *Histogramme et qqplot de la statistique* $X = t_\alpha S_{t_\alpha} \det(\mathbf{\Lambda})^{1/2}/\pi$ *comparée à une exponentielle d'espérance 2, dans le cas d'une grille* 30×30 *(à gauche) et* 60×60 *(à droite), et pour* $1-\alpha = 0.95$ *(première ligne) et* $1-\alpha = 0.995$ *(deuxième ligne).* 96

7.2 *Evolution du niveau* $1-\alpha_G$ *en fonction de la portée b d'un modèle de covariance exponentiel.* 103

7.3 *Histogramme et qqplot de la statistique* $X = t_\alpha S_{t_\alpha} \det(\mathbf{\Lambda})^{1/2}/\pi$ *comparée à une exponentielle d'espérance 2,* $E(2)$*, pour un ensemble de 100 simulations, dans le cas d'une grille* 30×30 *(à gauche) et* 60×60 *(à droite), et pour* $1-\alpha_{[i,j]}$ *(première ligne),* $1-\alpha_{moy}$ *(deuxième ligne),* $1-\alpha_G$ *(troisième ligne) et* $1-\hat{\alpha}$ *(dernière ligne).* 105

7.4 *ZCAs potentielles obtenues pour la simulation utilisée pour illustrer la méthode, en fonction du niveau* $1-\alpha$ *:* $1-\alpha_{[i,j]}$*,* $1-\alpha_{moy}$*,* $1-\alpha_G$ *et* $1-\hat{\alpha}$ *(de gauche à droite), et de la discrétisation :* 30×30 *(première ligne) et* 60×60 *(deuxième ligne).* 106

7.5 *Cconfigurations de points* $\mathbf{x}_i$*,* $\mathbf{x}_j$ *pour lesquels les couples* $\{Z(\mathbf{x}_i),\ Z(\mathbf{x}_j)\}$ *ne sont pas pris en compte dans le calcul du variogramme expérimental.* . . 108

7.6 *Application de la procédure itérative sur l'exemple utilisé pour illustrer la méthode. a) Echantillon de points considéré avec une discontinuité* $a = 2.5$ *introduite en* $x^1 = 0.4$*. Variogrammes obtenus à la première itération, b), et à la convergence, c). ZCAs détectées (significatives en noir, non significatives en gris) en considérant la covariance connue, d), à la première itération, e) et à la convergence, f).* 109

7.7 *a) Echantillon de points (proportionnels à leur valeur), avec introduction d'une valeur abérante (cercle plein). b) ZCAs (significatives en noir) obtenues à partir de cet échantillon.* 110

8.1 *ZCAs (significatives en noir et non significatives en gris) au niveau* $1-\hat{\alpha}$*, obtenues en utilisant le krigeage simple (*$1-\hat{\alpha}=0.9994$*), a), ou le krigeage ordinaire (*$1-\hat{\alpha}=0.9993$*), b), comme interpolateur, à partir de l'exemple utilisé pour illustrer la méthode.* 113

8.2 *a) ZCAs significatives (en noir) au niveau* $1-\hat{\alpha}$*. b) ZCAs significatives au niveau* $1-\hat{\alpha}$ *représentées au niveau* $1-\alpha_0=0.99$*.* 115

8.3 *Exemple de ZCA au niveau* $1-\alpha_0$ *(en gris) contenant une ZCA significative au niveau* $1-\alpha$ *(en noir).* 116

8.4 *Illustration de l'amélioration de la procédure itérative. Variogrammes et ZCAs à la première itération : a) et f), à la convergence pour* $1-\hat{\alpha}$ *: b) et g), pour* $1-\alpha_0=0.99$ *: c) h), pour* $1-\alpha_0=0.95$ *: d) et i), pour* $1-\alpha_0=0.90$ *: e) et j).* 117

9.1 *Puissance en* $x\in[0,1]$ *dans le cas d'un modèle de covariance exponentiel de portée* $b\in\{0.05,0.1,0.2,0.4,0.8\}$*, pour le krigeage simple (deux premières lignes) avec a)* $(a_1,a_2)=(1,-1)$*, b)* $(1.5,-1.5)$ *et c)* $(2,-2)$*, d)* $(a_1,a_2)=(2,0)$*, e)* $(2,-1)$ *et f)* $(3,-1)$ *et pour le krigrage ordinaire (dernière ligne) avec g)* $(a_1,a_2)=(1,-1)$*, h)* $(1.5,-1.5)$ *et i)* $(2,-2)$*.* 130

9.2 *Puissance en* $x\in[0,1]$ *dans le cas d'un modèle de covariance sphérique de portée* $b\in\{0.05,0.1,0.2,0.4,0.8\}$*, pour le krigeage simple (deux premières lignes) avec a)* $(a_1,a_2)=(1,-1)$*, b)* $(1.5,-1.5)$ *et c)* $(2,-2)$*, d)* $(a_1,a_2)=(2,0)$*, e)* $(2,-1)$ *et f)* $(3,-1)$ *et pour le krigrage ordinaire (dernière ligne) avec g)* $(a_1,a_2)=(1,-1)$*, h)* $(1.5,-1.5)$ *et i)* $(2,-2)$*.* 131

10.1 *Modèle utilisé pour l'alternative dans le calcul de la puissance en* $\mathbf{x}\in\mathcal{D}\subset\mathbb{R}^2$*.* 134

10.2 *a) Exemple d'échantillon de 100 points répartis aléatoirement dans un carré de côté un, issus d'un champ gaussien standard de covariance exponentielle de portée* $b=0.1$*. Une discontinuité est introduite en* $y=0.5$*. b) Puissance moyenne en un point sur 1000 échantillons du type a), le long de la discontinuité, pour* $a\in\{0,2,3,4\}$*.* 137

10.3 *Trait plein : puissance en fonction du nombre de directions et de la valeur de la discontinuité. Les échantillons considérés sont constitués de 100 points répartis aléatoirement dans un carré de côté un, issus d'un champ gaussien standard de covariance exponentielle de portée* $b = 0.1$. *Pointillés : les droites horizontales correspondent à la valeur de la puissance pour 4 directions, et la droite verticale à* $n_\theta = 4$. *a) moyenne de 100 échantillons, b) pour un échantillon.* 138

10.4 *a) Discontinuité et couples de points pour lesquels la puissance en deux points est calculée, avec un modèle de covariance exponentiel de portée* $b = 0.1$ *b) Trait plein : puissance moyenne en deux points sur 1000 échantillons. Pointillés : puissance en deux points en supposant les tests indépendants.* $a = 0$ *(noir),* $a = 2$ *(rouge),* $a = 3$ *(vert) et* $a = 4$ *(bleu).* 141

10.5 *Modèle de discontinuité utilisé pour l'alternative dans le calcul de la puissance conditionnelle locale.* 142

10.6 *a) Echantillon de points présentant une densité variable sur le domaine ; b), c), d) Puissance calculée pour une discontinuité* $a = 2.5$ *et pour différentes valeurs du paramètre* k *:* $k = 0, 1, 3$ *respectivement.* 144

10.7 *Courbe de puissance (en noir pour une discontinuité de* $a = 0$*, en rouge pour* $a = 2$*, en vert pour* $a = 2.5$ *et en bleu pour* $a = 3$*) en fonction du paramètre* k *(demie longueur de la fenêtre carrée), calculée en le centre d'un carré de côté un dans lequel ont été générés 100 points répartis aléatoirement, issus d'un champ gaussien de covariance exponentielle de portée* $b = 0.05$*, pour* $1-\hat{\alpha} = 0.999$. *La droite horizontale correspond à* $y = \eta$*, et la droite verticale à* $x =$ *portée pratique.* 145

10.8 *Erreur commise (pointillés verts) en fonction de la taille de la fenêtre (en gris), en considérant une discontinuité représentée par une droite passant par* $\mathbf{x}$ *(en rouge) au lieu du vrai modèle de discontinuité (en vert), modèle de type : a) parabole, b) elliptique.* 147

10.9 *Courbe de puissance en fonction du paramètre* k *(demi longueur de la fenêtre carrée), calculée en différents points (a) : points rouge, vert et bleu) d'un carré de côté un dans lequel ont été générés 100 points issus d'un champ gaussien de covariance exponentielle de portée* $b = 0.05$*, pour* $1-\hat{\alpha} = 0.9994$ *et différentes valeurs de la discontinuité : b)* $a = 0$*, c)* $a = 2$*, d)* $a = 2.5$ *et e)* $a = 3$. *Sur la figure b) la droite horizontale correspond à* $y = \eta$*, et la droite verticale à* $x =$ *portée pratique.* 148

10.10*a) : Echantillon de points issus d'un processus avec une discontinuité le long des pointillés ; b) : ZCAs (significatives en noir) détectées ; c) : les pointillés délimitent la zone permettant de définir les vraies ZCAs et la zone grise représente les bords que nous éliminons ; d) : ZCAs (significatives en noir) obtenues après élimination des bords.* 156

11.1 *Histogrammes du cumul des coordonnées i des pixels* $[i, j]$ *correspondants à des ZCAs significatives pour différents niveaux* $1 - \alpha$ *(*$1 - \hat{\alpha}$ *en blanc et 0.995 en noir) pour 100 simulations. Le cumul des pixels correspondant à la discontinuité (*$a = 2.5$*) en* $i = 24$ *a été tracé de couleur différente. Les bords ont été conservés figure a) et éliminés figure b).* 158

11.2 *Histogrammes des coordonnées i des pixels* $[i, j]$ *constituant les ZCAs significatives au niveau* $1-\hat{\alpha}$ *pour 100 simulations, en fonction de la densité d'échantillonnage : a)* $n = 50$*, b)* $n = 100$ *et c)* $n = 200$*. Le cumul des pixels correspondant à la discontinuité (*$a = 2.5$*) en* $i = 24$ *a été tracé de couleur différente.* 160

12.1 *Effet de la discrétisation sur la détection des ZCAs au niveau* $1-\hat{\alpha}$ *(significatives en noir, non significatives en gris).* 163

12.2 *Histogrammes des coordonnées i des pixels* $[i, j]$ *constituant les ZCAs significatives au niveau* $1-\hat{\alpha}$ *pour 100 simulations, en fonction de la discrétisation : a)* 15×15*, b)* 30×30 *et c)* 60×60*. Le cumul des pixels correspondant à la discontinuité (*$a = 2.5$*) a été tracé de couleur différente.* 165

12.3 *Histogrammes du cumul des coordonnées i des pixels* $[i, j]$ *correspondants à des ZCAs significatives au niveau* $1 - \hat{\alpha}$*, pour 100 simulations. Les interpolateurs utilisés sont le krigeage simple (première ligne) et le krigeage ordinaire (deuxième ligne). Le cumul des pixels correspondant à la discontinuité (*$i = 24$*) a été tracé de couleur différente. Les bords ont été conservés figures a) et c) et éliminés figures b) et d).* 167

13.1 *Carte pédologique [Nicoullaud* et al. *(2003)].* 176

13.2 *a) Carte pédologique simplifiée ; b) Schéma d'échantillonnage.* 177

13.3 *Points d'échantillonnage (proportionnels à leur valeur) et interpolation de la teneur en eau par krigeage ordinaire (valeurs fortes en blanc et faibles en noir). La même échelle est utilisée pour illustrer les interpolations.* . . . 178

13.4 *Points d'échantillonnage (proportionnels à leur valeur) et interpolation de la teneur en azote minéral (valeurs fortes en blanc, valeurs faibles en noir). Des échelles différentes sont utilisées pour illustrer les interpolations.* . . . 180
13.5 *Interpolation des variables permanentes (valeurs élevées en blanc et valeurs faibles en noir).* 182

14.1 *a) Points d'échantillonnage (proportionnels à leur valeur) de la teneur en eau en mars 2000. b) Interpolation sur une grille* 62×98 *par krigeage ordinaire. c) et d) (resp. e) et f), g) et h), i) et j), k) et l)) Ajustement du variogramme et ZCAs détectées à la première (resp. deuxième, troisième, quatrième, cinquième) itération. Les ZCAs significatives sont représentées en noir et les non significatives en gris.* 184
14.2 *ZCAs obtenues à la convergence pour la teneur en eau en mars 2000, avec : a)* $1 - \alpha = 1 - \alpha_G$*, b)* $1 - \hat{\alpha}$ *dans la procédure itérative.* 186
14.3 *ZCAs (significatives en noir, non significatives en gris) détectées pour la teneur en eau à la convergence.* 187
14.4 *ZCAs (significatives en noir, non significatives en gris) détectées pour les variables permanentes à la convergence.* 189
14.5 *ZCAs (significatives en noir, non significatives en gris) obtenues pour la teneur en eau en juillet 2002, avec un modèle de covariance a) exponentiel, au niveau* $1 - \hat{\alpha} = 0.9985$ *; b) sphérique, au niveau* $1 - \hat{\alpha} = 0.9964$. 190
14.6 *ZCAs (significatives en noir, non significatives en gris) obtenues pour la teneur en eau en juillet 2002, au niveau* $1 - \alpha_G$ *(première ligne) ou* $1 - \hat{\alpha}$ *(deuxième ligne), avec différentes discrétisations :* 31×49 *(a) et d)),* 62×98 *(b) et e)) et* 155×245 *(c) et f)).* 191
14.7 *ZCAs (significatives en noir) obtenues pour la teneur en eau en juillet 2002 au niveau* $1 - \alpha_G$*, en utilisant comme interpolateur a) le krigeage ordinaire, b) le krigeage simple.* 192
14.8 *Echantillon de points de la teneur en eau en juillet 2002 obtenus en éliminant* m *% des points de l'échantillon initial. a)* $m = 0$*, b)* $m = 10$*, c)* $m = 20$*, d)* $m = 30$*, e)* $m = 40$*, f)* $m = 50$*, g)* $m = 60$*, h)* $m = 70$*. Les points gris correspondent aux points éliminés et les points noirs aux points conservés.* . 192
14.9 *ZCAs (significatives en noir, non significatives en gris) obtenues à la convergence du cas I pour la teneur en eau en juillet 2002 en éliminant* m *% des points de l'échantillon initial. a)* $m = 0$*, b)* $m = 10$*, c)* $m = 20$*, d)* $m = 30$*, e)* $m = 40$*, f)* $m = 50$*, g)* $m = 60$*, h)* $m = 70$*.* 194

14.10 *ZCAs (significatives en noir) obtenues à la convergence du cas II pour la teneur en eau en juillet 2002 en éliminant m % des points de l'échantillon initial. a) $m = 0$, b) $m = 10$, c) $m = 20$, d) $m = 30$, e) $m = 40$, f) $m = 50$, g) $m = 60$, h) $m = 70$.* 195

14.11 *Puissance calculée en le centre d'une fenêtre de dimension 7×7, pour une discontinuité de 3 écarts-type pour l'échantillon de la teneur en eau en juillet 2002 en éliminant m % des points de l'échantillon initial. a) $m = 0$, b) $m = 10$, c) $m = 20$, d) $m = 30$, e) $m = 40$, f) $m = 50$, g) $m = 60$, h) $m = 70$.* 196

14.12 *Cartographie de la puissance calculée pour mars 2000 (première ligne) et pour octobre 2000 (deuxième ligne) pour une discontinuité de 3 écarts-type et pour différentes tailles de la fenêtre $dim(\mathcal{F})$: a) et e) puissance ponctuelle, b) et f) $dim(\mathcal{F}) = 3 \times 3$, c) et g) $dim(\mathcal{F}) = 7 \times 7$, d) et h) $dim(\mathcal{F}) = 11 \times 11$.* 197

Liste des tableaux

1.1 *Configurations de test possibles proposées par Gleyze et al. (2001).* 26

1.2 *Configurations des courbures en fonction des valeurs propres.* 31

6.1 *Surface S et valeur critique p des ZCAs détectées (figure 6.2c).* 93

7.1 *Nombre d'erreurs commises lors du test de N hypothèses.* 98

7.2 *Valeurs de* $1-\alpha_{[i,j]}$ *obtenues en chaque pixel de la grille d'interpolation pour un ensemble de 100 simulations.* 103

7.3 *Valeurs de* $1-\hat{\alpha}$ *pour une discrétisation* 30×30 *et* 60×60 *pour un ensemble de 100 simulations utilisant un modèle de covariance exponentiel de portée* $b=0.1$. 103

8.1 *Portée b et variance* σ^2 *estimées à la première itération et à la convergence pour différentes valeurs de* $1-\alpha_0$. 116

9.1 *Prédicteur de* $Z(x)$ *:* $Z^*(x)$*, gradient de* $Z^*(x)$ *:* $W(x)$ *et variance de* $W(x)$ *:* $\Sigma(x)$*, en fonction de l'interpolateur de* $Z(x)$. 124

10.1 *Dimension de la matrice de covariance* $\mathbf{\Sigma}_{\mathbf{V}_\theta}$ *en fonction de la taille de la fenêtre.* 144

11.1 *Nombre de simulations (sur 100) ayant des vraies ZCAs et/ou des autres ZCAs pour différentes valeurs de* $1-\alpha$ *(grille* 60×60*,* $n=100$*, covariance exponentielle de portée* $b=0.1$*).* 157

11.2 *Nombre de simulations, N, sur 100 présentant au moins une vraie ZCA et taille moyenne,* $\bar{s}$*, des ZCAs détectées, en fonction de la densité de points n et de la valeur de la discontinuité a.* 159

12.1 *Nombre, N, de simulations (sur 100) ayant des ZCAs significatives et taille moyenne, $\bar{s}$, des ZCAs détectées, en considérant comme modèle vrai un modèle exponentiel de portée $b = 0.1$, pour une grille 60×60 et $n = 100$. Les ZCAs ont été calculées pour cinq autres modèles (caractérisant une erreur dans l'estimation du paramètre de portée ou de la forme paramétrique).* . . 162

12.2 *Nombre, N, de simulations (sur 100) ayant des ZCAs significatives au niveau $1 - \hat{\alpha}$ et taille moyenne, $\bar{s}$, des ZCAs détectées, pour différentes discrétisation. La méthode utilise les paramètres d'une fonction de covariance exponentielle de portée $b = 0.1$ et des échantillons de $n = 100$ points.* 164

12.3 *Nombre, N, de simulations (sur 100) ayant des ZCAs et taille moyenne, $\bar{s}$, des ZCAs détectées pour deux interpolateurs du champ : krigeage simple et krigeage ordinaire (grille 60×60, $n = 100$, covariance exponentielle de portée $b = 0.1$).* 166

13.1 *Evolution culturale et nombre de points d'échantillonnage des variables dynamiques en fonction de la date de prélèvement.* 179

13.2 *Résumé statistique de la teneur en eau (en mm).* 179

13.3 *Résumé statistique de la teneur en azote minéral (en kg/ha).* 181

13.4 *Résumé statistique des variables permanentes.* 181

14.1 *Portée b (en m), variance (en mm^2) et niveau $1 - \alpha_G$ évalués lors de la procédure itérative pour la teneur en eau en mars 2000.* 185

14.2 *Niveaux de confiance associés au test local de détection obtenus aux différentes itérations pour la teneur en eau en mars 2000.* 185

14.3 *Portée b (en m), variance σ^2 (en mm^2) et niveau local de confiance $1-\alpha_G$ obtenus à la première itération et à la convergence pour la teneur en eau.* . 186

14.4 *Surfaces, S, (en m^2) et valeurs critiques, p, des ZCAs obtenues lors de la convergence pour la teneur en eau.* 188

14.5 *Portée b (en m) et variance σ^2 (en mm^2) des fonctions de covariance estimées aux différentes itérations, en fonction du modèle de covariance initial.* 190

14.6 *Portée, b (en m), variance, σ^2 (en mm^2) et niveau $1 - \alpha_G$ obtenus lors de la procédure itérative appliquée à la teneur en eau en juillet 2002, pour des échantillons obtenus en éliminant m % des points de l'échantillon initial.* . 193

Detecting Zones of Abrupt Change for spatial data and application to precision agriculture

---

Detecting Zones of Abrupt Change (ZACs), *i.e.* detecting ruptures in the plane, is a challenge in sparse and irregularly sampled spatial data.

We propose a two-steps method :

- After interpolating the variable by kriging, we estimate, *i.e.* detect, potential ZACs. We first define a test statistic, $T(\cdot)$, as a quadratic form of the interpolated gradient compared to its variance. Under a gaussian assumption for the variable of interest and under mild conditions about the regularity of its covariance function, $T(\cdot)$ has a non stationary χ^2 distribution. Potential ZACs are then defined as the excursion set of $T(\cdot)$ above the $(1-\alpha)$-quantile of the χ^2 distribution.
- In the second step, we test the statistical significance of the detected potential ZACs. We establish new results about the curvature at local maxima of the non stationary χ^2 field and give the asymptotic distribution of the size of a potential ZAC. The associated p-value is compared to a global level η. A significant potential ZAC define a ZAC.

The method requires two levels of significance : a fixed global level, η, and a local level, α, that we determine.

The power of the method being related to the local density of the samples, we assess the local power of the detection test, *i.e.* the probability to detect at a given point a discontinuity that passes through this point. Mapping the power allows to visualize the zones where the local sample density is not adapted to the ZACs detection.
We discuss the issues arised by the practical implementation of the method. The validation of the method on simulations and its application on soil data in a precision agriculture context brings to light a powerful method which is robust with respect to several parameters.

Keywords : Boundary estimation, Rupture detection, Excursion set, Gaussian random fields, Non stationary χ^2 fields, Geostatistics.

Ce travail s'inscrit dans le cadre de la détection de ruptures dans le plan, le problème étant de détecter des Zones de Changement Abrupt (ZCAs) dans des données faiblement et irrégulièrement échantillonnées dans un domaine de $\mathbb{R}^2$.

La méthode mise en œuvre se décompose en deux étapes :

- Après interpolation de la variable par krigeage, nous estimons, *i.e.* détectons, les ZCAs potentielles. Pour cela, nous définissons une statistique de test, $T(\cdot)$, comme étant une forme quadratique de l'interpolation du gradient du champ normalisé par sa variance. Sous l'hypothèse de gaussiannité du champ aléatoire et sous certaines conditions de régularité de sa fonction de covariance, $T(\cdot)$ a pour distribution une χ^2 non stationnaire. Les ZCAs potentielles sont alors définies comme l'ensemble d'excursion de $T(\cdot)$ au-dessus du $(1-\alpha)$-quantile d'une χ^2.
- Dans la deuxième étape, nous testons la significativité des ZCAs détectées. Nous établissons de nouveaux résultats sur la courbure d'un champ de χ^2 non stationnaire en des maxima locaux et la distribution asymptotique de la taille d'une ZCA potentielle. Cette distribution permet d'utiliser la taille d'une ZCA potentielle comme statistique de test. La valeur critique associée au test est alors comparée à un niveau global η. Une ZCA potentielle significative est appelée ZCA.

La méthode requiert deux niveaux de significativité : un niveau global, η, que nous fixons et un niveau local, α, que nous déterminons.

La puissance de la méthode dépendant de la densité locale de l'échantillon, nous calculons la puissance locale du test de détection de ZCAs, *i.e.* la probabilité de détecter en un point une courbe de discontinuité passant par ce point. La cartographie de cette puissance permet de visualiser les zones où la densité locale d'échantillonnage n'est pas adaptée à la détection de ZCAs.

Nous discutons des questions soulevées par l'implémentation de la méthode en pratique. Sa validation sur des simulations et son application à des données d'agriculture de précision révèlent une méthode puissante et robuste par rapport à de nombreux paramètres.

Mots-clés : Estimation de frontières, Détection de ruptures, Ensembles d'excursion, Champs aléatoires gaussiens, Champs de χ^2 non stationnaires, Géostatistique.

Discipline : Mathématiques appliquées

Institut National de la Recherche Agronomique,
Unité de Biométrie,
Domaine saint Paul - site Agroparc
84914 Avignon cedex 9, FRANCE.