ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n° 398 :
Géosciences, Ressources Naturelles et Environnement

Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité "Géostatistique"

présentée et soutenue publiquement par

Migraine Francky FOUEDJIO KAMENI

le 15 décembre 2014

Contributions à la modélisation et à l'inférence des fonctions aléatoires non-stationnaires de second ordre

Directeur de thèse : **Jacques RIVOIRARD**
Co-encadrant de thèse : **Nicolas DESASSIS**

Jury

Mme Liliane Bel,	Professeure, AgroParisTech	Présidente
Mme Christine Thomas,	Professeure, Université Toulouse 1 Capitole	Rapporteure
Mr Denis Allard,	Directeur de recherche, INRA Avignon	Rapporteur
Mr Jacques Rivoirard,	Directeur de recherche, MINES ParisTech	Examinateur
Mr Nicolas Desassis,	Chargé de recherche, MINES ParisTech	Examinateur
Mr Nicolas Jeannée,	Docteur, Geovariances	Examinateur

THÈSE

MINES ParisTech
Centre de Géosciences - Équipe Géostatistique
35, rue Saint-Honoré - 77305 Fontainebleau (France)

Remerciements

Je tiens à remercier tout spécialement Jacques Rivoirard et Nicolas Desassis pour avoir dirigé cette thèse, tout en me laissant une grande liberté quant aux choix d'orientation du travail. Leur disponibilité, leur enthousiasme et leur confiance en moi m'ont permis de réaliser ce travail dans des conditions tout à fait agréables.

Je souhaite également adresser ma reconnaissance aux membres de l'Équipe Géostatistique pour leur accueil chaleureux, pour m'avoir permis de réaliser ce travail, et accepté de me confier divers enseignements. Je remercie en particulier Thomas Romary et Christian Lantuéjoul pour leurs nombreuses suggestions, idées et conseils, Didier Renard et Fabien Ors pour leur précieuse aide en informatique. Je remercie également Jean-Paul Chilès, Hans Wackernagel, Christian Lajaunie, Serge Séguret, Chantal de Fouquet, Gaëlle Le loc'h et Hélène Beucher pour leur disponibilité et discussions enrichissantes. Un merci tout particulier à Isabelle Schmitt, Nathalie Dietrich et Dominique Vassiliadis pour leur incroyable gentillesse, ainsi qu'à mes collègues Jihane Belhadj et Clément Barthélemy pour leur agréable compagnie.

Je tiens à exprimer ma profonde gratitude aux membres du jury. Tout d'abord aux rapporteurs Christine Thomas Agnan et Denis Allard pour l'attention et l'intérêt qu'ils ont portés à ce travail et pour la pertinence de leurs remarques. Je suis également très reconnaissant à Liliane Bel et Nicolas Jeannée d'avoir bien voulu prendre de leur temps précieux pour faire partie du jury.

Je voudrais remercier l'entreprise CODELCO, ainsi que Alex McBratney pour m'avoir fourni les jeux de données permettant de développer la partie appliquée de ce travail.

Je voudrais remercier mes amis pour leur soutien accordé ces dernières années ; je pense particulièrement à Clorante Meuyou, Marcel Tchofach, Steve Gui-Diby, Steve Nana, Salomon Massoda, Mickael Kwekam et Delord Wandji.

Je remercie bien évidemment ma famille pour l'attention constante, le soutien sans faille et l'affection qu'elle m'a apportée jusqu'à ce jour. Puisse-t-elle trouver dans ce travail un motif supplémentaire de fierté.

Résumé

Les fonctions aléatoires stationnaires ont été utilisées avec succès dans les applications géostatistiques depuis plusieurs décennies. La structure de dépendance spatiale sous-jacente de la fonction aléatoire est alors représentée par un variogramme ou une covariance stationnaire. Cependant, dans certaines situations, il y a très peu de raisons de s'attendre à une structure de dépendance spatiale stationnaire sur l'ensemble du domaine d'intérêt. Dans cette thèse, deux approches de modélisation non-stationnaire de fonctions aléatoires sont considérées. Pour chacune d'elle, nous développons une méthodologie statistique d'estimation de la structure de dépendance spatiale non-stationnaire, dans le contexte d'une réalisation unique avec des données éventuellement espacées irrégulièrement. Par ailleurs, nous montrons également comment dans ce cadre non-stationnaire, les prédictions spatiales et les simulations conditionnelles peuvent être menées.

La première approche est celle par déformation d'espace consistant à transformer une fonction aléatoire non-stationnaire en stationnaire et isotrope via une déformation continue et bijective de l'espace d'indexation. Jusqu'à présent, cette approche n'a été fonctionnelle que dans le contexte de données avec réplication. Nous proposons une méthodologie d'inférence du modèle de déformation spatiale dans le cadre d'une réalisation unique. La méthode d'estimation est basée sur un estimateur non-paramétrique à noyau du variogramme non-stationnaire, servant de mesure de dissimilarité entre deux emplacements de la région d'intérêt. La procédure d'inférence proposée combine des aspects de lissage par noyau, positionnement multidimensionnel non-métrique pondéré et fonctions de base radiales de type spline plaque mince. La combinaison de ces outils permet de transformer la fonction aléatoire initialement non-stationnaire vers un nouvel espace où elle est stationnaire et isotrope. Les méthodes stationnaires pour la prédiction spatiale et la simulation s'appliquent alors dans cet nouvel espace. Les résultats prédits et simulés sont ensuite transposés vers l'espace d'origine, par simple correspondance.

La seconde approche est basée sur la convolution stochastique. Nous proposons un nouveau modèle pour les fonctions aléatoires non-stationnaires comme la convolution d'une mesure aléatoire orthogonale avec une fonction de pondération aléatoire variant

spatialement. Il s'agit là d'une généralisation de l'approche classique de convolution de fonctions aléatoires, basée sur une fonction de pondération jusqu'ici non-aléatoire. La classe de covariances non-stationnaires résultante est un mélange de modèles. Elle est très générale, flexible et permet pour des choix appropriés de la famille de fonctions de pondération aléatoires, de retrouver les expressions analytiques de classes de covariances non-stationnaires connues de la littérature. Ces dernières montrent localement un comportement stationnaire et leurs paramètres sont autorisés à varier en fonction de l'emplacement, donnant ainsi des variances, portées, anisotropies géométriques et régularités locales. Dans le contexte de réalisation unique et de quasi-stationnarité, nous développons une procédure d'estimation basée sur un estimateur non-paramétrique à noyau du variogramme local. Une approche par moindres carrés pondérés locaux en combinaison avec une méthode de lissage par noyau sont mises en œuvre pour estimer de manière efficace les paramètres qui régissent la structure de dépendance spatiale non-stationnaire.

Les deux modèles considérés fournissent une vue interprétable des fonctions aléatoires non-stationnaires. Les méthodes d'inférence développées permettent de capturer des structures de dépendance variables tout en garantissant la cohérence globale du modèle final. Elles sont faciles à mettre en œuvre et applicables à de grands ensembles de données. L'évaluation de leur performance selon plusieurs critères, sur des données synthétiques et réelles montre qu'elles donnent de meilleurs résultats de prédiction qu'une méthode stationnaire. Elles sont capables de prendre en compte les caractéristiques structurales locales de la régionalisation qui échappent aux méthodes stationnaires. Au delà de la prédiction, elles peuvent également servir comme outil pour une analyse exploratoire de la non-stationnarité.

Mots-clefs

Non-stationnarité, variogramme, covariance, déformation, convolution, noyau, krigeage, simulation.

Abstract

Stationary Random Functions have been sucessfully applied in geostatistical applications for decades. The underlying spatial dependence structure of the Random Function is represented by a stationary variogram or covariance. However, in some instances, there is little reason to expect the spatial dependence structure to be stationary over the whole region of interest. In this manuscript, two non-stationary modelling approaches for Random Functions are considered. For each of them, we develop a statistical methodology for estimating the non-stationary spatial dependence structure, in the context of a single realization with possibly irregularly spaced data. Moreover, we also show how spatial predictions and conditional simulations can be carried out in this non-stationary framework.

The first approach is the space deformation which consists in transforming a non-stationary Random Function to a stationary and isotropic one via a bijective continuous deformation of the index space. So far, this approach has been successfully applied in the context of replicated data. We propose an inference methodology of spatial deformation model, in the context of a single realization. The estimation method is based on a non-stationary variogram non-parametric kernel estimator which serves as a dissimilarity measure between two locations of the domain of interest. The proposed procedure combines aspects of kernel smoothing, weighted non-metric multi-dimensional scaling and thin-plate spline radial basis functions. These tools allow to transform the original non-stationary Random Function toward a new space where it is stationary and isotropic. Stationary techniques for spatial prediction and simulation can be applied in this new space. The predicted and simulated results are then mapped back into the original space, by simple correspondance.

The second approach is based on stochastic convolution. We propose a new model for non-stationary Random Functions as a convolution of an orthogonal random measure with a spatially varying random weighting function. This is a generalization of the common Random Functions convolution approach which use a non-random weighting function so far. The resulting class of non-stationary covariances is a mixture models. It is very general,

flexible and allows, to retrieve classes of closed-form non-stationary spatial covariance functions known from the literature, for a suitable choices of the random weighting functions family. These latter show locally a stationary behaviour and their parameters are allowed to vary with location, yielding local variances, ranges, geometric anisotropies and smoothnesses. Under a single realization and local stationarity framework, we develop an estimation procedure based on non-parametric kernel estimator of local variogram. A local weighted least-squares approach in combination with kernel smoothing method is implemented to estimate efficiently the parameters that govern the non-stationary spatial dependence structure.

Both considered models provide an interpretable view of non-stationary Random Functions. The proposed inference methods allow to capture varying spatial structures while guaranteeing the global consistency of the final model. They are easy to implement and feasible for large datasets. The assessment of their performance on both synthetic and real datasets show that they outperform stationary methods, according to several criteria. They take into account the local structural characteristics of the regionalization that the stationary approach is unable to retrieve. Beyond the prediction, they can also serve as a tool for exploratory analysis of the non-stationarity.

Keywords

Non-stationarity, variogram, covariance, deformation, convolution, kernel, kriging, simulation.

Table des matières

Introduction **13**

Notations **17**

1 Approches de modélisation non-stationnaire : état de l'art **19**
1.1 Introduction . . . 19
1.2 Points sources . . . 21
1.3 Déformation d'espace . . . 24
1.3.1 Modèle de base . . . 25
1.3.2 Développements . . . 28
1.3.3 Distance non-euclidienne . . . 31
1.4 Fonctions orthogonales empiriques . . . 32
1.5 Partitionnement et fenêtre glissante . . . 34
1.5.1 Partitionnement . . . 34
1.5.2 Fenêtre glissante . . . 35
1.6 Convolution . . . 37
1.6.1 Modèle continu . . . 37
1.6.2 Modèle discret . . . 43
1.6.3 Extensions . . . 45
1.7 Lissage par noyau . . . 49
1.7.1 Modèles locaux stationnaires pondérés . . . 49
1.7.2 Variogramme (covariance) empirique lissé(e) . . . 53
1.8 Équations aux dérivées partielles stochastiques . . . 54
1.9 Conclusion . . . 58

2 Modèle de déformation d'espace et inférence **61**
2.1 Introduction . . . 61
2.2 Description du modèle . . . 63
2.3 Inférence statistique . . . 65
2.3.1 Estimateur à noyau du variogramme . . . 65
2.3.2 Construction de l'espace déformé . . . 67
2.3.3 Estimation des paramètres fonctionnels . . . 73

2.3.4 Réglages des hyper-paramètres . . . 76
2.4 Prédiction . . . 77
2.4.1 Krigeage . . . 77
2.4.2 Simulation conditionnelle . . . 78
2.5 Illustration . . . 79
2.5.1 Exemple 1D . . . 79
2.5.2 Exemple 2D . . . 81
2.6 Conclusion . . . 88

3 Modèle de convolution généralisé et inférence 91
3.1 Introduction . . . 91
3.2 Définition du modèle . . . 94
3.3 Classes de covariance non-stationnaire . . . 95
3.4 Inférence statistique . . . 98
3.4.1 Modélisation . . . 98
3.4.2 Estimation des paramètres . . . 99
3.4.3 Choix des hyper-paramètres . . . 104
3.5 Prédiction . . . 105
3.5.1 Krigeage . . . 105
3.5.2 Simulation conditionnelle . . . 106
3.6 Illustration . . . 107
3.7 Conclusion . . . 113

4 Synthèse - trois exemples d'application 115
4.1 Introduction . . . 115
4.2 Application 1 : données de sol . . . 116
4.3 Application 2 : données de pluie . . . 127
4.4 Application 3 : données de topographie . . . 138
4.5 Conclusion . . . 150

Conclusions et perspectives 151

Liste des figures 154

Liste des tableaux 159

Bibliographie 161

A Rappels sur les fonctions aléatoires 171

B Scores de prédiction 177

C Démonstrations **181**

D Liste des publications et communications **187**

Introduction

"La géostatistique est l'application du formalisme des fonctions aléatoires à la reconnaissance et à l'estimation des phénomènes naturels." G. Matheron.

Le principal objectif de la géostatistique est la prédiction spatiale, encore appelée krigeage, consistant à prédire une variable régionalisée d'intérêt sur un domaine d'étude, à partir des données observées à certains emplacements. Le krigeage repose fondamentalement sur la modélisation et l'estimation de la structure de dépendance spatiale. La description de cette dernière se fait couramment à l'aide d'outils statistiques tels que le variogramme ou la covariance, calculés sur l'ensemble du domaine d'intérêt et sous une hypothèse de stationnarité [Chilès and Delfiner, 2012].

De la sorte, la complexité de la composante spatiale du phénomène analysé est donc limitée. L'hypothèse selon laquelle la structure de dépendance spatiale est invariante par translation sur l'ensemble du domaine d'intérêt peut être appropriée, lorsque ce dernier est de petite taille, lorsqu'il n'y a pas suffisamment de données pour justifier le recours à un modèle complexe ou simplement parce qu'il n'y a pas d'autres alternatives raisonnables. Cependant, il est de plus en plus admis que cette hypothèse est davantage motivée par des raisons de commodité mathématique que par la réalité [Guttorp and Schmidt, 2013]. La stationnarité n'est pas une généralité [Fuglstad, 2014] ; dans la pratique, il arrive souvent qu'elle soit douteuse, ceci en raison de nombreux facteurs, parmi lesquels les caractéristiques topographiques, les spécificités de la région d'intérêt ou d'autres influences locales. Ces influences locales peuvent être observées en calculant les variogrammes locaux dont les caractéristiques peuvent varier à travers le domaine d'étude. Dans un tel contexte, les modèles stationnaires sont incapables de prendre en compte certaines variations structurales comme une anisotropie locale variable. Par ailleurs, l'utilisation des modèles stationnaires serait susceptible de produire des prédictions moins précises, notamment une évaluation incorrecte de l'erreur d'estimation [Stein, 1999]. D'où la nécessité d'aller au-delà des modèles stationnaires. Dès lors se pose le problème du choix de la modélisation non-stationnaire à adopter et de son inférence.

Ce problème constitue le point de départ de notre réflexion : étant donnée une fonction aléatoire non-stationnaire en covariance (ou variogramme), n'y a t-il pas un paradigme statistique attrayant qui permette de la traiter au moyen de méthodes stationnaires ? Ce problème est bien connu dans la littérature géostatistique, avec les fonctions aléatoires intrinsèques d'ordre k (FAI-k) [Chilès and Delfiner, 2012]. Les FAI-k sont basées sur l'idée qu'une fonction aléatoire non-stationnaire en moyenne peut être réduite à une fonction aléatoire stationnaire par différenciation d'un ordre k.

Les deux stratégies de modélisation non-stationnaire considérées dans cette thèse s'inscrivent dans cette optique. Elles vont nous permettre d'analyser une fonction aléatoire non-stationnaire en covariance (ou variogramme) au moyen de techniques stationnaires plus accessibles à la fois d'un point de vue analytique et numérique. Ainsi, ces approches non-stationnaires restent connectées aux approches stationnaires, ce qui permet de conserver une certaine simplicité et interprétabilité.

S'agissant de l'inférence statistique de ces modèles non-stationnaires, nous nous plaçons dans le cadre fréquentiste univarié. Aucune hypothèse distributionnelle sur la loi spatiale n'est faite, en dehors de l'existence des moments d'ordre un et deux : les fonctions aléatoires ne sont pas nécessairement gaussiennes ou markoviennes. Nous proposons des procédures d'estimation ne nécessitant aucune inversion de matrice ou calcul de déterminant et applicables à de grands ensembles de données. Les modèles non-stationnaires considérés étant libres de distribution, l'estimation de leurs paramètres repose essentiellement sur la combinaison de méthodes de type moments et moindres carrés, relativement simples et rapides. Quant à la sélection des hyper-paramètres, elle repose sur des méthodes de type validation croisée. En outre, les prédictions spatiales et simulations conditionnelles peuvent être effectuées en utilisant tous les points de données (voisinage unique) si leur nombre est raisonnable.

La première approche étudiée est la déformation d'espace consistant à transformer une fonction aléatoire non-stationnaire en stationnaire et isotrope via une déformation continue et bijective de l'espace des coordonnées. Le variogramme non-stationnaire résultant s'exprime alors en fonction de la déformation et d'un variogramme stationnaire et isotrope. Jusqu'à présent, cette approche n'a été appliquée avec succès que dans le contexte de réalisations multiples. Nous proposons une méthodologie d'estimation de ce modèle dans le cadre d'une réalisation unique. La méthode proposée fournit une estimation non-paramétrique de la déformation en adoptant une approche par étapes. Tout d'abord, une matrice de dissimilarité est construite à partir d'un estimateur non-paramétrique à noyau du variogramme non-stationnaire. Ensuite, l'estimation de la déformation est déterminée en un ensemble réduit de points représentatifs (appelés points supports), par positionnement multidimensionnel non-métrique pondéré. Enfin, la déformation

est interpolée sur l'ensemble du domaine d'intérêt par des fonctions de base radiales de type spline plaque mince. L'approche fournit également une estimation rationnelle et automatique du variogramme stationnaire et isotrope dans l'espace déformé. Les méthodes stationnaires pour la prédiction spatiale et la simulation sont alors utilisées dans cet espace déformé. Les résultats prédits et simulés sont obtenus dans l'espace d'origine par simple correspondance. L'approche est illustrée sur des exemples simulés et appliquée à différents jeux de données réelles : sol, pluie et topographie.

La seconde approche considérée est basée sur l'idée de convolution de fonctions aléatoires. Nous présentons un nouveau modèle pour les fonctions aléatoires non-stationnaires comme la convolution d'une mesure aléatoire orthogonale avec une fonction de pondération stochastique variant spatialement. Il s'agit là d'une extension de l'approche classique de convolution de fonctions aléatoires, basée sur une fonction de pondération jusqu'ici déterministe. Par cette construction, la classe de covariances non-stationnaires obtenue est un mélange de modèles. Elle est très générale, flexible et permet pour certains choix de la famille de fonctions de pondération stochastiques, de retrouver les expressions analytiques de classes de covariances non-stationnaires existantes. Ces dernières montrent localement un comportement stationnaire et leurs paramètres sont autorisés à varier spatialement, donnant ainsi des variances, portées, anisotropies géométriques et régularités locales. De cette façon, cela permet de regrouper naturellement les paramètres de covariance locale stationnaire dans une covariance globale non-stationnaire valide, autorisant ainsi un krigeage en voisinage unique. L'estimation de ces paramètres se fait dans le contexte de réalisation unique et de stationnarité locale, par une approche étape par étape. D'abord, nous développons un estimateur non-paramétrique à noyau du variogramme local. Ensuite, il est utilisé dans une procédure de moindres carrés locaux pondérés pour estimer les paramètres en un ensemble réduit de points représentatifs dénommés points supports. Enfin, une technique de lissage par noyau est utilisée pour interpoler les estimations brutes des paramètres à tout emplacement d'intérêt. Les potentialités de l'approche proposée sont illustrées sur des données synthétiques et réelles (sol, pluie et topographie).

La modélisation et l'inférence des fonctions aléatoires d'ordre 2, à covariance ou variogramme non-stationnaire sont les principaux thèmes de cette thèse. On s'intéresse ici à la non-stationnarité d'un point de vue structural, c'est-à-dire celle liée à la structure de dépendance spatiale (non-stationnarité d'ordre 2). Ainsi, tout au long de ce manuscrit la non-stationnarité sera connotée généralement dans le sens de celle de la covariance ou du variogramme, de même que la stationnarité sera comprise comme celle d'ordre 2.

Le présent mémoire est constitué de quatre chapitres. Le premier chapitre est consacré à un tour d'horizon des différentes approches de modélisation non-stationnaire utilisées en géostatistique. Nous mettons en lumière les avantages de chaque approche de modélisation et relevons leurs limites. Nous nous attachons tout particulièrement aux problèmes d'inférence statistique des paramètres de chaque modèle, qui constitue un point crucial pour les applications. En effet, seule une inférence correcte peut permettre d'effectuer des prédictions optimales et d'obtenir des variances d'estimation ayant un sens. À la fin de ce premier chapitre, nous aurons pu mettre en évidence les différents enjeux et défis auxquels devront répondre toute nouvelle approche non-stationnaire. Dans les deuxième et troisième chapitres, nous proposons deux approches constructives de modélisation et d'estimation de la non-stationnarité. Chaque modèle est décrit à travers ses ingrédients de base et principales propriétés. Pour chacun des modèles, nous détaillons l'inférence statistique de ses paramètres. Nous montrons également comment la prédiction spatiale et la simulation conditionnelle basées sur de tels modèles peuvent être menées. Chaque approche est illustrée sur des exemples de données simulées. Le quatrième chapitre présente les résultats de l'application des approches proposées à trois jeux de données réelles, permettant ainsi d'apprécier leur efficacité et potentialité. Une comparaison en termes de performance de prédiction est faite avec l'approche stationnaire. Le mémoire se termine par une vue synoptique des méthodes proposées et identifie des développements ultérieurs à mener.

Notations

$Z(\mathbf{x}) \perp Z(\mathbf{y})$	$Z(\mathbf{x})$ indépendante de $Z(\mathbf{y})$
$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y}))$	Covariance de $Z(\mathbf{x})$ et $Z(\mathbf{y})$
$\mathbb{V}(Z(\mathbf{x}))$	Variance de $Z(\mathbf{x})$
$\mathbb{E}(Z(\mathbf{x}))$	Espérance de $Z(\mathbf{x})$
$\Gamma(.)$	Fonction Gamma
$\mathbb{1}$	Fonction indicatrice
$\mathcal{G}a(\alpha, \beta)$	Loi Gamma de paramètres α et β
$\mathcal{IG}(\alpha, \beta)$	Loi Inverse Gamma de paramètres α et β
$\mathcal{B}(k, \alpha)$	Loi Binomiale de paramètres k et α
$\mathcal{U}(G)$	Loi Uniforme sur G
$\mathcal{N}\left(m, \sigma^2\right)$	Loi Normale univariée de paramètres m et σ^2
$\mathcal{N}\left(\mathbf{m}, \mathbf{\Sigma}\right)$	Loi Normale multivariée de paramètres $\mathbf{m}$ et $\mathbf{\Sigma}$
$\|.\|$	Norme euclidienne dans $\mathbb{R}^p$
$\mathbf{h}^T$	Transposée du vecteur $\mathbf{h}$
$\mathbf{A}^T$	Transposée de la matrice $\mathbf{A}$
$\|\mathbf{A}\|$	Déterminant de la matrice $\mathbf{A}$
$\mathbf{1}$	Vecteur composé uniquement de 1
L^2	Ensemble des variables aléatoires de carré intégrable
$\mathcal{B}(\mathbb{R}^p)$	Ensemble des boréliens de $\mathbb{R}^p$
sup, min	Supremum, minimum
Nug(.)	Variogramme stationnaire pépitique, de palier 1
$\mathrm{Sph}_a(.)$	Variogramme stationnaire sphérique, de portée a, de palier 1
$\mathrm{Exp}_a(.)$	Variogramme stationnaire exponentiel, de portée pratique a, de palier 1
FAI-k	Fonctions Aléatoires Intrinsèques d'ordre k
NMDS	Positionnement Multidimensionel Non Métrique
EDPS	Équations aux Dérivées Partielles Stochastiques
ME	Moyenne des Erreurs
MAE	Moyenne des Erreurs Absolues
RMSE	Racine carrée de la Moyenne des Erreurs au Carré
NMSE	Moyenne des Carrés des Erreurs Standardisées
LogS	Score Logarithmique
CRPS	Indice Continu de Probabilité Ordonnée

Chapitre 1

Approches de modélisation non-stationnaire : état de l'art

"All is flux, nothing is stationary." Heraclitus.

Sommaire

1.1 Introduction **19**
1.2 Points sources **21**
1.3 Déformation d'espace **24**
1.4 Fonctions orthogonales empiriques **32**
1.5 Partitionnement et fenêtre glissante **34**
1.6 Convolution **37**
1.7 Lissage par noyau **49**
1.8 Équations aux dérivées partielles stochastiques **54**
1.9 Conclusion **58**

1.1 Introduction

Le problème classique en géostatistique est la prédiction d'une grandeur physique d'intérêt (variable régionalisée) sur un domaine d'étude à partir d'un ensemble fini d'observations éventuellement espacées irrégulièrement. Pour résoudre ce problème, les observations sont considérées comme étant issues de la réalisation d'une fonction aléatoire $Z = \{Z(\mathbf{x}), \mathbf{x} \in G \subseteq \mathbb{R}^p\}$ définie sur un domaine fixe et continu G de l'espace euclidien $\mathbb{R}^p$. La fonction aléatoire Z est très souvent modélisée sous une hypothèse de stationnarité de second ordre ou de stationnarité intrinsèque [Matheron, 1971]. Autrement dit, ses deux premiers moments peuvent s'écrire sous la forme : $\mathbb{E}(Z(\mathbf{x}) - Z(\mathbf{y})) = 0$, $\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = C(\mathbf{x} - \mathbf{y})$ ou $\frac{1}{2}\mathbb{V}(Z(\mathbf{x}) - Z(\mathbf{y})) = \gamma(\mathbf{x} - \mathbf{y})$. De la sorte, la moyenne est constante à travers le domaine d'intérêt et la structure de dépendance spatiale (covariance

ou variogramme) entre deux emplacements ne dépend que du vecteur de distance les séparant.

Dans certaines situations, il est approprié de modéliser la covariance ou le variogramme par une famille paramétrique stationnaire. Tandis que dans d'autres cas, les caractéristiques non-stationnaires sont apparentes et doivent être prises en considération afin d'obtenir une prédiction spatiale plus fiable. Par ailleurs, il a été remarqué que la stationnarité d'ordre 2 n'est pas toujours une hypothèse raisonnable même pour les fonctions aléatoires de moyenne constante et de variance homogène [Vanmarcke, 2010]. Un exemple remarquable est le processus de Wiener à 1D. La variance du processus augmentent linéairement avec la localisation. Et même après une standardisation du processus (moyenne nulle et variance unité), sa fonction de corrélation est non-stationnaire [Vanmarcke, 2010]. Le variogramme du processus quant à lui est stationnaire.

Dès lors que la structure de dépendance spatiale n'est plus stationnaire (varie spatialement), deux conséquences importantes surviennent. Premièrement, les modèles de variogramme ou covariance valides dans le cadre stationnaire ne peuvent plus s'appliquer directement. Deuxièmement, la fonction aléatoire ne s'auto-réplique plus dans les différentes parties du domaine d'intérêt. Or, c'est cette répétition dans l'espace qui fournit, d'une certaine manière, l'équivalent de plusieurs réalisations de la même fonction aléatoire, permettant ainsi une certaine inférence statistique.

L'idée de base des approches non-stationnaires est que les paramètres variographiques (effet de pépite, palier, portée, anisotropie, ...) peuvent varier spatialement. Plusieurs approches ont été proposées pour rendre compte de la non-stationnarité dans les données géostatistiques. La littérature à ce sujet est récente. Dans ce chapitre, nous la passons en revue sous les rubriques suivantes : points sources, déformation d'espace, fonctions orthogonales empiriques, partitionnement, fenêtre glissante, convolution, lissage par noyau et équations aux dérivées partielles stochastiques. Nous décrivons chaque approche de modélisation et mettons en exergue leurs avantages et limites. Une attention particulière est portée sur les hypothèses de travail et l'inférence statistique. Nous montrons également au moyen de simulations à quoi peuvent ressembler des réalisations issues de certains de ces modèles non-stationnaires. Ce chapitre se termine par une vue synoptique des différentes approches. Pour une revue bibliographique synthétique et non exhaustive des approches de modélisation non-stationnaire voir : Guttorp and Schmidt [2013], Sampson [2010], Sampson [2006] et Schabenberger and Gotway [2005a].

1.2 Points sources

Lorsqu'on connait les mécanismes qui contribuent à la non-stationnarité d'un processus spatial, ceux-ci peuvent être incorporés dans un modèle de covariance. Tel est le cas d'un processus spatial piloté par l'action d'un nombre limité de sites ou de points sources. L'intérêt peut se trouver non seulement dans la prédiction du processus spatial mais aussi dans l'évaluation de l'effet des points sources. Cette approche a été introduite par Hughes-Oliver et al. [1998a] qui proposent un modèle de corrélation pour un processus spatial conduit par un ou plusieurs points sources. Un point source étant défini comme une entité qui pilote le processus spatial directement ou indirectement. Cette définition attribue au moins un point source à un processus spatial. Le modèle à un point source de Hughes-Oliver et al. [1998a] s'écrit comme suit :

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = \exp\left(-\theta_1 \|\mathbf{x}-\mathbf{y}\| \exp\left(\theta_2 |c_{\mathbf{x}} - c_{\mathbf{y}}| + \theta_3 \min(c_{\mathbf{x}}, c_{\mathbf{y}})\right)\right), \tag{1.1}$$

où $c_{\mathbf{x}} = \|\mathbf{x}-\mathbf{c}\|$ et $c_{\mathbf{y}} = \|\mathbf{y}-\mathbf{c}\|$ sont les distances respectives entre les points $\mathbf{x}, \mathbf{y}$ et le point source $\mathbf{c}$.

Le modèle à un point source défini en (1.1) est circulaire, en ce sens qu'il n'y a pas de direction préférentielle de l'effet point source. Ce modèle est une adaptation du modèle de corrélation exponentiel stationnaire isotrope. En plus d'être une fonction de la distance entre deux emplacements, la fonction de corrélation est aussi une fonction de la distance du point source aux emplacements individuels, d'où la non-stationnarité. Ici, le modèle a deux paramètres additionnels comparativement à son homologue stationnaire. En posant ces paramètres égaux à zéro ($\theta_2 = \theta_3 = 0$), on obtient le modèle exponentiel stationnaire isotrope de portée pratique [1] $\alpha = 3/\theta_1$. Le modèle défini en (1.1) a une portée pratique :

$$\alpha(\mathbf{x}, \mathbf{y}) = 3\frac{\exp\left(-\theta_2 |c_{\mathbf{x}} - c_{\mathbf{y}}| - \theta_3 \min(c_{\mathbf{x}}, c_{\mathbf{y}})\right)}{\theta_1}. \tag{1.2}$$

Pour deux points $\mathbf{x}$ et $\mathbf{y}$ à égale distance du point source $\mathbf{c}$ ($c_{\mathbf{x}} = c_{\mathbf{y}}$), l'expression (1.2) se réduit à :

$$\alpha(\mathbf{x}, \mathbf{y}) = 3\frac{\exp\left(-\theta_3 \|\mathbf{x}-\mathbf{c}\|\right)}{\theta_1}. \tag{1.3}$$

Dans le modèle défini en (1.1), deux emplacements éloignés mais équidistants du point source peuvent être très corrélés [2]. Lorsque θ_1 est grand, les emplacements qui se trouvent sur la même bande radiale de la source ponctuelle sont fortement corrélés. Pour θ_3 grand, la corrélation décroît rapidement lorsqu'on s'éloigne de la source ponctuelle.

1. On appelle portée pratique d'une covariance $C(\mathbf{h})$ la plus petite distance α qui est telle que $C(\mathbf{h}) \leq 0,05C(\mathbf{0})$ lorsque $\|\mathbf{h}\| \geq \alpha$.
2. Les modèles stationnaires isotropes ne peuvent pas représenter une telle structure.

Si par exemple, la variable régionalisée d'intérêt est la quantité de monoxyde de carbone dans l'air, alors une cheminée produisant des fumées (de monoxyde de carbone) pourrait être considérée comme une source ponctuelle. Il est ainsi raisonnable de croire que la corrélation entre deux points à une distance donnée dépendra de la proximité à la cheminée.

Bien que le modèle défini en (1.1) soit intéressant, sa validité reste à vérifier. En effet, comme toute covariance, elle doit être définie positive. Autrement dit, toute matrice de covariance induite par celle-ci doit être semi-définie positive. Cette condition implique des contraintes sur les trois paramètres du modèle. Selon Hughes-Oliver et al. [1998a], ces contraintes dépendent de la dimension du domaine d'intérêt et de la métrique utilisée. Il n'est pas aisé de déterminer les conditions nécessaires et suffisantes sur les paramètres du modèle assurant une définie positivité de la structure de covariance non-stationnaire. Toutefois, des conditions nécessaires peuvent être énoncées : tous les paramètres doivent être positifs ($\theta_1 > 0, \theta_2 \geq 0, \theta_3 \geq 0$). Une astuce trouvée par Hughes-Oliver et al. [1998a] est de dériver une condition suffisante sur le domaine d'intérêt et sur la distance utilisée qui assure la condition de définie positivité du modèle estimé (aux points d'observation), même si le modèle théorique ne l'est pas forcément. En effet, cette condition assure que la matrice de covariance aux points d'observation est une matrice à diagonale dominante [3] et donc semi-définie positive.

L'inférence statistique des paramètres du modèle est faite par maximum de vraisemblance restreint. Le modèle défini en (1.1) est appliqué par Hughes-Oliver et al. [1998b] pour la modélisation de dépôts chimiques en cours de fabrication de semi-conducteurs. D'autres applications potentielles de cette approche de modélisation sont par exemple : la contamination aux métaux lourds causés par les sites de tir ou la pollution de l'eau causée par les eaux usées des usines.

Hughes-Oliver et al. [1998a] montrent qu'on peut généraliser le modèle paramétrique non-stationnaire à un seul point source de deux façons. Tout d'abord, le processus peut être affecté par de multiples points sources ; ensuite, l'effet d'un (des) point(s) source(s) peut être non-circulaire. En somme, le modèle est généralisable à de multiples points sources qui présentent une anisotropie de point source, qui est une structure non-circulaire autour de la source :

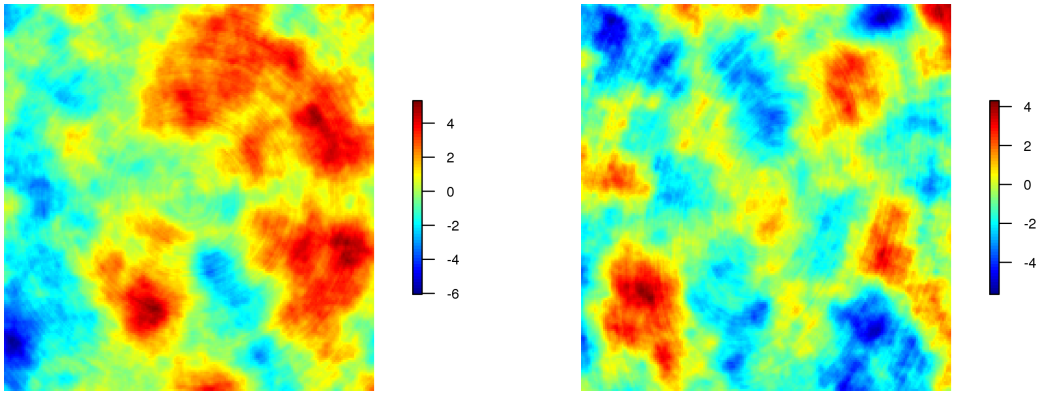
$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = \exp\left(-\theta_1 h_{\mathbf{xy}} \sum_{k=1}^{M} \exp\left(\theta_2^k |c_{\mathbf{x}}^k - c_{\mathbf{y}}^k| + \theta_3^k \min(c_{\mathbf{x}}^k, c_{\mathbf{y}}^k)\right)\right), \tag{1.4}$$

où $h_{\mathbf{xy}} = \|\mathbf{A}^0(\mathbf{x} - \mathbf{y})\|$ est la distance entre les points $\mathbf{x}$ et $\mathbf{y}$ après transformation par la

3. En algèbre linéaire, une matrice carrée à coefficients réels ou complexes est dite à diagonale dominante lorsque le module de chaque terme diagonal est supérieur ou égal à la somme des modules des autres termes de sa ligne.

matrice d'anisotropie $\mathbf{A}^0$; $c_{\mathbf{x}}^k = \|\mathbf{A}^k(\mathbf{x} - \mathbf{c}^k)\|$ est la distance entre le point $\mathbf{x}$ et le point source $\mathbf{c}^k$ après transformation par la matrice d'anisotropie $\mathbf{A}^k$.

Les figures 1.1a et 1.1b présentent deux réalisations de fonctions aléatoires gaussiennes issues d'un modèle paramétrique à un seul point source [Hughes-Oliver and Gonzàlez-Farìas, 1999]. La figure 1.1a illustre celle d'un modèle circulaire et la figure 1.1b celle d'un modèle non-circulaire. Le point source étant localisé dans les deux cas au centre du domaine.



(a) Circulaire (b) Non-circulaire

FIGURE 1.1 – Modèle à point source : réalisations de fonctions aléatoires non-stationnaires.

Hughes-Oliver and Gonzàlez-Farìas [1999] étendent le modèle défini en (1.1) en construisant une classe générale de modèles paramétriques non-stationnaires (à un seul point source) valides, où la mise à zéro de certains paramètres rend la classe stationnaire. Ils discutent des propriétés d'une famille particulière de cette catégorie générale (famille exponentielle). Pour cela, ils écrivent un processus spatial non-stationnaire comme le produit ou la somme de deux processus spatiaux indépendants de moyenne nulle dont l'un est stationnaire et l'autre non-stationnaire. La covariance résultante est la somme ou le produit des covariances de ces processus spatiaux et est donc définie positive. La covariance de la composante stationnaire est définie comme appartenant à la famille exponentielle et celle de la composante non-stationnaire est obtenue à partir du processus non-stationnaire de Wiener. Les auteurs optent pour une décomposition multiplicative. La procédure d'inférence s'effectue toujours par maximum de vraisemblance. Il est possible d'effectuer des tests pour évaluer la significativité de l'ensemble des paramètres, de même que l'impact du point source. Cette dernière peut être évaluée en prenant comme hypothèse nulle une covariance stationnaire et en utilisant le test du rapport de vraisemblance. Ainsi, l'intérêt des modèles paramétriques non-stationnaires à points sources ne réside pas seulement dans la prédiction mais aussi dans le fait qu'ils permettent de tester l'impact d'un point source sur la moyenne, la variance ou la corrélation d'un processus spatial ; ce qui permet *in fine* de comprendre la non-stationnarité. Hughes-Oliver and Gonzàlez-Farìas

[1999] appliquent leur modèle à un ensemble de mesures de potentiels électriques prises dans un champ contenant un poteau en métal.

Hughes-Oliver et al. [2009] étendent le modèle de décomposition d'un processus spatial suggéré par Hughes-Oliver and Gonzàlez-Farìas [1999] et proposent une approche hiérarchique bayésienne. Ecker and Oliveira [2008] présentent un modèle similaire à celui étudié par Hughes-Oliver and Gonzàlez-Farìas [1999], mais ils combinent les deux composantes du processus spatial de façon additive plutôt que multiplicative. Une approche bayésienne est utilisée pour tirer des conclusions sur les paramètres du modèle. Les covariances des deux processus spatiaux sous-jacents (stationnaire et non-stationnaire) sont définies dans la famille exponentielle. Le modèle est appliqué à une situation où il y a une externalité localisée dans une région qui influence le prix de vente des maisons (proximité de l'autoroute principale, d'une centrale nucléaire, d'un aéroport,...). Ecker et al. [2013] améliorent le modèle de Ecker and Oliveira [2008]. Leur modèle n'a qu'un seul paramètre de plus que le modèle à structure de covariance stationnaire et il s'adapte mieux que le modèle à trois paramètres supplémentaires de Ecker and Oliveira [2008]. Le modèle est estimé dans le contexte bayésien et la méthode est illustrée sur le jeu de données utilisé par Ecker and Oliveira [2008].

Dans le cas où un processus spatial ne dispose que d'un nombre réduit de points sources, ces derniers peuvent être identifiés et intégrés dans le modèle, ajoutant ainsi des renseignements précieux qui ne peuvent qu'améliorer le modèle. Si le processus spatial a un grand nombre de points sources, les intégrer tous dans la modélisation peut conduire à un modèle inutilement complexe et peu pratique. Dans les modèles à points sources, la position des points sources peut être supposée inconnue, mais cela conduit en général à des problèmes d'identification. La particularité des modèles paramétriques non-stationnaires à points sources est qu'ils peuvent être à la fois simples et informatifs dans certains cas de figures. Cependant, ils ne peuvent être utilisés que lorsque les sources de non-stationnarités ont été appréhendées.

1.3 Déformation d'espace

Généralement, si une fonction aléatoire ne possède pas les propriétés nécessaires à des fins d'inférence statistique, il est d'usage d'employer une transformation qui conduit aux propriétés souhaitées. L'anamorphose gaussienne et l'anisotropie géométrique sont deux exemples familiers de transformations en géostatistique. Une différence importante entre ces deux transformations est que la première transforme la variable régionalisée et la seconde l'espace géographique.

L'approche par déformation d'espace introduite par Sampson and Guttorp [1992] est la méthode la plus largement étudiée pour les fonctions aléatoires non-stationnaires. L'idée sous-jacente de la méthode est celle de la transformation du domaine d'observations d'une fonction aléatoire non-stationnaire dans un espace de dimension donnée, où elle peut être plus simplement décrite et modélisée ; c'est-à-dire où la stationnarité et l'isotropie seraient plausibles. À la différence de l'anisotropie géométrique où une transformation linéaire est effectuée, il s'agit ici d'une transformation non-linéaire.

1.3.1 Modèle de base

Soient $Z_{it} = Z(\mathbf{s}_i, t)$ les variables issues de la fonction aléatoire $Z = \{Z(\mathbf{x}, t) : \mathbf{x} \in G \subseteq \mathbb{R}^2, t \in T \subseteq \mathbb{R}\}$ aux n emplacements $\{\mathbf{s}_i, i = 1, \ldots, n\}$ et aux mêmes m points $\{t = 1, \ldots, m\}$. Sampson and Guttorp [1992] modélisent la fonction aléatoire Z comme suit :

$$Z(\mathbf{x}, t) = \mu(\mathbf{x}) + E(\mathbf{x}, t) + \epsilon(\mathbf{x}, t),\ \forall(\mathbf{x}, t) \in G \times T, \tag{1.5}$$

où $\mu(\mathbf{x})$ est la fonction moyenne ; $E(\mathbf{x}, t)$ est une fonction aléatoire d'espérance nulle, supposée temporellement stationnaire et continue en moyenne quadratique spatialement ; $\epsilon(\mathbf{x}, t)$ est une fonction aléatoire d'espérance nulle, de variance unité, indépendante à travers l'espace et le temps, et indépendante de E.

Sous la représentation définie en (1.5), Sampson and Guttorp [1992] considèrent le modèle spatial de second ordre suivant :

$$\gamma(\mathbf{x}, \mathbf{y}) \equiv \frac{1}{2}\mathbb{V}(Z(\mathbf{x}, t) - Z(\mathbf{y}, t)) = \gamma_0(\|f(\mathbf{x}) - f(\mathbf{y})\|),\ \forall(\mathbf{x}, \mathbf{y}) \in G \times G, \tag{1.6}$$

où $f : G \subseteq \mathbb{R}^2 \to D \subseteq \mathbb{R}^2$ est une fonction déterministe, bijective et continue de l'espace G vers l'espace D ; $\|.\|$ représente la norme euclidienne dans $\mathbb{R}^2$ et $\gamma_0(.)$ est un variogramme stationnaire et isotrope.

Ainsi, la structure de dépendance spatiale non-stationnaire (1.6) est exprimée comme une fonction de la distance euclidienne entre deux localisations, après une déformation bijective et continue du domaine d'étude.

Dans le modèle spatial de second ordre (1.6), les fonctions $\gamma_0(.)$ et $f(.)$ sont inconnues et donc à estimer. Plusieurs questions se posent : approche paramétrique ou non-paramétrique ? Quelle classe de transformation utiliser pour $f(.)$? Comment obtenir "le meilleur" membre de cette classe ? Sampson and Guttorp [1992] optent pour une approche non-paramétrique. Ils choisissent la déformation $f(.)$ dans la classe des fonctions de base radiales de type spline plaque mince.

L'approche utilisée par Sampson and Guttorp [1992] détermine $f(.)$ et $\gamma_0(.)$ en deux étapes. Premièrement, les dispersions spatiales $d_{ij}^2 = \frac{1}{2}\widehat{\mathbb{V}}(Z_{it} - Z_{jt}) = \frac{1}{2m}\sum_{t=1}^{m}(Z_{it} - Z_{jt})^2$ sont calculées et l'algorithme de positionnement multidimensionnel non-métrique [4] de Shepard-Kruskal [Mardia et al., 1979] est appliqué à la matrice $\mathbf{D} = [d_{ij}]$; ceci permet de représenter les points de données $\{\mathbf{s}_i\}$ de l'espace d'origine G par des points $\{\mathbf{s}_i^\star\}$ dans l'espace déformé D. À partir d'une configuration initiale de points, le résultat de cet algorithme est un ensemble de points $\{\mathbf{s}_i^\star\}$ et une fonction monotone $\phi(.)$ telle que : $\phi(d_{ij}) \approx \|\mathbf{s}_i^\star - \mathbf{s}_j^\star\|$. Une fois l'algorithme de positionnement multidimensionnel appliqué à la matrice $\mathbf{D}$, un modèle $\gamma_0(.)$ est ajusté sur le nuage de points $(\|\mathbf{s}_i^\star - \mathbf{s}_j^\star\|, d_{ij}^2)$ dans l'espace déformé D, comme dans le cadre stationnaire classique.

Pour effectuer la prédiction spatiale aux emplacements non-observés, l'estimation d'un modèle spatial de second ordre entre des emplacements arbitraires du domaine d'origine est nécessaire. À cet effet, une représentation des coordonnées de tout point arbitraire de l'espace G vers l'espace D est requise. D'où la deuxième étape de la procédure. Elle consiste à estimer la déformation $f(.)$ en interpolant l'ensemble des points d'observation $\{\mathbf{s}_i\}$ et leurs images $\{\mathbf{s}_i^\star\}$ obtenus à l'étape précédente. La classe des splines plaques minces est utilisée par Sampson and Guttorp [1992] pour approcher la déformation $f(.)$. Spécifiquement, la fonction bivariée $f(.)$ a la forme suivante :

$$f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}))^T = \mathbf{c} + \mathbf{A}\mathbf{x} + \mathbf{V}^T\boldsymbol{\sigma}(\mathbf{x}),\ \forall \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2, \tag{1.7}$$

avec les matrices $\mathbf{c}(2\times 1), \mathbf{A}(2\times 2), \mathbf{V}(n\times 2)$ et $\boldsymbol{\sigma}(\boldsymbol{x}) = (\sigma(\mathbf{x}-\mathbf{s}_1), \ldots, \sigma(\mathbf{x}-\mathbf{s}_n))^T$, où $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ sont les points de données vus comme les centres de la fonction de base radiale $\sigma(\mathbf{h}) = \|\mathbf{h}\|^2 \log(\|\mathbf{h}\|)\mathbb{1}_{\|\mathbf{h}\|>0}$.

Dans le modèle (1.7), les paramètres $\{\mathbf{c}, \mathbf{A}, \mathbf{V}\}$ sont inconnus et donc à estimer. Un paramètre de lissage est incorporé pour permettre de choisir le degré de régularité souhaité pour $f(.)$. Pour une valeur donnée du paramètre de lissage λ, les paramètres $\{\mathbf{c}, \mathbf{A}, \mathbf{V}\}$ sont choisis en minimisant la quantité suivante :

$$\sum_{i=1}^{n}(\mathbf{s}_i^\star - f(\mathbf{s}_i))^T(\mathbf{s}_i^\star - f(\mathbf{s}_i)) + \lambda J(f), \tag{1.8}$$

où $\lambda > 0$ et $J(f) = \sum_{j=1}^{2}\int_{\mathbb{R}^2}\left(\frac{\partial^2 f_j}{\partial x_1^2}\right)^2 + \left(\frac{\partial^2 f_j}{\partial x_1 x_2}\right)^2 + \left(\frac{\partial^2 f_j}{\partial x_2^2}\right)^2 dx_1 dx_2$.

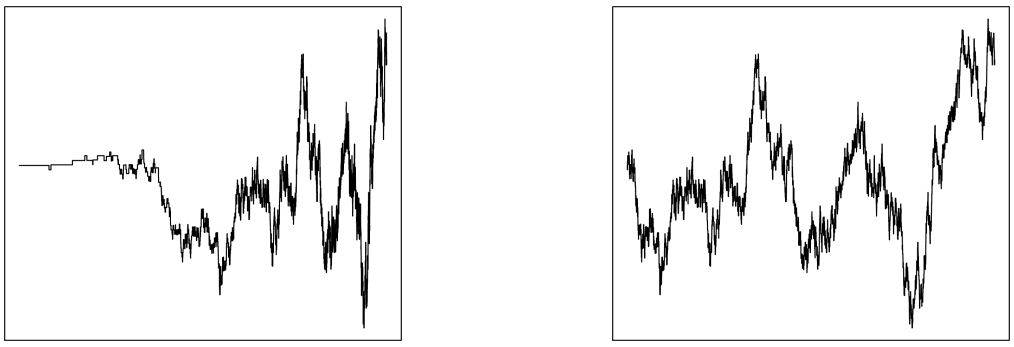
Pour $\lambda = 0$, le résultat obtenu est une spline d'interpolation, tandis que $\lambda \to +\infty$ conduit $\mathbf{V}$ vers une matrice nulle, de telle sorte que $f(.)$ soit juste une transformation

4. Le positionnement multidimensionnel est un outil statistique qui utilise une mesure de proximité ou dissimilarité entre des objets pour produire une représentation spatiale de ces objets. Il est présenté en détail au chapitre 2.

linéaire. Sampson and Guttorp [1992] utilisent une spline de lissage plutôt qu'une spline d'interpolation en raison de la possibilité de repliement de la déformation $f(.)$ qui elle n'est pas garantie bijective. Le paramètre de lissage dans la procédure d'ajustement est réglé pour s'assurer que la déformation $f(.)$ ne se replie pas. Le paramètre λ est choisi a priori ou déterminé au moyen d'une procédure de validation croisée.

Comme décrit par Perrin and Meiring [1999], lorsque $\gamma_0(.)$ est non-décroissante, la déformation spatiale $f(.)$ opère concrètement comme suit : dans les régions de corrélation spatiale relativement faible, la transformation étire l'espace géographique G ; tandis que dans les régions de corrélation spatiale relativement élevée, elle le contracte ; de sorte qu'une structure de dépendance spatiale stationnaire et isotrope puisse être modélisée dans l'espace déformé D.

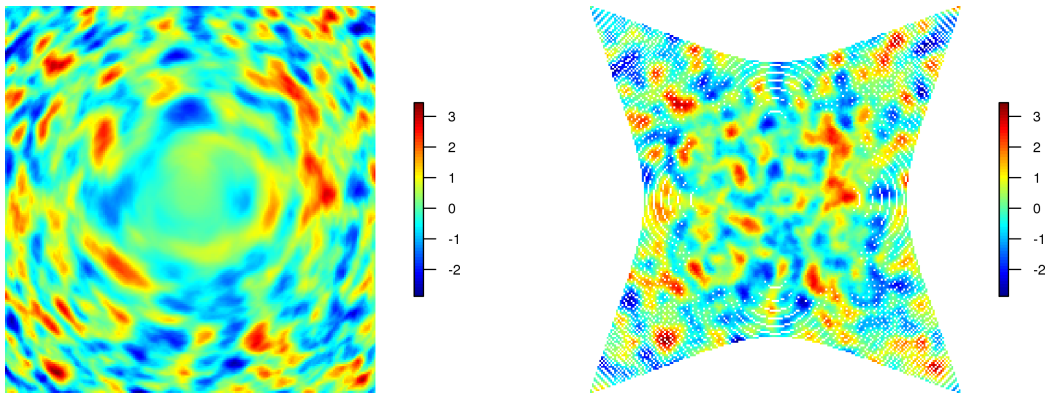
Les figures 1.2a et 1.2b montrent une réalisation d'une fonction aléatoire gaussienne en 1D par déformation spatiale dans les espaces géographique et déformé. La fonction aléatoire non-stationnaire est transformée en une fonction aléatoire stationnaire et isotrope, par simple contraction et étirement de l'axe des abscisses. En effet, dans l'espace géographique, les points du segment qui sont situés près de l'origine sont fortement corrélés avec leurs voisins et ceux à proximité de l'extrémité droite sont légèrement corrélés avec leurs voisins.



(a) Non-stationnaire (b) Stationnaire

FIGURE 1.2 – Modèle de déformation d'espace (1D) : réalisation d'une fonction aléatoire non-stationnaire.

Les figures 1.3a et 1.3b illustrent l'approche par déformation d'espace en 2D. Les points du domaine géographique qui sont situés près du centre sont fortement corrélés avec leurs voisins, tandis que ceux qui sont éloignés du centre le sont moins. Ainsi partant d'une fonction aléatoire non-stationnaire, la déformation étire le domaine dans les régions de corrélation spatiale relativement faible (aux extrémités), tandis qu'elle contracte le domaine dans les régions de corrélation spatiale relativement élevée (au centre).



(a) Non-stationnaire (b) Stationnaire

FIGURE 1.3 – Modèle de déformation d'espace (2D) : réalisation d'une fonction aléatoire non-stationnaire.

Sampson and Guttorp [1992] appliquent leur approche sur des données de radiation solaire du sud-ouest de la Colombie-Britannique, au Canada. Leur approche soulève quelques remarques. Un variogramme empirique est nécessaire comme entrée dans l'algorithme de positionnement multidimensionnel non-métrique. Comment obtenir les estimations d_{ij}^2 sans hypothèse de stationnarité ou de réplication ? Les mesures répétées à travers le temps en chaque emplacement et l'hypothèse de stationnarité temporelle permettent ici l'estimation des dispersions spatiales. Par ailleurs, au delà de la portée finie, les valeurs observées aux points de données ne sont pas corrélées. Ainsi, les d_{ij} peuvent être quasi-constants et donc la représentation par positionnement multidimensionnel non-métrique peut ne pas reproduire fidèlement les d_{ij}. La solution obtenue par l'algorithme de positionnement multidimensionnel non-métrique dépend de la configuration initiale des points et est obtenue à une similitude près (translation, rotation, symétrie, homothétie). Les sites d'observation sont un choix logique pour l'initialisation de l'algorithme.

1.3.2 Développements

L'utilisation de la famille non-paramétrique de splines plaques minces pour ajuster la déformation a deux inconvénients : la propriété de bijectivité n'est pas assurée et l'ajustement du modèle peut être un problème numérique difficile. Pour pallier à ces inconvénients, Perrin and Meiring [1999] proposent une approche paramétrique. Ils modélisent la déformation en utilisant une composition d'un petit nombre de déformations radiales élémentaires et bijectives. Puis, ils estiment les paramètres par moindres carrés. Cependant, un choix rationnel du nombre pertinent de déformations élémentaires n'a pas été mise en œuvre. Par ailleurs, le fait de savoir si toutes les fonctions bijectives peuvent être approchées de cette manière reste une question ouverte. L'approche est appliquée à des données de pluie de la région Languedoc-Roussillon, au sud de la France.

Iovleff and Perrin [2004] implémentent un algorithme de recuit simulé pour l'ajustement du modèle de déformation spatiale. Ils proposent d'estimer à la fois la déformation et la corrélation stationnaire isotrope en utilisant une version continue du recuit simulé avec contraintes. Cette méthode fournit une estimation non-paramétrique de la déformation qui a la propriété requise d'être bijective. Pour garantir la bijectivité de la déformation estimée, plutôt que d'imposer une contrainte de régularité analytique sur la fonction de déformation (comme avec les splines plaques minces), Iovleff and Perrin [2004] imposent des contraintes sur les perturbations aléatoires en utilisant une triangulation de Delaunay. L'approche est illustrée sur des données de pluie.

Perrin and Meiring [1999] établissent des propriétés théoriques du modèle de déformation. Ils prouvent que si la fonction $f(.)$ est différentiable ainsi que son inverse et si la fonction $\gamma_0(.)$ est aussi différentiable, alors le modèle de déformation spatiale est identifiable à une homothétie près pour $\gamma_0(.)$ et à une similitude (translation, rotation, symétrie, homothétie) près pour $f(.)$. Perrin and Senoussi [2000] donnent une caractérisation des fonctions de corrélation non-stationnaire réductibles en stationnaire ou en stationnaire isotrope, sous certaines conditions de régularité. Porcu et al. [2010] traitent ce problème de réductibilité d'une fonction aléatoire non-stationnaire dans un cadre plus général. Ainsi, il est important de noter que la covariance ou le variogramme d'une fonction aléatoire non-stationnaire n'est pas toujours réductible à une covariance ou un variogramme stationnaire via une déformation de l'espace.

Les méthodes d'inférence par maximum de vraisemblance du modèle de déformation ont été développées par Mardia and Goodall [1993] et Smith [1996]. Damian et al. [2001] et Schmidt and O'Hagan [2003] formulent indépendamment des approches similaires d'inférence bayésienne du modèle de déformation spatiale ; ce qui permet la prise en compte de l'incertitude dans l'estimation de la déformation qui sous-tend la structure de corrélation spatiale. Une fois de plus, le caractère bijectif n'est pas assuré. Damian et al. [2001] adoptent une approche bayésienne pour estimer la fonction de déformation par splines plaques minces. Schmidt and O'Hagan [2003] choisissent de ne pas modéliser directement la fonction de déformation mais plutôt de modéliser les localisations transformées. Les emplacements transformés sont modélisés comme des réalisations d'une fonction aléatoire bivariée gaussienne, avec des lois a priori sur les paramètres. Ainsi, la fonction de déformation $f(\mathbf{x})$ est vue comme une réalisation d'une fonction aléatoire bivariée en $\mathbf{x}$ au lieu d'une valeur en $\mathbf{x}$ d'une transformation bivariée aléatoire. L'approche est également illustrée avec les données de pluie de la région Languedoc-Roussillon, au sud de la France.

Dans la majorité des applications, c'est la représentation en 2D qui est étudiée dans les modèles de déformation d'espace, bien qu'en principe on puisse opter pour une représentation en dimension supérieure. Une représentation exacte de n points par l'algorithme de positionnement multidimensionnel est atteinte au plus en dimension $n-1$. Lorsque le nombre de points d'observation augmente, il est généralement plus difficile d'obtenir une configuration qui décrit les données avec précision en dimension 2 [Vera et al., 2008]. D'après Barber [2002], il n'est pas évident qu'une telle réorganisation des localisations en 2D soit suffisante pour tenir compte du degré de régularité du processus. Une transformation en dimension supérieure peut être appropriée aux dépens de la lisibilité du modèle. Vera et al. [2008, 2009] proposent une modification consistant à partitionner les données spatiales en classes (avec inclusion des contraintes géographiques spatiales) et de représenter les centres de classe dans un espace de faible dimension, tandis que les localisations et les classes conservent leurs relations spatiales. Cette méthode est appliquée sur des données de vitesse du vent dans l'ouest du Pacifique tropical.

Schmidt et al. [2011] proposent une approche permettant d'intégrer des covariables dans le modèle de déformation spatiale pour mieux estimer la déformation. L'estimation du modèle est réalisée sous le paradigme bayésien. Schmidt et al. [2011] appliquent leur approche sur des données de radiation solaire du sud-ouest de la Colombie-Britannique, au Canada [Sampson and Guttorp, 1992] et aux données de température dans le Colorado, au États-Unis d'Amérique [Paciorek and Schervish, 2006].

Bornn et al. [2012] présentent une méthode qui consiste à plonger le domaine d'intérêt dans un espace de dimension supérieure afin de récupérer la stationnarité du processus spatial. Ils combinent les aspects du positionnement multidimensionnel, du groupe lasso, et des modèles à variables latentes, pour trouver l'espace dans lequel la fonction aléatoire initialement non-stationnaire présente une stationnarité. Le point de départ de cette approche est que les fonctions aléatoires non-stationnaires peuvent être représentées comme des projections en faible dimension de fonctions aléatoires stationnaires en grande dimension [Perrin and Meiring, 2003]. Cette approche est appliquée d'une part sur le jeu de données de radiation solaire utilisé par Sampson and Guttorp [1992], et d'autre part sur des données de concentration de fumée noire au Royaume-Uni.

Jusqu'à présent, et comme l'ont souligné Meiring et al. [1997], Perrin and Monestiez [1998], Iovleff and Perrin [2004], aucun choix rationnel n'a été encore développé pour le modèle de variogramme stationnaire et isotrope dans les modèles de déformation. Ce choix est souvent fait par une inspection visuelle.

À notre connaissance, Anderes and Stein [2008], Anderes and Chatterjee [2009] sont les premiers et seuls auteurs jusqu'à présent à avoir traité de l'estimation du modèle de déformation spatiale basé sur une réalisation unique. Ils présentent une méthodologie

d'estimation de la déformation basée sur les transformations conformes dans le contexte gaussien. Un calcul approximatif de la vraisemblance est dérivé du partitionnement des observations denses en voisinages et en supposant l'indépendance de la fonction aléatoire sur les partitions. Cependant, cette approche n'a pas été appliquée sur des exemples de données réelles et nécessite des données denses.

1.3.3 Distance non-euclidienne

La non-stationnarité de la structure de dépendance spatiale peut également survenir parce que la distance euclidienne n'est pas une mesure appropriée de la séparation spatiale [Diggle and Ribeiro, 2007]. La structure de dépendance spatiale peut par exemple suivre des formes géologiques qui conduisent à des corrélations non-stationnaires. Cependant, l'utilisation d'une distance autre que la distance euclidienne peut rendre invalides les modèles de second ordre [Schabenberger and Gotway, 2005b]. Pour corriger cette situation, le positionnement multidimensionnel métrique est utilisé pour incorporer les emplacements du domaine d'intérêt dans un espace euclidien où la distance euclidienne entre les points approche celle non-euclidienne de l'espace des coordonnées d'origine ; la définie positivité est ainsi garantie.

Almendral et al. [2008] définissent une nouvelle mesure de distance entre les points dans l'espace d'origine. Cette distance est telle qu'elle honore le domaine d'intérêt en termes d'anisotropies locales variables. Dans ce contexte, une nouvelle distance géodésique est introduite par l'application d'une méthode de recherche sur graphe de voisinages pondérés. Cette distance est utilisée comme entrée dans l'algorithme de positionnement multidimensionnel métrique, afin de déterminer l'espace déformé dans lequel la stationnarité et l'isotropie sont assurées. Cependant, le calcul de cette distance nécessite la donnée des paramètres d'anisotropie à chaque emplacement du domaine d'étude ; ils sont calculés sur la base de variogrammes locaux. Dans leur application, Almendral et al. [2008] modélisent les propriétés pétro-physiques telles que la porosité et la perméabilité.

Boisvert et al. [2009], Boisvert and Deutsch [2011] proposent une approche similaire pour modéliser l'anisotropie locale variable. Leur méthode nécessite aussi des paramètres d'anisotropie à chaque endroit. Par ailleurs, la mise en œuvre de cette technique exige que les paramètres d'anisotropie (directions et les ratios d'anisotropie) soient fournis sur une grille régulière. Leur méthodologie utilise des algorithmes d'optimisation (Newton, Dijkdtra) pour minimiser la distance d'anisotropie entre deux points dans l'espace, ou de manière équivalente à maximiser la covariance. Cette approche est appliquée à des données de sondages d'une mine de cuivre. McBratney and Minasny [2013] proposent également une approche similaire basée sur le calcul d'une certaine distance géographique. Leur méthode est appliquée sur des données de sol. Malgré l'absence d'un modèle

de second ordre, ces approches basées sur une distance non-euclidienne peuvent être considérées comme des approches par déformation d'espace.

1.4 Fonctions orthogonales empiriques

Cette approche introduite par Cohen and Jones [1969] consiste à représenter une fonction aléatoire gaussienne centrée $Z = \{Z(\mathbf{x}), \mathbf{x} \in G \subseteq \mathbb{R}^p\}$ sous la forme d'une décomposition de Karhunen-Loève (K-L) de sa fonction de covariance. Cela conduit à une représentation sous la forme :

$$Z(\mathbf{x}) = \sum_{k=1}^{+\infty} w_k \phi_k(\mathbf{x}), \tag{1.9}$$

où pour tout $k \in \mathbb{N}^\star$ les variables w_k sont gaussiennes, centrées et mutuellement indépendantes de variance λ_k avec $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$; $\{\phi_k(.)\}$ sont des fonctions de base orthogonales en ce sens que $\int_G \phi_k(\mathbf{x})\phi_{k'}(\mathbf{x})d\mathbf{x} = \delta_{kk'}$, δ étant le symbole de Kronecker.

Notons que $w_k = \int_G Z(\mathbf{x})\phi_k(\mathbf{x})d\mathbf{x}$ est la projection orthogonale de Z sur l'espace engendré par les $\phi_k(.)$. La covariance de Z s'écrit :

$$C(\mathbf{x}, \mathbf{y}) \equiv \mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = \sum_{k=1}^{+\infty} \lambda_k \phi_k(\mathbf{x})\phi_k(\mathbf{y}). \tag{1.10}$$

Ainsi, les $\phi_k(.)$ sont les solutions de l'équation intégrale de Fredholm du 2$^{\text{nd}}$ ordre :

$$\int_G C(\mathbf{x}, \mathbf{y})\phi_k(\mathbf{y})d\mathbf{y} = \lambda_k \phi_k(\mathbf{x}), \tag{1.11}$$

où $\phi_k(.)$ est la k-ième fonction propre de la covariance $C(.,.)$ et λ_k est la valeur propre correspondante.

La représentation continue K-L (1.9) n'est généralement pas utilisée dans les applications en raison de la nature discrète des données et de la difficulté de résoudre l'équation intégrale (1.11). Les expressions analytiques pour les fonctions propres ne sont connues que dans quelques cas simples.

En pratique, $C(\mathbf{x}, \mathbf{y})$ est remplacée par l'approximation $\sum_{k=1}^{K} \lambda_k \phi_k(\mathbf{x})\phi_k(\mathbf{y})$ pour un certain entier K. On obtient alors une approximation de Z sous la forme :

$$Z(\mathbf{x}) = \sum_{k=1}^{K} w_k \phi_k(\mathbf{x}), \tag{1.12}$$

qui est la meilleure approximation d'ordre K par rapport à n'importe quel autre ensemble de fonctions de base orthogonales, au sens de l'erreur quadratique moyenne intégrée sur le domaine G.

Dans certaines situations, des approximations numériques des fonctions propres peuvent être obtenues en effectuant une analyse en composantes principales sur la matrice de covariance empirique qui est estimée à partir des données. Tel est le cas lorsqu'on dispose d'observations répétées de la fonction aléatoire Z (par exemple, à travers le temps) qui permettent de calculer la matrice de covariance empirique. Ainsi, si $\mathbf{C}$ est la matrice de covariance empirique, il suffit de résoudre le problème aux valeurs propres :

$$\mathbf{C\Phi} = \mathbf{\Phi\Lambda}, \tag{1.13}$$

où $\mathbf{\Phi}$ est la matrice de vecteurs propres et $\mathbf{\Lambda}$ la matrice diagonale avec les valeurs propres correspondantes sur la diagonale.

Comme l'ont souligné Gelfand et al. [2010], ces approximations discrètes peuvent cependant ne pas bien approcher les vraies fonctions propres puisque la matrice de covariance empirique est affectée par le bruit de mesure. Par ailleurs, parce qu'elles sont discrètes, elles doivent être interpolées d'une manière quelconque afin de produire une approximation du modèle continu.

Une autre approche pour estimer les fonctions orthogonales empiriques et les valeurs propres correspondantes est de restreindre les $\phi_k(.)$ dans un espace fonctionnel généré par des fonctions de base pré-spécifiées. Ensuite, la version classique ou spatialement pondérée d'une analyse en composantes principales peut être appliquée directement.

Bien que la représentation en fonctions orthogonales empiriques (1.9) soit une approche très générale, flexible et puisse être considérée comme optimale dans le sens mentionné en (1.12), elle n'est pas facile à interpréter d'un point de vue géostatistique du fait de l'absence de connexion avec les approches classiques basées sur le variogramme. Par ailleurs, dans le cadre d'une réalisation unique le modèle est difficile à estimer et donc la réplication des observations est nécessaire pour utiliser ce type d'approche.

Nychka and Saltzman [1998], Holland et al. [1999] proposent de représenter la fonction de covariance comme la somme d'un modèle de covariance stationnaire isotrope et une décomposition (finie) de fonctions orthogonales empiriques. Ceci correspond à une décomposition de la fonction aléatoire comme une somme d'une fonction aléatoire stationnaire isotrope et d' une combinaison linéaire de fonctions de base supplémentaires à coefficients aléatoires ; la dernière somme représentant la déviation de la structure spatiale à la stationnarité. Ces approches sont utilisées pour la prédiction spatiale de dioxyde de soufre dans l'est des États-Unis [Nychka and Saltzman, 1998] et de l'ozone [Holland et al., 1999]. Chang et al. [2010] étendent le modèle de Holland et al. [1999] en remplaçant la fonction aléatoire stationnaire isotrope par une somme de plusieurs fonctions aléatoires stationnaires isotropes. Leur méthodologie est appliquée à des données de précipitation

aux États-Unis d'Amérique.

Nychka et al. [2002] décrivent une approche dans laquelle les fonctions orthogonales empiriques sont remplacées par des ondelettes. L'accent est mis sur les problèmes de calcul pour les grands jeux de données. La représentation en ondelettes est motivée par la non-stationnarité. Cependant, les données doivent être sur une grille et les données irrégulières doivent être mises en correspondance avec une grille régulière avant d'appliquer cette méthode. Leur modèle de covariance est appliqué à la sortie d'un modèle d'ozone. Des travaux récents de Matsuo et al. [2011] et Nychka et al. [2014] étendent cette approche pour prendre en compte des grands ensembles de données irrégulièrement espacées. L'approche de Matsuo et al. [2011] est illustrée sur des données d'ozone et celle de Nychka et al. [2014] sur des données de pluie.

Pintore and Holmes [2004], Stephenson et al. [2005] ont aussi travaillé avec les décompositions de Karhunen-Loève et de Fourier. La non-stationnarité est introduite en faisant évoluer spatialement la densité spectrale dans le domaine de fréquence. Les modèles qui en résultent sont valides en termes de covariance, mais avec des paramètres locaux variant spatialement. L'inférence se fait dans le cadre bayésien et la méthode est appliquée à des données de précipitation de l'Écosse.

Cressie and Johannesson [2008] proposent une méthode de krigeage à rang fixe destinée à des grands ensembles de données. La méthode proposée utilise une combinaison linéaire d'un petit nombre de fonctions de base et estime la matrice de covariance pour les coefficients de la combinaison linéaire. Cette approche est appliquée sur des données d'ozone.

1.5 Partitionnement et fenêtre glissante

1.5.1 Partitionnement

Dans certaines applications géostatistiques, la variable régionalisée est fortement liée à la structure géologique sous-jacente. Dans ce type de situations, on peut espérer voir des transitions nettes dans la structure de dépendance spatiale entre les points à travers les strates. Autrement dit, la structure de dépendance spatiale montre des changements brusques aux interfaces entre les roches. Ainsi, les modèles non-stationnaires qui supposent généralement que la covariance ou le variogramme entre deux points varie lentement avec la distance sont inappropriés. Dans ce contexte, il convient de procéder à une partition du domaine d'intérêt et d'ajuster des modèles stationnaires dans chaque sous-domaine. Le domaine d'intérêt peut être divisé en utilisant par exemple des informations a priori ou un algorithme de classification. Les méthodes qui utilisent cette stratégie s'appellent des méthodes de partitionnement [Lloyd, 2011].

Atkinson and Lloyd [2007] modélisent la structure de covariance non-stationnaire en divisant la région d'intérêt en petits domaines dans lequel la covariance est supposée stationnaire. Cette approche est appliquée à des données topographiques. Kim et al. [2005] proposent un modèle qui peut faire face à ces transitions nettes dans la structure de covariance. La méthode qu'ils adoptent repose sur un modèle bayésien de partition aléatoire. Le domaine spatial est partitionné en utilisant la tesselation de Voronoi de telle sorte qu'à l'intérieur des sous-régions, la fonction aléatoire est supposée stationnaire. L'indépendance est supposée à travers les sous-régions. Au sein de chaque partition, Kim et al. [2005] supposent que les données sont issues d'une fonction aléatoire gaussienne, de fonction de corrélation stationnaire isotrope. L'incertitude sur le nombre de régions disjointes, leurs formes et le modèle au sein des régions est traitée dans le cadre bayésien. Le modèle est appliqué à un problème de perméabilité du sol où cette structure non-stationnaire discrète semble justifiée.

Banerjee et al. [2004] proposent un modèle à partition fixe avec l'hypothèse gaussienne au sein des partitions. Cette approche est appliquée à la modélisation des prix de maison à Stockton, en Californie. Gosoniu et al. [2009] adoptent le modèle de Banerjee et al. [2004] en relâchant l'hypothèse d'indépendance entre les régions et la gaussiannité. Gosoniu and Vounatsou [2011] étendent le travail de Gosoniu et al. [2009] en supposant une partition aléatoire. Ils partitionnent le domaine spatial en sous-régions, à l'aide de la tesselation de Voronoi, et supposent une structure de covariance spatiale stationnaire au sein des sous-régions et une corrélation entre elles. Leur approche est appliquée aux données sur la prévalence du paludisme en Afrique de l'Ouest.

L'approche par partitionnement appelle quelques remarques. Dans certaines partitions, les données peuvent ne pas être suffisamment denses pour estimer correctement les covariances locales. Il est aussi difficile de justifier du choix des paramètres de covariance à la frontière de deux sous-régions. Remarquons enfin que dans cette approche l'apparition des discontinuités est tout à fait naturelle et normale, puisque les partitions ne se chevauchent pas ; un modèle global cohérent n'existe pas toujours.

1.5.2 Fenêtre glissante

Une autre approche pour l'adaptation locale largement utilisée est celle de la fenêtre glissante. Cette technique a été introduite par Haas [1990a,b] pour améliorer la prédiction spatiale dans le cas des données non-stationnaires. L'idée de l'approche par fenêtre glissante est d'estimer la structure de dépendance spatiale (représentée par un modèle de variogramme) dans une fenêtre circulaire autour de chaque point cible (au sein de laquelle il est raisonnable de supposer un modèle stationnaire) où une prédiction spatiale (krigeage) est calculée. Cette opération est exécutée en chaque point cible. Ainsi, la méthode

conserve toutes les techniques statistiques des fonctions aléatoires stationnaires, tout en n'ayant pas à supposer que la stationnarité s'applique à l'ensemble de la région d'intérêt. Cette approche fournit une alternative au partitionnement de l'ensemble du domaine en un petit nombre de sous-régions disjointes.

La prédiction est faite au centre de la fenêtre en utilisant les paramètres estimés du variogramme local stationnaire. Le voisinage pour le krigeage local peut éventuellement être différent de celui utilisé pour estimer les paramètres du variogramme mais ils sont généralement identiques en pratique. Le choix de la taille des fenêtres doit équilibrer la nécessité d'avoir un nombre suffisant de paires pour estimer de façon satisfaisante les paramètres du variogramme et d'utiliser des fenêtres aussi petites que possible de telle sorte que l'hypothèse de stationnarité dans ces fenêtres soit tenable. Haas [1990b] décrit une approche heuristique pour déterminer la taille du voisinage : agrandir un cercle autour du point cible jusqu'à ce qu'au moins 35 sites soient inclus, ensuite inclure à chaque étape 5 sites jusqu'à ce qu'il y ait au moins une paire de points dans chaque classe de distance sur laquelle le variogramme est estimé et de sorte que l'estimation par moindres carrés non-linéaires du variogramme local converge. L'ajustement automatique des variogrammes locaux en chaque position peut augmenter considérablement les exigences en calcul. Les exemples d'application de l'approche par fenêtre glissante sont la prédiction de dépôts de sulfate et d'acide sulfurique aux États-Unis d'Amérique [Haas, 1990b,a].

Plusieurs méthodes ont été développées dans l'esprit de la fenêtre glissante. Harris et al. [2010] remplacent l'estimateur par fenêtre glissante du variogramme local par un estimateur pondéré. Cette approche est appliquée à des données de pluie, de sol et de radioactivité ambiante. Magneron et al. [2010] proposent une méthodologie fournissant un ensemble de techniques d'optimisation pour les paramètres locaux impliqués dans les modèles de variogramme, en particulier les paramètres structuraux tels que la portée et la direction d'anisotropie. Le calcul des paramètres structuraux locaux basé sur une validation croisée se fait par un calcul des paramètres des variogrammes locaux dans les zones adjacentes du champ de données, suivi d'un lissage des paramètres obtenus afin de les rendre disponibles en chaque point. Cette approche est appliquée sur des données de bathymétrie de la côte Marenne-Oléron, à l'ouest de la France. Felder [2011] développe des méthodes de traitement d'images pour la détermination des paramètres variographiques locaux, en faisant l'hypothèse d'un comportement localement stationnaire de la variable étudiée. Son approche est illustrée sur des données de porosité et de bathymétrie. Machuca-Mory and Deutsch [2012] proposent une méthodologie où ils construisent des modèles locaux de fonctions aléatoires qui sont considérés comme stationnaires seulement au voisinage de points où ils sont définis. Les distributions spatiales locales sont supposées être des anamorphoses gaussiennes et leurs propriétés statistiques sont inférées par une

approche de distance pondérée (inversement proportionnelle) aux points cibles. Les paramètres de variogrammes locaux et les paramètres de transformation ainsi calculés sont utilisés dans les algorithmes de prédiction spatiale localement stationnaire. Cette procédure est appliquée à des données de teneur en or.

Une source de préoccupation dans l'approche par fenêtre glissante est le choix de la taille des fenêtres. Ceci ajouté à la difficulté d'estimer de manière fiable les variogrammes locaux lorsque les données sont rares ou très variables. Par ailleurs, comme les points de données sont inclus et soudainement exclus du voisinage lorsque la cible se déplace, il peut résulter des discontinuités sur la carte des paramètres locaux et de krigeage. Rappelons aussi qu'en voisinage glissant, l'estimateur par krigeage n'est plus l'estimateur linéaire optimal, du fait de l'exclusion de certaines observations. L'approche par fenêtre glissante est locale (elle définit une fonction aléatoire stationnaire en chaque voisinage de prédiction), et les interprétations locales ne s'étendent pas facilement à des interprétations globales, de même qu'aucune approche locale ne garantit d'avoir de bonnes propriétés globales. Ainsi, cette approche ne conduit pas à un modèle unique pour décrire l'ensemble des données. Par conséquent, aucun modèle global de covariance [5] n'existe pour cette approche.

1.6 Convolution

L'idée générale derrière la représentation par convolution d'une fonction aléatoire est que des variables corrélées peuvent s'écrire comme des combinaisons linéaires de variables non-corrélées. Soit l'exemple suivant proposé par Schabenberger and Gotway [2005b] :

Exemple 1.1. Considérons $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées suivant une loi de Bernoulli de paramètre α. Alors, $U = \sum_{i=1}^{k} X_i \sim \mathcal{B}(k, \alpha)$ et $V = \sum_{i=1}^{k+m} X_i \sim \mathcal{B}(k+m, \alpha)$.

On peut voir que les variables aléatoires U et V sont corrélées puisqu'elles partagent k variables : $\mathbb{C}\mathrm{ov}(U, V) = \min(k, k+m)\alpha(1-\alpha)$.

1.6.1 Modèle continu

L'approche par convolution introduite par Higdon [1998], Higdon et al. [1999] consiste à représenter une fonction aléatoire non-stationnaire (gaussienne) comme la convolution d'un noyau déterministe variant spatialement avec un bruit blanc (gaussien). Higdon [1998], Higdon et al. [1999] définissent une fonction aléatoire gaussienne $Z = \{Z(\mathbf{x}), \mathbf{x} \in$

5. L'absence d'un modèle global ne permet pas par exemple, de calculer la covariance en toute paire de points non-observés.

$G \subseteq \mathbb{R}^p\}$ comme suit :

$$Z(\mathbf{x}) = \int_{\mathbb{R}^p} k_{\mathbf{x}}(\mathbf{u}) X(\mathbf{u}) d\mathbf{u}, \ \forall \mathbf{x} \in G, \tag{1.14}$$

où $k_{\mathbf{x}}(.) = k(\mathbf{x}, .)$, $k(.,.)$ étant un noyau [6] non-aléatoire dont les paramètres varient spatialement ; $X(.)$ est le bruit blanc gaussien standard. Par bruit blanc gaussien standard, on entend une fonction aléatoire telle que : $W(A) \equiv \int_A X(\mathbf{u}) d\mathbf{u} \sim \mathcal{N}(0, \nu(A))$ et $\mathbb{C}\mathrm{ov}(W(A), W(B)) = \nu(A \cap B)$, pour tout ensemble mesurable $A, B \subset \mathbb{R}^p$; $\nu(.)$ étant la mesure de Lebesgue sur $\mathbb{R}^p$.

Même si dans la pratique, la fonction $k(.,.)$ est souvent prise dans la classe des fonctions symétriques et définies positives, il important de noter que ce n'est pas une nécessité. En effet, il suffit que $\sup_{\mathbf{x}} \int_{\mathbb{R}^p} k_{\mathbf{x}}^2(\mathbf{u}) d\mathbf{u} < +\infty$ pour que le modèle spécifié en (1.14) conduit à une covariance non-stationnaire définie positive sur $\mathbb{R}^p \times \mathbb{R}^p$:

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) \equiv C(\mathbf{x}, \mathbf{y}) = \int_{\mathbb{R}^p} k_{\mathbf{x}}(\mathbf{u}) k_{\mathbf{y}}(\mathbf{u}) d\mathbf{u}. \tag{1.15}$$

Le noyau $k(.,.)$ contrôle la régularité de la fonction aléatoire Z [Abrahamsen, 1997]. En effet, si la fonction $k_{\mathbf{x}}(.)$ est différentiable en tout $\mathbf{x} \in G$, alors la fonction aléatoire Z est différentiable et :

$$\frac{\partial Z(\mathbf{x})}{\partial \mathbf{x}_i} = \int_{\mathbb{R}^p} \frac{\partial k_{\mathbf{x}}(\mathbf{u})}{\partial \mathbf{x}_i} X(\mathbf{u}) d\mathbf{u}, \quad \mathbb{C}\mathrm{ov}\left(\frac{\partial Z(\mathbf{x})}{\partial \mathbf{x}_i}, \frac{\partial Z(\mathbf{y})}{\partial \mathbf{y}_j}\right) = \int_{\mathbb{R}^p} \frac{\partial k_{\mathbf{x}}(\mathbf{u})}{\partial \mathbf{x}_i} \frac{\partial k_{\mathbf{y}}(\mathbf{u})}{\partial \mathbf{y}_j} d\mathbf{u}. \tag{1.16}$$

De cette manière, il est possible d'obtenir des fonctions aléatoires gaussiennes régulières en utilisant un noyau régulier. Si le noyau $k(.,.)$ est invariant par translation (stationnaire), c'est-à-dire $k_{\mathbf{x}}(\mathbf{u}) = k_0(\mathbf{x} - \mathbf{u})$ pour tout $\mathbf{x}$, alors Z est de covariance stationnaire (auto-convoluée) :

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) \equiv C_0(\mathbf{h}) = \int_{\mathbb{R}^p} k_0(\mathbf{u}) k_0(\mathbf{u} - \mathbf{h}) d\mathbf{u}. \tag{1.17}$$

Dans ce cas, la relation entre la covariance $C_0(.)$, sa densité spectrale $S_0(.)$ et le noyau $k_0(.)$ est donnée par le théorème de convolution pour les transformées de Fourier :

$$(2\pi)^p |\mathcal{F}(k_0)|^2 = \mathcal{F}(C_0) = S_0, \tag{1.18}$$

où $\mathcal{F}(.)$ est la transformée de Fourier.

Ainsi, $C_0(.)$ est la transformée de Fourier inverse du carré de la transformée de Fourier de $k_0(.)$. Réciproquement, pour $C_0(.)$ donnée, $k_0(.)$ est obtenue comme l'inverse de la

6. Le terme noyau a plusieurs significations distinctes en statistique. Sauf mention contraire, tout au long du document, un noyau (que nous appellerons aussi fonction de pondération) sur $\mathbb{R}^p \times \mathbb{R}^p$ est une fonction positive $k : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}_+$.

transformée de Fourier de la racine carrée de la densité spectrale de $C_0(.)$. Si en plus $k_0(.)$ est isotrope, $C_0(.)$ l'est aussi et la correspondance entre $C_0(.)$ et $k_0(.)$ est bijective [Gaetan and Guyon, 2009]. Une famille flexible de noyaux stationnaires isotropes est obtenue à partir de la classe de corrélations stationnaires isotropes de type Matérn [Stein, 1999]. La densité spectrale d'une fonction de corrélation stationnaire isotrope de type Matérn dans $\mathbb{R}^2$, de paramètres d'échelle $a > 0$ et de régularité $\nu > 0$ est donnée par $f(\omega) \propto 1/(a^2+\omega^2)^{\nu/2+1}$. Le noyau correspondant est l'inverse de la transformée de Fourier de $1/(a^2+\omega^2)^{\nu/4+1/2}$ qui est proportionnel à :

$$(a\|\mathbf{u}\|)^{\nu} K_{\nu}(a\|\mathbf{u}\|),\ a>0, \nu>0, \tag{1.19}$$

où $K_\nu(.)$ est la fonction de Bessel modifiée de second espèce d'ordre ν [Gradshteyn and Ryzhik, 2007].

La relation entre le noyau $k_0(.)$ et la covariance $C_0(.)$ est discutée en détail par Matérn [1986], Oliver [1995], Ver Hoef and Barry [1998] et Kern [2000]. Le tableau 1.1 présente des exemples de noyaux (stationnaires et isotropes) et leur covariances (stationnaire et isotrope) associées [Matérn, 1986].

	Noyaux $k_0(.)$	Covariances $C_0(.)$
Gauss	$\exp(-2a^2\|\mathbf{u}\|^2)$	$\propto \exp(-a^2\|\mathbf{h}\|^2)$
Cauchy	$(1+4\|\mathbf{u}\|^2/a^2)^{-\frac{p+1}{2}}$	$\propto (1+\|\mathbf{h}\|^2/a^2)^{-\frac{p+1}{2}}$
Matérn	$(a\|\mathbf{u}\|)^{\nu}K_{\nu}(a\|\mathbf{u}\|)$	$\propto (a\|\mathbf{h}\|)^{2\nu+p/2}K_{2\nu+p/2}(a\|\mathbf{h}\|)$

TABLEAU 1.1 – Modèle de convolution : exemples de noyaux et leur covariance associée.

Dans le cas où $k_0(.)$ est un noyau stationnaire isotrope, on peut dériver une covariance elliptique en remplaçant $k_0(\|\mathbf{u}\|)$ par $k_0(\sqrt{\mathbf{u}^T\mathbf{A}\mathbf{u}})$ où $\mathbf{A}$ est une matrice définie positive. La covariance stationnaire isotrope $C_0(\|\mathbf{h}\|)$ correspondant alors à $k_0(\|\mathbf{u}\|)$ est ainsi changée en $C_0(\sqrt{\mathbf{h}^T\mathbf{A}\mathbf{h}})$.

La représentation par convolution présente quelques atouts. Elle permet de modéliser le noyau $k(.,.)$ plutôt que la covariance $C(.,.)$ qui doit être définie positive. Ainsi, la formule de convolution permet d'obtenir les modèles du second ordre sans avoir à vérifier la condition de définie positivité d'une covariance. Par ailleurs, elle permet de générer des modèles non-gaussiens dès que la convolution est valide. Une simple extension du modèle (1.14) est de modifier la spécification régissant la fonction aléatoire latente $X(.)$. Par exemple, Wolpert and Ickstadt [1998] spécifient $X(.)$ comme un processus de Lévy (dont un cas particulier est celui où $X(.)$ est une mesure aléatoire de Poisson). Ainsi, le processus résultant n'est pas gaussien. Une telle approche peut être appropriée dans la modélisation

de taux ou de concentrations qui ne suivent généralement pas une distribution gaussienne. De plus, l'approche par convolution fournit un moyen direct et simple de simulation de fonctions aléatoires sans recourir à des inversions de matrices ou calcul de déterminants.

Un point crucial dans cette approche est la définition ou le choix du noyau défini en tout point. Les noyaux généralement utilisés sont des noyaux isotropes ou elliptiques. Contrairement à ce à quoi on pourrait s'attendre, on est limité dans le choix de la famille de noyaux. En effet, l'inférence des modèles de convolution n'est aisé que dans le contexte de stationnarité locale. Dans ce contexte, l'estimation n'est possible qu'avec des noyaux qui localement sont connectés à des covariances stationnaires paramétriques. Tel est le cas des noyaux présentés dans le tableau 1.1. De la sorte, l'estimation des paramètres de la covariance locale stationnaire permet d'obtenir ceux du noyau correspondant.

Pour une fonction aléatoire non-stationnaire sur $\mathbb{R}^2$, Higdon [1998], Higdon et al. [1999] choisissent $k_{\mathbf{x}}(.)$ comme une densité gaussienne bivariée centrée en $\mathbf{x}$, de matrice de covariance $\mathbf{\Sigma_x}$ (appelée matrice de noyau) variant spatialement :

$$k_{\mathbf{x}}(\mathbf{u}) = \frac{1}{2\pi}|\mathbf{\Sigma_x}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{u}-\mathbf{x})^T(\mathbf{\Sigma_x})^{-1}(\mathbf{u}-\mathbf{x})\right). \tag{1.20}$$

La covariance non-stationnaire de la fonction aléatoire Z a une expression analytique donnée par :

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = \frac{1}{\pi}\left|\frac{\mathbf{\Sigma_x}+\mathbf{\Sigma_y}}{2}\right|^{-\frac{1}{2}} \exp\left(-Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})\right), \tag{1.21}$$

où $Q_{\mathbf{xy}}(\mathbf{h}) = \mathbf{h}^T\left(\frac{\mathbf{\Sigma_x}+\mathbf{\Sigma_y}}{2}\right)^{-1}\mathbf{h}$, $\forall \mathbf{h} \in \mathbb{R}^2$.

Pour $\mathbf{\Sigma_x} = \mathbf{\Sigma}$ une matrice constante, on retrouve la covariance gaussienne à anisotropie géométrique où les axes principaux de $\mathbf{\Sigma}$ déterminent les directions de la structure anisotrope.

Pour paramétrer la matrice de noyau $\mathbf{\Sigma_x}$, Higdon [1998] et Higdon et al. [1999] utilisent la correspondance bijective qui existe entre une ellipse standard d'un noyau gaussien et le noyau lui-même. Une ellipse peut être définie de plusieurs façons ce qui implique différentes formes de paramétrisation. D'une part, une ellipse peut être définie à travers un ensemble de trois paramètres : centre, un des deux foyers, et aire ; d'autre part, par la donnée de l'excentricité et la direction du grand axe. Bien que ces propriétés géométriques peuvent être dérivées l'une de l'autre, cette approche ne se limite pas à la paramétrisation qu'ils ont choisie. D'autres paramétrisations sont possibles et peuvent être plus facilement applicables dans certaines situations.

En utilisant la définition bi-focale de l'ellipse et en considérant $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}))^T$ et $-\mathbf{F}(\mathbf{x}) = (-F_1(\mathbf{x}), -F_2(\mathbf{x}))^T$, $\forall \mathbf{x} \in G$, les points focaux (foyers) de l'ellipse centrée à l'origine et d'aire fixe A, Higdon et al. [1999] modélisent la matrice de covariance $\mathbf{\Sigma_x}$ comme suit :

$$\mathbf{\Sigma}_{\mathbf{x}}^{\frac{1}{2}} = \tau_z \begin{pmatrix} \left[\frac{\sqrt{4A^2+\|\mathbf{F}\|^4\pi^2}}{2\pi} + \frac{\|\mathbf{F}\|^2}{2}\right]^{\frac{1}{2}} & 0 \\ 0 & \left[\frac{\sqrt{4A^2+\|\mathbf{F}\|^4\pi^2}}{2\pi} - \frac{\|\mathbf{F}\|^2}{2}\right]^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \cos\{\theta(\mathbf{x})\} & \sin\{\theta(\mathbf{x})\} \\ -\sin\{\theta(\mathbf{x})\} & \cos\{\theta(\mathbf{x})\} \end{pmatrix}, \tag{1.22}$$

où $\|\mathbf{F}(\mathbf{x})\|^2 = F_1(\mathbf{x})^2 + F_2(\mathbf{x})^2$, $\theta(\mathbf{x}) = \tan^{-1}\{F_2(\mathbf{x})/F_1(\mathbf{x})\}$ et τ_z est un facteur d'échelle du noyau.

Pour écrire $\mathbf{\Sigma_x}$ en fonction de ces paramètres, on utilise certaines propriétés géométriques des ellipses. Cela permet d'utiliser une ellipse comme un moyen plus intuitif de paramétrer $\mathbf{\Sigma_x}$. Ainsi, l'expression (1.22) fournit une correspondance entre une ellipse définie par son centre, foyer et aire, et le noyau gaussien bivarié qu'elle représente. Travaillant dans le contexte bayésien, Higdon et al. [1999] spécifient une distribution spatiale a priori du noyau $k_{\mathbf{x}}(.)$ en modélisant $F_1(\mathbf{x})$ et $F_2(\mathbf{x})$ comme des fonctions aléatoires gaussiennes, indépendantes, chacune de covariance stationnaire isotrope gaussienne : $\exp(-r^2/\tau_F^2)$. Ceci signifie que $\mathbf{F}(\mathbf{x})$ a une distribution bivariée gaussienne standard en tout point $\mathbf{x}$. Le choix de τ_F dépend de l'application et celui de A est fixé à une valeur qui donne des ellipses raisonnables lorsque $F_1(\mathbf{x})$ et $F_2(\mathbf{x})$ sont tirés de la distribution gaussienne. Ces choix se font après une analyse exploratoire des données. Swall [1999] a constaté des problèmes de sur-apprentissage lorsque l'aire de l'ellipse A est autorisée à varier. C'est la raison pour laquelle elle fixe l'aire de l'ellipse à une valeur constante commune à tous les emplacements. Un facteur d'échelle τ_z est autorisé à varier en chaque endroit.

Une autre paramétrisation de la matrice de noyau $\mathbf{\Sigma_x}$ peut se faire par décomposition spectrale. L'anisotropie locale est alors capturée en permettant une rotation et un étirement ou contraction des axes de coordonnées [Higdon, 1998] :

$$\mathbf{\Sigma}_{\mathbf{x}}^{\frac{1}{2}} = \begin{pmatrix} \lambda_1(\mathbf{x}) & 0 \\ 0 & \lambda_2(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \cos\{\theta(\mathbf{x})\} & \sin\{\theta(\mathbf{x})\} \\ -\sin\{\theta(\mathbf{x})\} & \cos\{\theta(\mathbf{x})\} \end{pmatrix}, \tag{1.23}$$

où $\lambda_1(.)$ est la longueur du grand axe, $\lambda_2(.)$ celle du petit axe et $\theta(.)$ la direction du grand axe.

Toujours dans le contexte bayésien, afin d'assurer que les noyaux varient lentement sur le domaine d'intérêt, Neto et al. [2013] intègrent une certaine régularité dans le comportement de $\mathbf{\Sigma_x}$. Les distributions a priori de $\log\{\lambda_j(.)\}$ sont modélisées comme des fonctions aléatoires gaussiennes stationnaires, indépendantes de moyenne μ_{λ_j}, de variance

σ_λ^2 et de covariance stationnaire isotrope gaussienne : $\sigma_\lambda^2 \exp(-r^2/\tau_\lambda^2)$. Le paramètre $\theta(.)$ lui est modélisé comme suit : $\theta(\mathbf{x}) = \frac{\pi}{2}\Phi\{\Xi(\mathbf{x})\}$, où $\Phi(.)$ est la fonction de répartition d'une loi gaussienne standard et $\Xi(.)$ est une fonction aléatoire gaussienne de moyenne μ_Ξ et de covariance stationnaire isotrope gaussienne : $\sigma_\Xi^2 \exp(-r^2/\tau_\Xi^2)$.

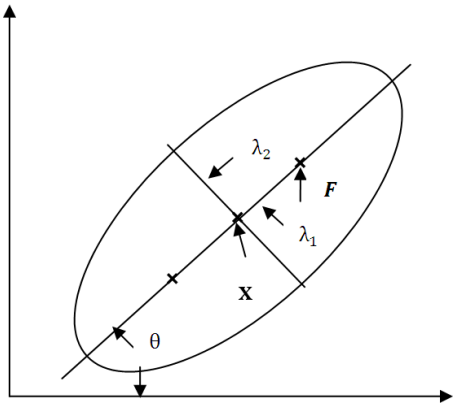


FIGURE 1.4 – Modèle de convolution : représentation de la matrice de noyau.

La structure de dépendance spatiale définie en (1.21) est caractérisée par les paramètres de la matrice d'anisotropie $\mathbf{\Sigma_x}$ définie en tout point $\mathbf{x} \in G$. L'estimation de ces paramètres est requise en tout point du domaine d'intérêt. Pour que l'inférence soit possible, Higdon [1998] s'appuie sur l'hypothèse de stationnarité locale et de la connexion entre le variogramme gaussien et le noyau de convolution gaussien. Tout d'abord, les paramètres du noyau sont estimés pour un sous-ensemble de sites pour limiter le temps de calcul. En chaque site du sous-ensemble, les paramètres de variogramme sont estimés dans le voisinage local et transformés pour correspondre aux paramètres du noyau gaussien. Il s'agit d'une simple transformation puisqu'une fonction aléatoire spécifiée par un variogramme stationnaire gaussien est équivalente à une fonction aléatoire obtenue par convolution d'un bruit blanc avec un noyau gaussien. Enfin, les paramètres estimés du noyau sont interpolés de telle sorte que le noyau varie lentement sur l'ensemble du domaine. Plus précisément, les paramètres du noyau gaussien aux points non-estimés sont calculés par une moyenne pondérée des paramètres aux points estimés. Higdon [1998] applique cette méthode à la prédiction de températures dans l'Océan Atlantique.

Habituellement, l'inférence des modèles de convolution se fait dans le contexte bayésien. Higdon et al. [1999] considèrent le noyau de convolution comme une fonction inconnue que l'on contraint à être lisse. Ils prennent en compte l'incertitude dans la spécification de la non-stationnarité en développant un modèle hiérarchique qui peut intégrer cette incertitude dans l'inférence. Le processus sous-jacent est considéré comme un paramètre du modèle, et sa distribution a posteriori peut être trouvée directement dans une version simple du modèle ou en ayant recours à des algorithmes de Monte Carlo

par chaîne de Markov. L'exemple d'application est l'estimation de la concentration en dioxines sur un site dans le Missouri, aux États-Unis d'Amérique. Higdon [2002] discute de l'utilisation de la procédure de maximum de vraisemblance restreint pour ajuster les modèles de convolution.

Higdon [1998], Higdon et al. [1999] choisissent un noyau gaussien principalement pour sa commodité de calcul et de tractabilité analytique. Ce choix conduit à une expression explicite de la covariance non-stationnaire, et par conséquent de la vraisemblance. La méthode de paramétrisation s'étend au-delà de l'utilisation du noyau gaussien, mais les auteurs n'ont pas implémenté d'autres formes de noyaux. Une conséquence de l'utilisation du noyau gaussien est que le modèle de covariance locale est également gaussien. Ceci implique que la fonction aléatoire est infiniment différentiable. Dans le cadre stationnaire, Stein [1999] prévient que la modélisation des phénomènes réels avec une telle covariance n'est pas réaliste. En effet, il montre qu'en utilisant un développement de Taylor d'une telle fonction aléatoire à une origine arbitraire, il est facile de voir qu'on peut prédire exactement la fonction aléatoire en tout autre point du domaine simplement en connaissant la fonction aléatoire dans un certain voisinage arbitrairement petit.

Calder and Cressie [2007] discutent d'un certain nombre de sujets liés à la modélisation par convolution, y compris des problèmes numériques pour les grands ensembles de données. D'Hondt et al. [2007] applique le modèle de convolution à noyau gaussien dans l'analyse de la texture d'images radar.

1.6.2 Modèle discret

Si un noyau autre que le noyau gaussien est utilisé dans l'équation (1.14), il arrive généralement qu'il ne conduise pas à une expression analytique de la covariance. Dans ce cas, l'usage est de remplacer le modèle de convolution continu (1.14) par une approximation en somme finie [Higdon, 1998] :

$$Z(\mathbf{x}) = \sum_{j=1}^{L} k_{\mathbf{x}}(\mathbf{u}_j) X(\mathbf{u}_j), \tag{1.24}$$

où $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_L\} \in G$ est un support de convolution permettant d'approcher raisonnablement l'intégrale (1.14).

Le résultat de la discrétisation de la fonction aléatoire sous-jacente est que les variables latentes $\{X(\mathbf{u}_j), j = 1, \ldots, L\}$ contrôlent complètement la fonction aléatoire Z. Cette réduction de la dimension s'est avérée fructueuse pour la modélisation de différents types de données géostatistiques telles que : les données de température de l'Océan Atlantique ([Higdon, 1998]), les données de concentration en azote (Kern [2000]) et les problèmes

inverses en hydrologie d'eaux souterraines (Higdon [2002]).

Une décision importante de modélisation est l'emplacement de la fonction aléatoire sous-jacente $X(.)$, c'est-à-dire le choix du support $\mathcal{U}$. Un choix possible pour $\mathcal{U}$ proposé par Gaetan and Guyon [2009] est un réseau régulier triangulaire de maille δ ; δ doit réaliser un compromis entre d'une part un bon ajustement aux données (maille petite) et d'autre part des calculs simples (maille plus grande). Une solution intermédiaire consiste à utiliser un modèle multi-résolution à deux mailles ou plus [Higdon, 2002]. D'après Higdon [2002], la convolution sur une grille de maille δ est une bonne approximation de la convolution continue de la fonction aléatoire qui en résulte, dès que la maille δ est inférieure au minimum des valeurs propres de la matrice de noyau $\mathbf{\Sigma_x}$.

Dans la version discrète du modèle de convolution, la covariance est donnée par :

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = \sum_{j=1}^{L} k_{\mathbf{x}}(\mathbf{u}_j) k_{\mathbf{y}}(\mathbf{u}_j). \tag{1.25}$$

L'utilisation de convolutions discrètes génère un problème très important celui des effets de bord illustrés par la figure 1.5. Ainsi, le nombre de point de $\mathcal{U}$ contenu dans le support du noyau est en moyenne plus faible sur les bordures. Une manière de traiter cela est d'étendre légèrement le support de la fonction aléatoire latente au-delà de la région d'intérêt, comme présentée à la figure 1.6.

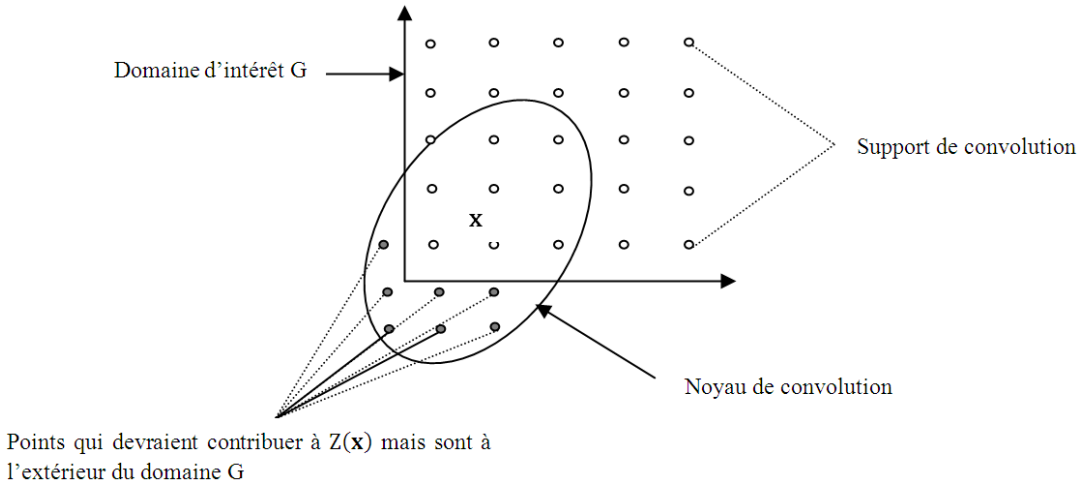


FIGURE 1.5 – Modèle de convolution : présence d'effets de bord.

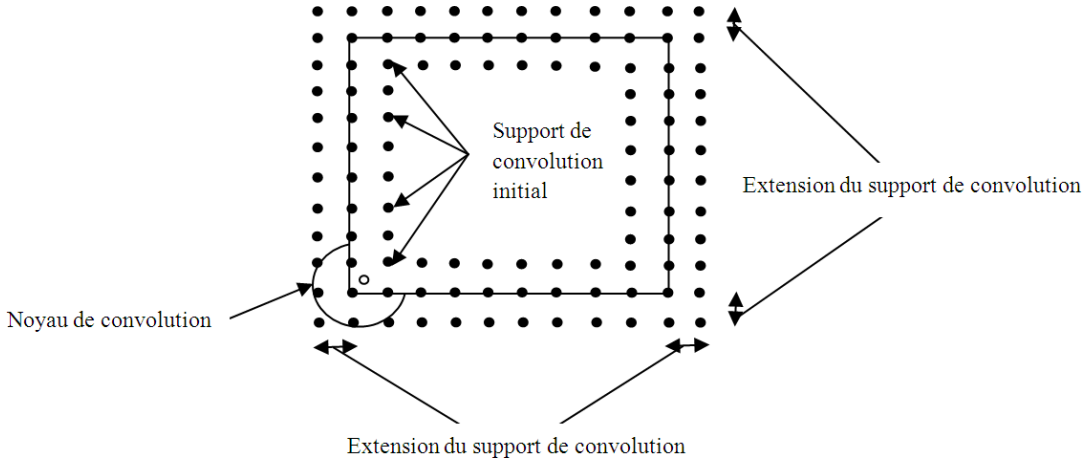
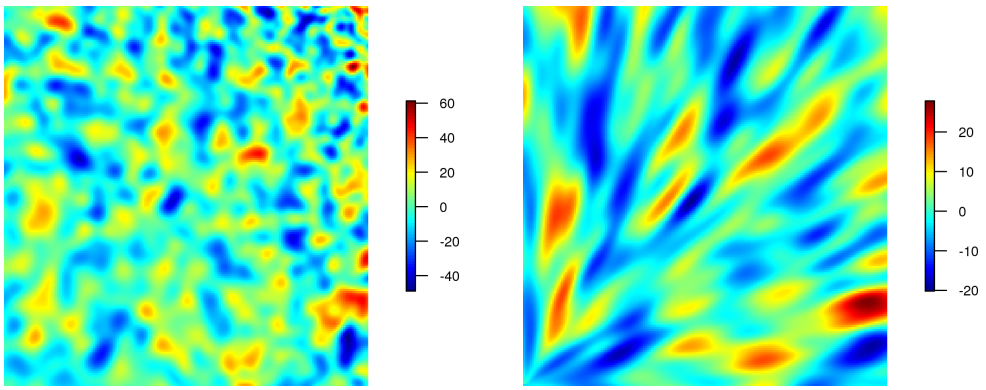


FIGURE 1.6 – Modèle de convolution : correction d'effets bord.

Les figures 1.7a et 1.7b présentent des réalisations de fonctions aléatoires gaussiennes issues du modèle de convolution. Sur figure 1.7a, la structuration est isotrope mais les tailles des structures locales varient progressivement : elles sont grandes en bas à gauche de l'image (sud-ouest) puis deviennent de plus en plus petites jusqu'en en haut à droite (nord-est). Sur la figure 1.7b, les structures locales sont orientées horizontalement en bas à droite (sud-est) puis varient progressivement jusqu'à atteindre une orientation verticale en haut à gauche (nord-ouest).



(a) Portée variable (b) Anisotropie variable

FIGURE 1.7 – Modèle de convolution : réalisations de fonctions aléatoires non-stationnaires.

1.6.3 Extensions

Zhu and Wu [2010] généralisent l'approche par convolution de Higdon [1998], en utilisant un noyau variant spatialement donné par les fonctions de Bessel modifiées de second espèce. Leur modèle produit des fonctions aléatoires avec un comportement local semblable aux fonctions aléatoires à covariance stationnaire isotrope de Matérn. Leur modèle ne conduit pas à une expression analytique pour la covariance, par conséquent la

prédiction spatiale par krigeage n'est pas aisée. Par ailleurs, l'anisotropie locale variable n'est pas prise en compte dans leur approche. Cette dernière est appliquée à des données de précipitation aux États-Unis d'Amérique.

Étant donnée que la classe de covariances non-stationnaires (analytiques) obtenue à partir du modèle de convolution à noyau gaussien de Higdon [1998] est indéfiniment différentiable, Paciorek and Schervish [2006] ont étendu cette classe de covariance, pour permettre d'obtenir différents degrés de régularité à l'origine. Cette nouvelle classe comprend entre autre une version non-stationnaire de la covariance stationnaire de Matérn.

Pour un noyau gaussien dans le modèle de convolution (1.14), la covariance non-stationnaire résultante a une expression analytique de la forme suivante :

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = (\sqrt{\pi})^{-p} \left| \frac{\mathbf{\Sigma_x} + \mathbf{\Sigma_y}}{2} \right|^{-\frac{1}{2}} R^S \left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x} - \mathbf{y})} \right), \tag{1.26}$$

où $R^S(.)$ est la fonction de corrélation stationnaire, isotrope et gaussienne.

Paciorek and Schervish [2006] montrent alors qu'en substituant la fonction de corrélation gaussienne $R^S(.)$ par toute autre fonction de corrélation définie positive dans toute dimension, la covariance résultante est toujours valide. D'où la proposition 1.2 donnant une représentation plus large de classe de corrélations non-stationnaires valides sur $\mathbb{R}^p \times \mathbb{R}^p$.

Proposition 1.2. *Paciorek and Schervish [2006]*

Si une fonction de corrélation stationnaire et isotrope $R^S(.)$ est définie positive sur $\mathbb{R}^p$ pour tout $p \in \mathbb{N}^\star$, alors la fonction définie par :

$$R^{NS}(\mathbf{x}, \mathbf{y}) = |\mathbf{\Sigma_x}|^{\frac{1}{4}} |\mathbf{\Sigma_y}|^{\frac{1}{4}} \left| \frac{\mathbf{\Sigma_x} + \mathbf{\Sigma_y}}{2} \right|^{-\frac{1}{2}} R^S \left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x} - \mathbf{y})} \right), \tag{1.27}$$

est une fonction de corrélation non-stationnaire définie positive sur $\mathbb{R}^p \times \mathbb{R}^p$ pour tout $p \in \mathbb{N}^\star$.

Paciorek and Schervish [2006] introduisent ainsi une nouvelle classe de corrélations non-stationnaires valides, de formules explicites, à anisotropie géométrique locale variable. La classe comprend une version non-stationnaire de la covariance stationnaire de Matérn, dans laquelle la régularité de la fonction aléatoire est contrôlée par un paramètre. La classe permet d'insérer des paramètres de covariance locale stationnaire dans une covariance globale non-stationnaire valide, quelle que soit la façon dont la structure de covariance locale est estimée. Paciorek and Schervish [2006] emploient cette nouvelle classe de covariances non-stationnaires sur des données de précipitation dans le Colorado et sous une formulation bayésienne ; la fonction aléatoire a une distribution a priori gaussienne de covariance non-stationnaire appartenant à cette classe. Ils suggèrent aussi des approches

non-bayésiennes.

Pour une application pratique, Paciorek and Schervish [2006] proposent la séquence générale suivante pour estimer la structure de covariance non-stationnaire :

1. partitionner le domaine d'intérêt non-stationnaire en petites parties qui soient stationnaires autant que possible ;
2. estimer les paramètres de covariance localement stationnaire anisotrope, en utilisant soit une approche par variogramme ou une maximisation de la vraisemblance marginale dans laquelle seules les données dans une sous-région sont utilisées pour estimer les paramètres spécifiques de la sous-région ;
3. définir $\mathbf{\Sigma}_{\mathbf{x}} = \mathbf{\Sigma}_{G(i)}$, où $G(i)$ représente la sous-région dans laquelle l'emplacement $\mathbf{x}$ tombe. $\mathbf{\Sigma}_{G(i)}$ est construite pour chaque sous-région à partir des paramètres de la structure de covariance localement stationnaire anisotrope estimée pour la sous-région.

Une autre possibilité, toujours suggérée par Paciorek and Schervish [2006] consiste à estimer la matrice de noyau en chaque emplacement par fenêtre glissante ou variogramme pondéré. En effet, si les éléments de la matrice de noyau varient lentement, la covariance non-stationnaire est localement approximativement stationnaire et anisotrope. Ainsi, en tout point cible $\mathbf{x}$ définissant un certain voisinage, on peut estimer des paramètres locaux $\boldsymbol{\theta}_{\mathbf{x}}$ à partir desquels la matrice de noyau $\mathbf{\Sigma}_{\mathbf{x}}$ sera construite. Pour l'estimation de ces paramètres basés sur le variogramme, on peut utiliser une fenêtre glissante pour inclure seulement les paires d'emplacements proches de $\mathbf{x}$. On peut aussi attribuer des poids à chaque paire de sites et estimer $\boldsymbol{\theta}_{\mathbf{x}}$ sur la base de variogrammes pondérés. Par ailleurs, on pourrait paramétrer la covariance non-stationnaire en fonction de covariables clés afin de réduire la dimension du problème d'estimation. Cette idée a été mise en œuvre par exemple par Neto et al. [2013]. Paciorek and Schervish [2006] suggèrent aussi d'estimer $\boldsymbol{\theta}_{\mathbf{x}}$ par maximum de vraisemblance, en utilisant la vraisemblance marginale seulement pour les observations dont l'emplacement est proche de $\mathbf{x}$ ou une approche par maximum de vraisemblance locale pondérée.

Néanmoins, la classe de covariances non-stationnaires proposée par Paciorek and Schervish [2006] ne permet pas de faire varier spatialement d'autres aspects de la structure de covariance tels que la régularité ou l'indice de dépendance de longue portée. En utilisant des spectres variant spatialement, Pintore and Holmes [2004] montrent que l'on peut obtenir des covariances non-stationnaires explicites dont le degré de différentiabilité varie dans l'espace. Stein [2005] montre comment ces deux idées peuvent être combinées pour produire des classes très flexibles de covariances non-stationnaires explicites. C'est l'objet de la proposition 1.3.

Proposition 1.3. *Stein [2005]*

Si $\mu(.)$ est une mesure positive finie sur $\mathbb{R}_+$ et pour tout $\mathbf{x} \in \mathbb{R}^p$, $g(.;\mathbf{x}) \in L^2(\mu)$, alors la fonction définie par :

$$C^{NS}(\mathbf{x},\mathbf{y}) = |\mathbf{\Sigma_x}|^{\frac{1}{4}}|\mathbf{\Sigma_y}|^{\frac{1}{4}}\left|\frac{\mathbf{\Sigma_x}+\mathbf{\Sigma_y}}{2}\right|^{-\frac{1}{2}}\int_0^{+\infty} e^{-wQ_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}g(w;\mathbf{x})g(w;\mathbf{y})\mu(dw), \qquad (1.28)$$

est une fonction de covariance non-stationnaire définie positive sur $\mathbb{R}^p \times \mathbb{R}^p$ pour tout $p \in \mathbb{N}^\star$.

Pour un choix approprié de la fonction $g(.;\mathbf{x})$ et de la mesure $\mu(.)$, l'expression (1.28) permet d'obtenir une version non-stationnaire de la covariance stationnaire de Matérn où la régularité est autorisée à varier spatialement. On obtient aussi une version non-stationnaire de la covariance stationnaire de Cauchy où le paramètre de longue portée est autorisé à varier spatialement.

Dans le même esprit, Mateu et al. [2010] présentent une nouvelle classe de covariances non-stationnaires à support compact. Les covariances à support compact permettent de réduire considérablement la charge de calcul du krigeage, par l'utilisation de techniques efficaces de calcul de matrices creuses, devenant ainsi un aspect essentiel dans la prédiction spatiale lorsqu'on traite de grands jeux de données. Toutefois, il y a lieu de préciser que cette classe de covariances non-stationnaires n'est pas analytique, ce qui limite son utilisation.

En utilisant la version non-stationnaire de la covariance stationnaire de Matérn, Anderes and Stein [2011] développent une estimation des paramètres qui régissent la dépendance spatiale locale d'une fonction aléatoire localement stationnaire par vraisemblance locale pondérée. L'avantage de cette estimation par vraisemblance locale pondérée est qu'elle diminue lentement le poids de l'influence des observations lointaines et fonctionne sur des emplacements d'échantillonnage irrégulier.

L'approche de modélisation par convolution fournit une manière attractive d'introduire la non-stationnarité tout en conservant une interprétation claire. Elle permet un calcul analytique dans certains cas. L'estimation repose sur l'hypothèse de stationnarité locale qui impose aux paramètres de la structure de dépendance spatiale de varier régulièrement et lentement dans l'espace. Ainsi, l'approche est incapable de gérer un changement brutal de la structuration spatiale des données. Elle n'est donc adaptée que pour des changements réguliers de la structuration.

1.7 Lissage par noyau

1.7.1 Modèles locaux stationnaires pondérés

Une approche pour traiter la non-stationnarité est de partir de modèles locaux stationnaires puis de les lisser empiriquement sur l'ensemble du domaine d'étude. Dans l'approche par fenêtre glissante pour l'estimation des fonctions aléatoires non-stationnaires des estimations locales de la structure de covariance sont calculées mais elles ne sont pas intégrées dans un modèle global. Les travaux de Fuentes [2001, 2002b,a] développent cette idée pour représenter la structure de covariance non-stationnaire en termes de combinaisons spatialement pondérées de fonctions de covariance stationnaire supposées représenter la structure locale dans les différentes régions du domaine d'étude.

Fuentes [2001] suppose qu'une fonction aléatoire non-stationnaire $Z = \{Z(\mathbf{x}), \mathbf{x} \in G \subseteq \mathbb{R}^p\}$ définie sur une région d'intérêt G découpée en M sous-régions disjointes G_i ($G = \cup_{i=1}^{M} G_i$) peut s'écrire comme un mélange de fonctions aléatoires localement stationnaires et orthogonales $Z_1(\mathbf{x}), \ldots, Z_M(\mathbf{x})$:

$$Z(\mathbf{x}) = \sum_{i=1}^{M} w_i(\mathbf{x}) Z_i(\mathbf{x}), \ \forall \mathbf{x} \in G, \tag{1.29}$$

où $\{Z_i, i = 1, \ldots, M\}$ sont des fonctions aléatoires sous-jacentes (non-observables) stationnaires, orthogonales et définies sur G ; $\{w_i(\mathbf{x}), i = 1, \ldots, M\}$ sont des poids dépendants d'une distance entre une sous-région $G_i \subset G$ et le point $\mathbf{x}$ (par exemple, inverse de la distance au carré). La distance peut par exemple être la distance au centre de gravité d'une sous-région ou la moyenne des distances à tous les points à l'intérieur d'une sous-région. La partition du domaine d'intérêt est supposée connue a priori.

Les fonctions aléatoires localement stationnaires étant non-corrélées, on a $\mathbb{C}\mathrm{ov}(Z_i(\mathbf{x}), Z_j(\mathbf{x})) = 0, \forall (i \neq j)$ et chacune a une covariance $\mathbb{C}\mathrm{ov}(Z_i(\mathbf{x}), Z_i(\mathbf{y})) \equiv C(\mathbf{x} - \mathbf{y}; \theta_i)$. Le résultat est une covariance non-stationnaire de la fonction aléatoire Z définie comme suit :

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) \equiv C(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{M} w_i(\mathbf{x}) w_i(\mathbf{y}) C(\mathbf{x} - \mathbf{y}; \theta_i). \tag{1.30}$$

La représentation (1.29) est basée sur l'idée qu'il y a M sous-domaines de stationnarité pouvant être représentés par M fonctions aléatoires stationnaires et orthogonales Z_i chacune définie sur G. Les poids $w_i(\mathbf{x})$ agissent comme des fenêtres ou des filtres qui donnent plus d'importance à la fonction aléatoire $Z_i(\mathbf{x})$ pour la localisation $\mathbf{x} \in G_i$ et moins d'importance pour les localisations extérieures à G_i. En supposant que sur G_i, $w_i(\mathbf{x}) \approx 1$ et $w_j(\mathbf{x}) \approx 0$ pour $j \neq i$, la fonction aléatoire non-stationnaire Z en $\mathbf{x} \in G_i$ est

approximée par la fonction aléatoire stationnaire $Z_i(\mathbf{x})$. La covariance entre $Z(\mathbf{x})$ et $Z(\mathbf{y})$, $\forall(\mathbf{x},\mathbf{y}) \in G_i \times G_i$ est alors approximée par $C(\mathbf{x}-\mathbf{y};\theta_i)$. La covariance entre $Z(\mathbf{x})$ et $Z(\mathbf{y})$, avec $\mathbf{x}$ et $\mathbf{y}$ appartenant respectivement aux sous-régions adjacentes G_i et G_j, est alors approximée par la combinaison $w_i(\mathbf{x})w_i(\mathbf{y})C(\mathbf{x}-\mathbf{y};\theta_i) + w_j(\mathbf{x})w_j(\mathbf{y})C(\mathbf{x}-\mathbf{y};\theta_j)$.

Les fonctions aléatoires stationnaires Z_i sont locales uniquement dans le sens où leur covariance associée $C(.;\theta_i)$ est estimée localement, avec les données de la sous-région G_i. Elles sont orthogonales par hypothèse en vue de représenter la covariance globale non-stationnaire comme une somme pondérée de covariances stationnaires. Fuentes [2001] détermine le nombre de sous-régions par le critère d'information bayésien (BIC) ou le critère d'information d'Akaike (AIC). Chaque sous-région est supposée avoir un nombre suffisant de données pour permettre l'estimation de la covariance locale stationnaire. Les paramètres de cette dernière peuvent être estimés par maximum de vraisemblance ou par moindres carrés.

Fuentes [2002b] propose une extension de la décomposition finie (1.29) en une convolution continue de fonctions aléatoires localement stationnaires :

$$Z(\mathbf{x}) = \int_G k(\mathbf{x}-\mathbf{u}) Z_{\boldsymbol{\theta}(\mathbf{u})}(\mathbf{x}) d\mathbf{u},\ \forall \mathbf{x} \in G, \tag{1.31}$$

où $k(.)$ est une fonction de pondération ; $\boldsymbol{\theta}(.)$ est une fonction multivariée qui induit la non-stationnarité dans Z. $\{Z_{\boldsymbol{\theta}(\mathbf{u})}(\mathbf{x}), \mathbf{x} \in G\}$ est une famille de fonctions aléatoires stationnaires d'ordre 2 (gaussiennes), centrées, indépendantes (à travers $\mathbf{u}$), de covariances $C(.;\boldsymbol{\theta}(\mathbf{u}))$. Par exemple, $C(.;\boldsymbol{\theta}(\mathbf{u})) = \sigma^2(\mathbf{u}) \exp(-\phi(\mathbf{u})\|.\|^{\alpha(\mathbf{u})})$, avec $\boldsymbol{\theta}(\mathbf{u}) = (\sigma^2(\mathbf{u}), \phi(\mathbf{u}), \alpha(\mathbf{u}))$.

L'expression (1.31) est un mélange d'un nombre indénombrable de fonctions aléatoires stationnaires (indexées par $\boldsymbol{\theta}(.)$) en chaque emplacement $\mathbf{x}$. Formellement, Z est de moyenne nulle et de covariance s'exprimant comme une convolution de covariances locales :

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = \int_G k(\mathbf{x}-\mathbf{u}) k(\mathbf{y}-\mathbf{u}) C(\mathbf{x}-\mathbf{y};\boldsymbol{\theta}(\mathbf{u})) d\mathbf{u}. \tag{1.32}$$

L'expression (1.32) montre que Z est une fonction aléatoire non-stationnaire. Comme $\boldsymbol{\theta}(\mathbf{u})$ peut varier sensiblement sur l'ensemble du domaine d'étude, ce modèle permet également une forte non-stationnarité.

Si $k(.)$ est une fonction fortement piquée (décroît très rapidement) et $\boldsymbol{\theta}(\mathbf{u})$ varie lentement avec $\mathbf{u}$, on a la propriété que pour $\mathbf{x}$ près de $\mathbf{u}$, la fonction aléatoire non-stationnaire Z est proche d'une fonction aléatoire stationnaire régie par $\boldsymbol{\theta}(\mathbf{u})$. Ainsi, l'approche proposée par Fuentes [2002b] se réfère à la classe des modèles non-stationnaires qui présente une stationnarité locale.

L'estimation du modèle (1.32) requiert que le paramètre $\boldsymbol{\theta}(.)$ indexant les fonctions aléatoires stationnaires varie lentement. Fuentes [2001] applique une méthode non-linéaire par moindres carrés pondérés dans le domaine des fréquences en utilisant des périodogrammes lorsque des données sont régulièrement espacées et propose une méthode de maximum de vraisemblance restreint dans le cas contraire.

En pratique, on ne travaille pas directement avec l'expression (1.31) ; une approximation par somme finie est employée. L'intégrale (1.32) et la représentation spectrale de sa covariance sont approximées avec une somme discrète impliquant un ensemble fini de localisations indépendantes $\{\mathbf{u}_1, \ldots, \mathbf{u}_L\} \in G$ et les fonctions aléatoires stationnaires correspondantes $Z_{\boldsymbol{\theta}_j}(\mathbf{x})$ comme dans l'équation suivante :

$$Z(\mathbf{x}) = \sum_{j=1}^{L} k(\mathbf{x} - \mathbf{u}_j) Z_j(\mathbf{x}), \forall \mathbf{x} \in G,\ Z_{\boldsymbol{\theta}(\mathbf{u}_j)} \equiv Z_j. \tag{1.33}$$

D'où,

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) = \sum_{j=1}^{L} k(\mathbf{x} - \mathbf{u}_j) k(\mathbf{y} - \mathbf{u}_j) C_j(\mathbf{x} - \mathbf{y}). \tag{1.34}$$

Dans l'expression (1.33), il peut arriver que le point $\mathbf{x}$ soit très éloigné de chaque $\mathbf{u}_j$ de sorte que $k(\mathbf{x} - \mathbf{u}_j) \approx 0$, d'où $Z(\mathbf{x}) \approx 0$; ce qui ne peut pas se produire avec la version continue définie en (1.31). Une manière possible de remédier à cela a été proposée par Banerjee et al. [2004]. Il s'agit de remplacer l'expression (1.33) par la suivante :

$$Z(\mathbf{x}) = \sum_{j=1}^{L} \alpha(\mathbf{x}, \mathbf{u}_j) Z_j(\mathbf{x}),\ \forall \mathbf{x} \in G, \tag{1.35}$$

où $\alpha(\mathbf{x}, \mathbf{u}_j) = \beta(\mathbf{x}, \mathbf{u}_j) / \sqrt{\sum_{j=1}^{L} \beta^2(\mathbf{x}, \mathbf{u}_j)}$. $\beta(\mathbf{x}, \mathbf{u})$ est une fonction décroissante de la distance entre $\mathbf{x}$ et $\mathbf{u}$, qui peut changer avec $\mathbf{x}$ ($\beta(\mathbf{x}, \mathbf{u}) = k_{\mathbf{x}}(\|\mathbf{x} - \mathbf{u}\|)$).

Ainsi, $\sum_{j=1}^{L} \alpha^2(\mathbf{x}, \mathbf{u}_j) = 1$, pour tout $\mathbf{x} \in G$ et les poids de (1.35) sont non tous nuls. D'autres normalisations pour $\alpha(.,.)$ sont possibles. La normalisation ci-dessus permet d'avoir $\mathbb{V}(Z(\mathbf{x})) = \sigma^2$ lorsque tous les $C_j(0) = \sigma^2$ sont égaux. Ainsi, si tous les $Z_j(\mathbf{x})$ ont la même variance, alors cette variance est attachée à $Z(\mathbf{x})$. Par ailleurs, en supposant que deux points $\mathbf{x}$ et $\mathbf{y}$ sont proches, on a : $\beta(\mathbf{x}, \mathbf{u}_j) \approx \beta(\mathbf{y}, \mathbf{u}_j)$ et donc $\alpha(\mathbf{x}, \mathbf{u}_j) \approx \alpha(\mathbf{y}, \mathbf{u}_j)$. Si de plus, tous les paramètres des covariances locales sont égaux ($\boldsymbol{\theta}_j = \boldsymbol{\theta}$), alors $\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) \approx \sigma^2 \rho(\mathbf{x} - \mathbf{y}; \boldsymbol{\theta})$.

Une alternative de normalisation beaucoup plus intuitive est de prendre $\alpha(\mathbf{x}, \mathbf{u}_j) = \beta(\mathbf{x}, \mathbf{u}_j) / \sum_{j=1}^{L} \beta(\mathbf{x}, \mathbf{u}_j)$. De tels poids préservent une moyenne constante. Notons que $L = 1$ est autorisé dans l'expression (1.35) et Z est toujours non-stationnaire.

L'approche par fonctions aléatoires stationnaires pondérées est relativement proche de celle par convolution. La différence importante entre les deux est l'hypothèse suivant laquelle la fonction de pondération (noyau) évolue spatialement. Dans l'approche par convolution, les paramètres du noyau varient spatialement, tandis que dans l'approche par fonctions aléatoires stationnaires pondérées, les paramètres de covariance des fonctions aléatoires stationnaires varient spatialement mais les paramètres du noyau sont maintenus constants. Par conséquent, la non-stationnarité de la covariance ne provient pas du changement local du noyau, mais des différences entre les covariances stationnaires. Cette approche a des traits communs avec l'approche par fenêtre glissante, mais ici il n'y a aucun problème quant à la définition d'un modèle global cohérent. Elle est affectée par des mêmes problèmes que l'approche par fenêtre glissante, c'est-à-dire la détermination de la taille des partitions et la fiabilité des covariances locales estimées lorsque les données sont rares. Un inconvénient majeur de cette approche est le manque d'une formule explicite pour le calcul de la covariance et la nécessité de calculer celle-ci par des méthodes numériques. En outre, certains éléments clés de l'approche, comme le nombre de composantes localement stationnaires, la taille des partitions pour l'ajustement des modèles locaux, et la nature du noyau doivent être déterminés par des méthodes ad hoc.

La figure 1.8 montre une réalisation d'une fonction aléatoire non-stationnaire par modèles stationnaires locaux pondérés. Quatre sous-domaines de stationnarité de même taille (sous forme de rectangles) sont définis.

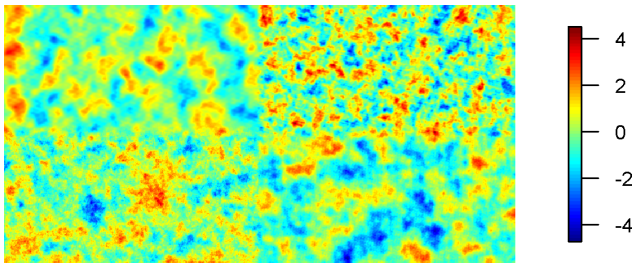


FIGURE 1.8 – Modèles locaux stationnaires pondérés : réalisation d'une fonction aléatoire non-stationnaire.

Une approche similaire à celle de Fuentes [2001] a été développée par Nott and Dunsmuir [2002]. Elle prend comme point de départ une matrice de covariance empirique pour un ensemble de sites d'observation (matrice potentiellement non-stationnaire) et l'étend à une fonction de covariance non-stationnaire valable sur toute la région d'intérêt ; ceci en utilisant une collection de fonctions aléatoires stationnaires avec des covariances choisies pour approximer au niveau local celle de la fonction aléatoire non-stationnaire aux sites d'observation. La même notation que dans celle de l'expression (1.30) est utilisée, mais pour Nott and Dunsmuir [2002] i indexe les sites d'observation (ou les sources probables de la non-stationnarité) plutôt qu'un petit nombre de sous-régions et $C(.;\theta_i)$ représentent

la structure de covariance locale résiduelle après conditionnement sur les valeurs aux sites d'observation. Celles-ci sont dérivées de l'ajustement de modèles localement stationnaires. La technique est illustrée sur des données des vents de Sydney, en Australie. Bien que les modèles introduits par Fuentes [2001] et Nott and Dunsmuir [2002] se ressemblent, les détails sont sensiblement différents. Nott and Dunsmuir [2002] supposent que la matrice de covariance empirique est calculée à partir de données spatio-temporelles, tandis que l'aproche de Fuentes [2001] s'applique à des données purement spatiales. Les deux méthodes ne prennent pas en compte l'ajustement de modèles de covariance localement anisotrope, mais faire cette extension ne devrait pas compliquer les calculs.

1.7.2 Variogramme (covariance) empirique lissé(e)

Il existe d'autres approches de lissage par noyau qui consistent à lisser la matrice de covariance (ou variogramme) empirique. Guillot et al. [2001] et Bel [2004] proposent des estimateurs à noyau similaires à celui introduit par Oehlert [1993].

Soient $Z_{it} = Z(\mathbf{s}_i, t)$ les variables issues de la fonction aléatoire $Z = \{Z(\mathbf{x}, t) : \mathbf{x} \in G \subseteq \mathbb{R}^2, t \in T \subseteq \mathbb{R}\}$ aux n emplacements $\{\mathbf{s}_i, i = 1, \ldots, n\}$ et aux mêmes m points $\{t = 1, \ldots, m\}$. Considérons $\mathbf{C} = [c_{ij}]$ et $\mathbf{\Gamma} = [\gamma_{ij}]$ les matrices de covariance spatiale et de variogramme spatial empiriques, calculées aux sites d'observation $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. Plus précisément :

$$c_{ij} = \frac{1}{m}\sum_{t=1}^{m}(Z(\mathbf{s}_i, t) - \bar{Z}(\mathbf{s}_i))(Z(\mathbf{s}_j, t) - \bar{Z}(\mathbf{s}_i)), \tag{1.36}$$

$$\gamma_{ij} = \frac{1}{2m}\sum_{t=1}^{m}\left((Z(\mathbf{s}_i, t) - Z(\mathbf{s}_j, t)) - (\bar{Z}(\mathbf{s}_i) - \bar{Z}(\mathbf{s}_j))\right)^2, \tag{1.37}$$

où $\bar{Z}(\mathbf{s}_i) = \frac{1}{m}\sum_{t=1}^{m} Z(\mathbf{s}_i, t)$ est la moyenne empirique en tout site $i = 1, \ldots, n$.

Oehlert [1993] a introduit une méthode à noyau pour estimer la covariance spatiale. Elle est fondée sur l'idée que la covariance spatiale devrait varier lentement. Son estimateur non-paramétrique est défini comme suit :

$$\hat{C}_\epsilon(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i,j} c_{ij} K_\epsilon((\mathbf{x}, \mathbf{y}), (\mathbf{s}_i, \mathbf{s}_j))}{\sum_{i,j} K_\epsilon((\mathbf{x}, \mathbf{y}), (\mathbf{s}_i, \mathbf{s}_j))}, \ \forall (\mathbf{x}, \mathbf{y}) \in G \times G, \tag{1.38}$$

où $K_\epsilon((\mathbf{x}, \mathbf{y}), (\mathbf{s}_i, \mathbf{s}_j)) = \exp(-(\|\mathbf{x} - \mathbf{s}_i\|^2 + \|\mathbf{y} - \mathbf{s}_j\|^2)/\epsilon^2)(\mathbb{1}_{\mathbf{x}\neq\mathbf{y}, \mathbf{s}_i\neq\mathbf{s}_j} + \mathbb{1}_{\mathbf{x}=\mathbf{y}, \mathbf{s}_i=\mathbf{s}_j})$.

Oehlert [1993] n'établit pas si cette méthode conduit toujours à une structure de covariance valide (définie positive). L'approche a été appliquée aux données annuelles de sulfate sur plusieurs réseaux de surveillance des pluies acides dans le nord-est des États-Unis d'Amérique et le sud-est du Canada.

L'estimateur non-paramétrique proposé par Guillot et al. [2001] s'écrit :

$$\hat{C}_\epsilon(\mathbf{x},\mathbf{y}) = \sum_{i,j} c_{ij} \int_{G_i \times G_j} K_\epsilon(\mathbf{x}-\mathbf{u},\mathbf{y}-\mathbf{v}) d\mathbf{u} d\mathbf{v},\ \forall (\mathbf{x},\mathbf{y}) \in G \times G, \tag{1.39}$$

où $K_\epsilon(\mathbf{u},\mathbf{v}) = \epsilon^{-4} K(\mathbf{u}/\epsilon, \mathbf{v}/\epsilon), \forall \epsilon > 0$ et $K(.,.)$ est un noyau défini sur $G \times G$, symétrique, défini positif et d'intégrale unité. $\cup_{i=1}^{n} G_i$ est une partition de G telle que $\mathbf{s}_i \in G_i$ pour tout $i = 1, \ldots, n$ (par exemple une tesselation de Voronoi basée sur les $\{\mathbf{s}_i\}$).

Guillot et al. [2001] montrent la définie positivité de cet estimateur pour les noyaux symétriques et définis positifs (noyau gaussien, noyau exponentiel, ...) et discutent de la sélection du paramètre de fenêtre ϵ. Leur méthode est appliquée à des données de précipitation en Afrique de l'Ouest. Étonnamment, le krigeage avec un modèle stationnaire donne des meilleurs résultats que celui avec leur covariance non-stationnaire.

Dans le même esprit, Bel [2004] propose un estimateur non-paramétrique du variogramme s'exprimant comme suit :

$$\hat{\Gamma}_\epsilon(\mathbf{x},\mathbf{y}) = \sum_{i,j} \gamma_{ij} \frac{K_\epsilon(\mathbf{x}-\mathbf{s}_i,\mathbf{y}-\mathbf{s}_i)}{\sum_{k,l} K_\epsilon(\mathbf{x}-\mathbf{s}_k,\mathbf{y}-\mathbf{s}_l)},\ \forall (\mathbf{x},\mathbf{y}) \in G \times G, \tag{1.40}$$

où $K_\epsilon(\mathbf{x},\mathbf{y}) = K(\mathbf{x}/\epsilon, \mathbf{y}/\epsilon), \forall \epsilon > 0$ et $K(.,.)$ est un noyau défini sur $G \times G$, symétrique, défini positif et séparable.

En se basant sur la propriété de séparabilité du noyau $K(.,.)$, Bel [2004] montre la définie négativité conditionnelle de l'estimateur non-paramétrique défini en (1.40). L'approche est utilisée pour produire des cartes de concentration d'ozone sur la région de Paris, en France.

1.8 Équations aux dérivées partielles stochastiques

Une autre stratégie de modélisation des fonctions aléatoires non-stationnaires est celle par équations aux dérivées partielles stochastiques (EDPS). Cette approche récente introduite par Lindgren et al. [2011] spécifie une fonction aléatoire gaussienne non-stationnaire comme la solution d'une EDPS non-homogène dont la structure de covariance est contrôlée par les coefficients de l'EDPS. Lindgren et al. [2011] approximent la solution au moyen de fonctions aléatoires gaussiennes markoviennes. L'idée principale consiste à obtenir une approximation par éléments finis de la solution en utilisant le développement en base $\sum_i w_i \varphi_i(\mathbf{x})$ où les $\{w_i\}$ sont des poids aléatoires. Pour certains choix de fonctions de base $\{\varphi_i\}$, en particulier celles à support compact, les poids induisent des fonctions aléatoires gaussiennes markoviennes. Ainsi, grâce à cette propriété markovienne des poids

(possédant une matrice de précision creuse [7]), des méthodes numériques rapides pour les matrices creuses peuvent être utilisées lors de l'estimation des paramètres et la prédiction spatiale dans ces modèles. Ceci améliore considérablement l'applicabilité à des problèmes impliquant de grands ensembles de données, pour lesquelles les méthodes traditionnelles échouent à cause des problèmes de calcul.

Cette idée est due à Whittle [1963] ; il prouve que la solution de l'EDPS suivante :

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau Z(\mathbf{x})) = X(\mathbf{x}),\ \forall \mathbf{x} \in \mathbb{R}^p,\ \alpha = \nu + p/2,\ \kappa > 0, \nu > 0, \tag{1.41}$$

est une fonction aléatoire gaussienne de covariance stationnaire de type Matérn [Stein, 1999]. $X(.)$ étant le bruit blanc gaussien standard ; $\Delta = \sum_{i=1}^{p} \frac{\partial^2}{\partial \mathbf{x}_i^2}$ est l'opérateur de Laplace ; $(\kappa^2 - \Delta)^{\alpha/2}$ est l'opérateur pseudo-différentiel défini au moyen de sa transformée de Fourier par : $\{\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi\}(\mathbf{k}) = (\kappa^2 + \|\mathbf{k}\|^2)^{\alpha/2}(\mathcal{F}\phi)(\mathbf{k})$, pour une fonction ϕ sur $\mathbb{R}^p$.

Au lieu de définir les fonctions aléatoires de type Matérn à travers leur covariance, Lindgren et al. [2011] utilisent la solution de l'équation (1.41) comme définition. Cette définition a l'avantage d'être valide non seulement sur $\mathbb{R}^p$, mais aussi sur des variétés lisses telle qu'une sphère. Par ailleurs, elle facilite l'extension à des fonctions aléatoires non-stationnaires en permettant aux paramètres κ^2 et τ de varier spatialement. D'où l'EDPS (1.42) suivante :

$$(\kappa^2(\mathbf{x}) - \Delta)^{\alpha/2}(\tau(\mathbf{x})Z(\mathbf{x})) = X(\mathbf{x}),\ \forall \mathbf{x} \in \mathbb{R}^p. \tag{1.42}$$

Sous réserve que les paramètres $\kappa^2(.)$ et $\tau(.)$ varient lentement, l'EDPS non-stationnaire (1.42) conduit à une covariance localement stationnaire isotrope, alors que le modèle de covariance global lui est non-stationnaire.

Afin de pouvoir capturer des anisotropies locales variables, l'EDPS (1.42) peut être étendue, en autorisant un Laplacien anisotrope. Ce qui conduit à l'EDPS ci-après ;

$$(\kappa^2(\mathbf{x}) - \nabla \cdot \mathbf{H}(\mathbf{x})\nabla)^{\alpha/2}(\tau(\mathbf{x})Z(\mathbf{x})) = X(\mathbf{x}),\ \forall \mathbf{x} \in \mathbb{R}^p, \tag{1.43}$$

où $\kappa^2(.)$ est une fonction continue, strictement positive ; $\mathbf{H}(.)$ est une fonction continûment différentiable telle que pour tout $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{H}(\mathbf{x})$ est une matrice définie positive ; ∇ désigne le gradient.

Lindgren et al. [2011] proposent également une modélisation par déformation d'espace sous forme d'EDPS, établissant ainsi un lien avec l'approche de Sampson and Guttorp [1992]. Dans l'approche par déformation d'espace, une fonction aléatoire non-stationnaire

7. La matrice de précision est définie comme l'inverse de la matrice de covariance.

est représentée comme étant la déformation d'une fonction aléatoire stationnaire. Ainsi, l'EDPS non-stationnaire correspondante peut être interprétée comme une EDPS stationnaire dans l'espace déformé.

Soit $f : G \subseteq \mathbb{R}^p \rightarrow \tilde{G} \subseteq \mathbb{R}^p$, $\mathbf{x} \mapsto \tilde{\mathbf{x}}$, la fonction de déformation. On s'intéresse au cas $\alpha = 2$. Considérons l'EDPS stationnaire dans l'espace déformé $\tilde{G}$ suivant :

$$(\kappa^2 - \tilde{\nabla} \cdot \tilde{\nabla})\tilde{Z}(\tilde{\mathbf{x}}) = \tilde{X}(\tilde{\mathbf{x}}),\ \forall \tilde{\mathbf{x}} \in \tilde{G}, \tag{1.44}$$

dont la solution est une fonction aléatoire stationnaire de covariance de type Matérn.

Lindgren et al. [2011] montrent qu'un changement de variables vers l'espace non-déformé G conduit au modèle suivant :

$$\frac{1}{\det\{\mathbf{F}(\mathbf{x})\}}\left[\kappa^2 - \det\{\mathbf{F}(\mathbf{x})\}\nabla \cdot \frac{\mathbf{F}(\mathbf{x})\mathbf{F}(\mathbf{x})^T}{\det\{\mathbf{F}(\mathbf{x})\}}\nabla\right] Z(\mathbf{x}) = \frac{1}{\det\{\mathbf{F}(\mathbf{x})\}^{1/2}} X(\mathbf{x}),\ \forall \mathbf{x} \in G, \tag{1.45}$$

où $\mathbf{F}(\mathbf{x})$ est le Jacobien de la fonction de déformation $f(.)$.

La solution de cette EDPS a pour fonction de covariance non-stationnaire celle donnée par le modèle de déformation d'espace de Sampson and Guttorp [1992] (basé sur une covariance stationnaire de type Matérn). Comme le soulignent Lindgren et al. [2011], les paramètres de l'EDPS résultante ne dépendent pas directement de la fonction de déformation elle-même, mais seulement de son Jacobien. Ainsi, pour Lindgren et al. [2011] une option possible pour paramétrer le modèle sans une construction explicite de la fonction de déformation est de contrôler le grand axe de la déformation locale donnée par $\mathbf{F}(\mathbf{x})$ à travers un champ de vecteurs donné par des informations sur des covariables par exemple. Une addition ou une soustraction d'un terme de dérivée directionnelle généralise davantage le modèle. Permettre à tous les paramètres, y compris la variance du bruit blanc, de varier spatialement, donne un modèle non-stationnaire très général qui comprend à la fois le modèle de déformation et celui défini en (1.43).

L'approche par EDPS a une certaine connexion avec l'approche par convolution. Les solutions aux EDPS présentées précédemment peuvent se représenter sous forme de convolution. En effet, ces EDPS peuvent s'écrire sous la forme générale :

$$\mathcal{L}(\mathbf{x})Z(\mathbf{x}) = X(\mathbf{x}),\ \forall \mathbf{x} \in G, \tag{1.46}$$

où $\mathcal{L}(.)$ est un opérateur différentiel linéaire.

Le lien entre le modèle de convolution et le modèle d'EDPS est la fonction de Green de l'opérateur différentiel linéaire $\mathcal{L}(.)$ dans l'équation (1.46). Une fonction de Green $G(\mathbf{x}, \mathbf{u})$ au point $\mathbf{u} \in \mathbb{R}^p$ correspondant à $\mathcal{L}(.)$ est une fonction solution de l'équation :

$$\mathcal{L}(\mathbf{x})G(\mathbf{x}, \mathbf{u}) = \delta(\mathbf{x} - \mathbf{u}), \tag{1.47}$$

où $\delta(.)$ est la fonction de Dirac.

En multipliant (1.47) par le bruit blanc gaussien $X(\mathbf{u})$ et en intégrant par rapport à $\mathbf{u}$ on obtient :

$$\int_{\mathbb{R}^p} \mathcal{L}(\mathbf{x})G(\mathbf{x}, \mathbf{u})X(\mathbf{u})d\mathbf{u} = \int_{\mathbb{R}^p} \delta(\mathbf{x} - \mathbf{u})X(\mathbf{u})d\mathbf{u} = X(\mathbf{x}). \tag{1.48}$$

$\mathcal{L}(.)$ étant un opérateur linéaire et n'agissant que sur $\mathbf{x}$, il s'ensuit que :

$$\mathcal{L}(\mathbf{x}) \left(\int_{\mathbb{R}^p} G(\mathbf{x}, \mathbf{u})X(\mathbf{u})d\mathbf{u} \right) = X(\mathbf{x}). \tag{1.49}$$

Des équations (1.46) et (1.49) on obtient :

$$Z(\mathbf{x}) = \int_{\mathbb{R}^p} G(\mathbf{x}, \mathbf{u})X(\mathbf{u})d\mathbf{u}. \tag{1.50}$$

D'où la représentation en convolution de la solution à l'EDPS (1.46). Ainsi, la fonction de Green associée au modèle d'EDPS correspond au noyau dans le modèle de convolution (1.14). La figure 1.9 présente une réalisation d'une fonction aléatoire gaussienne solution d'un modèle non-stationnaire d'EDPS [Lindgren et al., 2011]. L'anisotropie locale variable est assez visible.

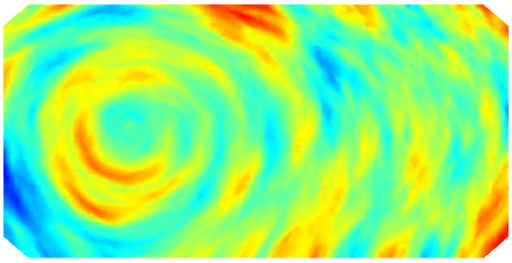


FIGURE 1.9 – Modèle d'EDPS : réalisation d'une fonction aléatoire non-stationnaire. [8]

Lindgren et al. [2011] montrent que les avantages de la représentation de fonctions aléatoires gaussiennes comme solutions d'EDPS ne sont pas seulement d'ordre numérique. En utilisant la représentation par EDPS, l'introduction de la non-stationnarité se fait aisément (en autorisant les paramètres dans l'EDPS à varier spatialement), et la classe de modèle peut être généralisé pour inclure des structures de covariance plus générales

8. Source : `http://www.math.ntnu.no/inla/r-inla.org/papers/spde-rss-presentation.pdf`

(non-isotropes, oscillantes, spatio-temporelles, ...). Ce sont des atouts d'un point de vue de l'application à des phénomènes nécessitant des modèles non-stationnaires complexes pour capturer habilement la structure de covariance dans les données.

Une limite de l'approche par EDPS est d'être connectée seulement aux covariances de type Matérn. Il n'est pas possible d'obtenir des fonctions aléatoires de covariance de type Cauchy, par exemple. Par ailleurs, les EDPS permettant une variabilité spatiale du degré de régularité n'ont pas encore été considérées.

L'inférence des modèles d'EDPS (estimation des coefficients) est généralement faite sous une formulation bayésienne. Ces modèles ont été appliquée à différents jeux de données : données météorologiques sur les continents [Lindgren et al., 2011], données d'ozone [Bolin and Lindgren, 2011], données de précipitation annuelles pour le sud de la Norvège [Ingebrigtsen et al., 2014] et les États-Unis d'Amérique [Fuglstad, 2013, 2014].

1.9 Conclusion

Dans ce chapitre, nous avons passé en revue un certain nombres d'approches permettant la prise en compte de la non-stationnarité structurale. Force est de constater qu'il est difficile de généraliser la manière dont une fonction aléatoire non-stationnaire peut être modélisée. Introduire la non-stationnarité dans la covariance (respectivement le variogramme) est problématique. En effet, il est difficile d'étendre les covariances (respectivement les variogrammes) stationnaires à des versions non-stationnaires tout en conservant la propriété essentielle de définie positivité (respectivement de définie négativité conditionnelle). Nous avons pu constater cela avec l'approche par points sources dans laquelle il est difficile de prouver la définie positivité du modèle de covariance non-stationnaire.

La majorité des approches de spécification de modèles non-stationnaires n'utilise pas la représentation en covariance (ou variogramme) directement, mais plutôt une représentation en terme de modèle de fonctions aléatoires qui induit une certaine covariance (ou variogramme) non-stationnaire valide. De la sorte, la structure de second ordre résultante est naturellement globale et cohérente en ce sens qu'elle est définie pour toute paire de points du domaine d'intérêt. De plus, de telles approches fournissent des algorithmes directs de simulation de fonctions aléatoires non-stationnaires. Tel est le cas des approches par déformation d'espace, convolution, fonctions orthogonales empiriques et équations aux dérivés partielles stochastiques.

Dans la plupart des approches, l'hypothèse de stationnarité est présente, soit en terme local comme dans les approches par convolution, fenêtre glissante et partitionnement, soit en terme global comme dans l'approche par déformation d'espace. Ainsi, ces approches

sont la preuve du rôle important que joue le concept de stationnarité en géostatistique. En outre, plusieurs de ces approches ont la particularité d'avoir naturellement les modèles stationnaires comme cas particuliers. Nous pouvons citer les approches par déformation d'espace, convolution, équations aux dérivés partielles stochastiques et points sources.

En ce qui concerne l'inférence, certaines approches nécessitent des réplications (par exemple à travers le temps) pour estimer la variabilité spatiale. Tandis que d'autres fonctionnent en réalisation unique, mais ont recours à une hypothèse de stationnarité locale. Dans la pratique, les approches par déformation d'espace, fonctions orthogonales empiriques et lissage par noyau de la covariance (ou du variogramme) empirique nécessitent d'avoir des réalisations multiples ; celles par convolution, fenêtre glissante et équations aux dérivés partielles stochastiques se contentent d'une réalisation unique.

Les méthodes d'estimation utilisées pour les modèles présentés se font soit dans le contexte fréquentiste soit dans le cadre bayésien. La plupart des méthodes sont semi-paramétriques : elles sont non-paramétriques par la façon dont la variation spatiale dans la structure de dépendance spatiale est décrite, mais la structure de dépendance spatiale locale est décrite à l'aide de modèles paramétriques classiques. D'après Sampson [2010], bien que le contexte bayésien permette de prendre en compte l'incertitude dans l'estimation de la structure de dépendance spatiale, les effets pratiques de cette incertitude n'ont pas encore été démontrés.

Selon Sampson [2010] les méthodes présentées ne sont pas encore matures à plusieurs égards. Tout d'abord, la majorité des approches examinées ne sont pas simples d'application. En effet, il y a très peu de logiciels disponibles pour l'estimation de ces modèles non-stationnaires. Bien que les modèles présentés soient attrayants et intuitifs, les méthodes d'estimation quant à elles ne sont pas toujours très claires. Un certain nombre de questions pratiques restent à régler de manière adéquate par le biais d'analyses et d'applications.

Il est important de souligner que les approches non-stationnaires sont intensives en calcul par rapport aux techniques stationnaires. Mais s'il y a suffisamment de données pour permettre une inférence fiable, elles surpassent souvent ces dernières en terme de prédiction et de précision. Parmi les principales approches examinées, il semble peu probable de prouver qu'il y aurait une meilleure approche non-stationnaire pour toutes les applications. Toutefois, certaines approches se distinguent par leur simplicité, leur interprétabilité et leur élégance.

Notons que la stationnarité ou la non-stationnarité ne peuvent pas être testées avec les données en tant que tel (en réalisation unique). Cependant, lorsque l'objectif est la prédiction spatiale, l'utilisation des scores de prédiction peut permettre de choisir la

stratégie de modélisation (stationnaire ou non-stationnaire) la plus appropriée.

Les méthodes non-stationnaires proposées dans cette thèse reposent sur les approches par déformation d'espace et convolution. La première méthode est motivée par la quasi-inexistence dans la littérature géostatistique, d'une procédure d'inférence du modèle de déformation d'espace en réalisation unique. La seconde méthode se justifie d'une part, par l'absence d'une approche constructive permettant d'obtenir les expressions explicites de covariances non-stationnaires existantes. D'autre part, l'inférence statistique de ces classes de covariance non-stationnaire est un problème crucial. Les chapitres suivants sont consacrés à ces deux méthodes.

Chapitre 2

Modèle de déformation d'espace et inférence

"Le simple est toujours faux. Ce qui ne l'est pas est inutilisable." P. Valery.

Sommaire

2.1 Introduction **61**
2.2 Description du modèle **63**
2.3 Inférence statistique **65**
2.4 Prédiction **77**
2.5 Illustration **79**
2.6 Conclusion **88**

2.1 Introduction

Nous avons vu au chapitre 1 qu'une manière d'introduire la non-stationnarité d'ordre 2 est la déformation d'espace [Sampson and Guttorp, 1992]. Elle consiste à prendre une fonction aléatoire stationnaire, puis à déformer son espace d'indexation de façon régulière pour construire une fonction aléatoire non-stationnaire. Les méthodes d'inférence par maximum de vraisemblance et bayésiennes de cette approche ont été développées par Mardia and Goodall [1993], Smith [1996], Damian et al. [2001] et Schmidt and O'Hagan [2003]. Perrin and Meiring [1999], Perrin and Senoussi [2000], Perrin and Meiring [2003], Genton and Perrin [2004] et Porcu et al. [2010] ont établi certaines propriétés théoriques sur l'identifiabilité et la richesse de cette classe de modèles. Certaines adaptations ont été proposées récemment par Castro Morales et al. [2013], Bornn et al. [2012], Schmidt et al. [2011] et Vera et al. [2008, 2009].

Une limitation fondamentale à toutes les méthodes d'inférence citées précédemment réside dans le fait que la mise en œuvre nécessite de multiples réalisations indépendantes de la fonction aléatoire, ceci afin d'obtenir une matrice de covariance (ou variogramme) empirique. L'idée d'avoir plusieurs réalisations indépendantes du phénomène étudié paraît irréaliste, car il n'y a pas plusieurs mondes physiques parallèles. Dans la pratique, l'approche est possible quand des données temporelles sont recueillies en chaque emplacement. Cela permet d'obtenir des réplications qui peuvent être dépendantes. Cependant, il est préférable d'intégrer la dimension temporelle dans la modélisation plutôt que de tenter des réparations (différenciation ou détendancialisation) afin d'obtenir approximativement des réalisations indépendantes. Par ailleurs, de nombreuses applications géostatistiques impliquent seulement une mesure à chaque site ou de façon équivalente, une seule réalisation d'une fonction aléatoire. À notre connaissance, Anderes and Stein [2008], Anderes and Chatterjee [2009] sont les premiers et seuls auteurs jusqu'à présent à avoir traité de l'estimation du modèle de déformation d'espace dans le cas d'une seule réalisation. Ils présentent une méthodologie d'estimation de la déformation basée sur les transformations conformes dans le cadre gaussien. Un calcul approximatif de la vraisemblance est dérivé du partitionnement des observations denses en voisinages et en supposant l'indépendance de la fonction aléatoire sur les partitions. Cependant, cette approche n'a pas été appliquée sur des exemples de données réelles et nécessite des données denses.

Dans ce chapitre, nous optons pour l'idée originale de Sampson and Guttorp [1992], tout en s'affranchissant de l'hypothèse forte de réplication. De plus, nous prenons en compte d'autres manquements associés à cette approche que sont : la non-garantie pour la fonction de déformation d'être bijective et le problème numérique pour ajuster le modèle sur des jeux de données de taille modeste à grande. Pour ce faire, nous proposons une procédure d'estimation basée d'une part, sur l'inclusion de contraintes spatiales et d'autre part, sur l'utilisation d'un ensemble réduit de points représentatifs que nous appelons points supports. Ces derniers servent à la détermination de la déformation, plutôt que l'ensemble des points d'observation. La méthode proposée est libre de distribution (*distribution free*) et fournit une estimation non-paramétrique de la déformation via une approche par étapes. Tout d'abord, une matrice de dissimilarité est construite à partir d'un estimateur non-paramétrique à noyau du variogramme non-stationnaire. Ensuite, l'estimation de la déformation est effectuée aux points supports par positionnement multidimensionnel non-métrique pondéré. Enfin, la déformation est interpolée sur l'ensemble du domaine d'intérêt par des fonctions de base radiales de type spline plaque mince. L'approche fournit également une estimation rationnelle et automatique du variogramme stationnaire et isotrope dans l'espace déformé. Les potentialités de l'approche sont illustrées en premier lieu sur des données simulées.

Le chapitre est organisé comme suit : dans la section 2.2 le modèle de déformation spatiale est décrit à travers ses ingrédients de base et principales propriétés. Dans les sections 2.3 et 2.4, nous abordons le problème de l'estimation du modèle et comment les prédictions spatiales et simulations conditionnelles peuvent être effectuées. Deux jeux de données synthétiques sont utilisés dans la section 2.5 pour illustrer les performances de la nouvelle approche. Enfin, dans la section 2.6 une synthèse est faite.

2.2 Description du modèle

L'idée sous-jacente à l'approche par déformation d'espace est celle du plongement isométrique d'une fonction aléatoire non-stationnaire dans un nouvel espace de dimension égale ou supérieure où elle peut être facilement décrite et modélisée, c'est-à-dire où la stationnarité et l'isotropie sont plausibles [Schabenberger and Gotway, 2005a].

Soit $Z = \{Z(\mathbf{x}) : \mathbf{x} \in G \subseteq \mathbb{R}^p, p \geq 1\}$ une fonction aléatoire de moyenne constante définie sur un domaine d'intérêt G fixe et continu de l'espace euclidien $\mathbb{R}^p$ et reflétant le phénomène étudié. On considère que Z est gouvernée par le modèle suivant :

$$Z(\mathbf{x}) = Y(f(\mathbf{x})),\ \forall \mathbf{x} \in G, \tag{2.1}$$

qui peut s'écrire de manière équivalente :

$$Z(f^{-1}(\mathbf{u})) = Y(\mathbf{u}),\ \forall \mathbf{u} \in D, \tag{2.2}$$

où $Y = \{Y(\mathbf{u}) : \mathbf{u} \in D \subseteq \mathbb{R}^q, q \geq p\}$ représente une fonction aléatoire stationnaire et isotrope ; $f : G \to D$ représente une fonction déterministe, non-linéaire en général, bijective et continue de l'espace G vers l'espace D. En principe, on autorise $q \geq p$ bien que le plus souvent dans la littérature $q = p$. Dans la suite de l'exposé et sans perte de généralité, on suppose $q = p$.

Le modèle spécifié en (2.1) conduit à modéliser le variogramme de Z sous la forme :

$$\gamma(\mathbf{x}, \mathbf{y}) \equiv \frac{1}{2}\mathbb{V}(Z(\mathbf{x}) - Z(\mathbf{y})) = \frac{1}{2}\mathbb{E}(Z(\mathbf{x}) - Z(\mathbf{y}))^2 = \gamma_0(\|f(\mathbf{x}) - f(\mathbf{y})\|),\ \forall (\mathbf{x}, \mathbf{y}) \in G \times G, \tag{2.3}$$

où $\|.\|$ désigne la norme euclidienne dans $\mathbb{R}^p$; $\gamma_0(.)$ est le variogramme stationnaire et isotrope de Y. Il dépend seulement de la distance euclidienne entre les emplacements dans l'espace D.

Le modèle de second ordre obtenu en (2.3) conduit à un variogramme valide, c'est-à-dire conditionnellement défini négatif [Matheron, 1971]. Sa validité repose sur la proposition 2.1 inspirée de Meiring et al. [1997] et facile à établir (voir annexe C).

Proposition 2.1. *Si $\gamma_0 : D \times D \to \mathbb{R}^+$ est un variogramme valide, alors $\gamma_0 \circ (f \times f)$ est un variogramme valide sur $G \times G$, pour toute fonction $f : G \to D$.*

Ainsi, il est possible de s'appuyer sur un variogramme valide dans un autre espace, par le biais d'une fonction qui lie les deux espaces. Par conséquent, au lieu de travailler dans l'espace d'origine G, le variogramme de Z est défini par rapport à l'espace déformé D, où la stationnarité et l'isotropie sont supposées. Tout problème faisant intervenir la fonction aléatoire observée Z est transposé par la déformation $f(.)$ à Y, fonction aléatoire stationnaire et isotrope. Les techniques géostatistiques classiques telles que le krigeage et la simulation conditionnelle s'appliquent directement à cette dernière. Les résultats obtenus sur Y se transposent ensuite à Z par la déformation inverse $f^{-1}(.)$.

Remarque 2.2. L'approche par déformation d'espace autorise d'utiliser tous les modèles de variogramme valides dans le cadre stationnaire isotrope. Il tente ainsi de conserver la simplicité de la structure de variogramme stationnaire isotrope.

Comme décrit par Perrin and Meiring [1999], lorsque $\gamma_0(.)$ est non-décroissante, la déformation spatiale $f(.)$ opère concrètement comme suit : dans les régions de corrélation spatiale relativement faible, la transformation étire l'espace géographique G ; tandis que dans les régions de corrélation spatiale relativement élevée, elle le contracte de sorte qu'une structure de dépendance spatiale stationnaire et isotrope puisse être modélisée dans l'espace déformé D. En outre, remarquons que la déformation n'agit que sur la structure de dépendance spatiale : elle conserve l'espérance et la variance de la fonction aléatoire.

Il est important de noter que le modèle de déformation spatiale défini en (2.3) n'est identifiable qu'à une homothétie près pour $\gamma_0(.)$ et une similitude près pour $f(.)$. Ce résultat est basé sur la proposition 2.3, équivalente à celle établie par Perrin and Senoussi [2000] pour le modèle de déformation défini en terme de corrélation. La preuve est donnée en annexe C.

Proposition 2.3. *Si (γ_0, f) est une solution de (2.3), alors pour toute matrice carré régulière $\mathbf{A}$ et tout vecteur $\mathbf{b}$, $(\tilde{\gamma_0}, \tilde{f})$ avec $\tilde{f}(\mathbf{x}) = \mathbf{A}f(\mathbf{x}) + \mathbf{b}$ et $\tilde{\gamma_0}(\|\mathbf{u}\|) = \gamma_0(\|\mathbf{A}^{-1}\mathbf{u}\|)$ est aussi une solution.*

Remarque 2.4. La classe des modèles de déformation spatiale (2.3) englobe celles des modèles stationnaires isotropes et anisotropes (géométrique). En effet, si $f(.)$ est la fonction identité, on retrouve le modèle stationnaire isotrope et si $f(.)$ est une transformation linéaire, on retrouve le modèle stationnaire anisotrope (géométrique).

2.3 Inférence statistique

Soit $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$ le $(n \times 1)$ vecteur d'observations issu de la fonction aléatoire Z aux emplacements $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset G \subseteq \mathbb{R}^p$. Dans le modèle de second ordre défini en (2.3), les fonctions $f(.)$ et $\gamma_0(.)$ sont inconnues et donc à estimer. La méthodologie d'estimation que nous proposons comporte quatre étapes principales : tout d'abord, nous définissons un estimateur non-paramétrique à noyau du variogramme non-stationnaire $\gamma(.,.)$; celui-ci servant de mesure de dissimilarité entre les points de l'espace géographique G. Ensuite, nous construisons l'espace déformé D en utilisant une procédure de positionnement multidimensionnel non-métrique pondéré appliquée à la matrice de dissimilarité. Cette dernière est construite à partir de l'estimateur non-paramétrique du variogramme non-stationnaire. Puis, nous estimons la déformation $f(.)$ par interpolation entre une configuration de points de l'espace G et l'estimation de leur déformation dans l'espace D. Pour cela, nous utilisons la classe de fonctions de base radiales de type spline plaque mince. Enfin, l'estimation du variogramme stationnaire et isotrope $\gamma_0(.)$ est effectuée à partir du calcul du variogramme expérimental sur les données positionnées dans l'espace déformé D et en utilisant un mélange de modèles de variogramme de base, offrant ainsi une plus grande flexibilité pour s'adapter à la structure observée.

2.3.1 Estimateur à noyau du variogramme

Nous commençons par une estimation du variogramme non-stationnaire, en toute paire d'emplacements arbitraires. Pour cela, nous utilisons une moyenne locale pondérée des accroissements au carré de la variable régionalisée. Notre proposition est d'utiliser la nuée variographique $\gamma_{ij}^{\star} = \frac{1}{2}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$ qui est un estimateur sans biais de $\gamma(\mathbf{s}_i, \mathbf{s}_j)$ [Chilès and Delfiner, 2012] comme donnée d'entrée pour l'estimateur non-paramétrique du variogramme non-stationnaire. Ce faisant, un estimateur empirique intuitif du variogramme non-stationnaire entre deux emplacements quelconques est donné par l'estimateur non-paramétrique à noyau défini comme suit :

$$\widehat{\gamma}(\mathbf{x}, \mathbf{y}; \lambda) = \frac{\sum_{i,j=1}^{n} K_\lambda\left((\mathbf{x}, \mathbf{y}), (\mathbf{s}_i, \mathbf{s}_j)\right)\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2}{2\sum_{i,j=1}^{n} K_\lambda\left((\mathbf{x}, \mathbf{y}), (\mathbf{s}_i, \mathbf{s}_j)\right)} \mathbb{1}_{\{\mathbf{x} \neq \mathbf{y}\}}, \quad \forall (\mathbf{x}, \mathbf{y}) \in G \times G \quad (2.4)$$

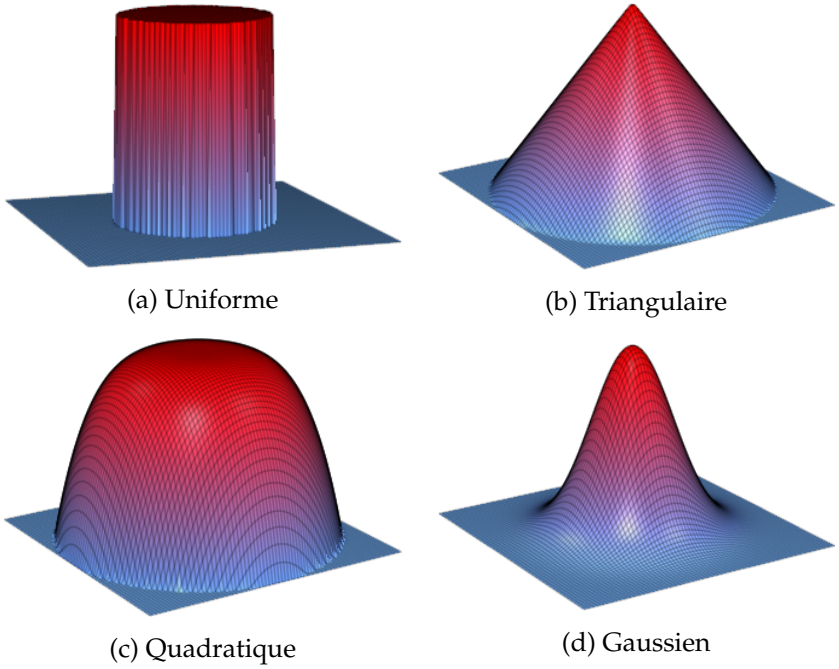
où $K_\lambda(.,.)$ est défini par : $K_\lambda\left((\mathbf{x}, \mathbf{y}), (\mathbf{s}_i, \mathbf{s}_j)\right) = K(\mathbf{x}, \mathbf{s}_i; \lambda) K(\mathbf{y}, \mathbf{s}_j; \lambda)$, $K(., .; \lambda)$ étant un noyau symétrique sur $\mathbb{R}^p \times \mathbb{R}^p$ et de largeur de fenêtre $\lambda \in \mathbb{R}_+^*$.

L'expression (2.4) mesure la dissimilarité spatiale entre deux implantations arbitraires de l'espace géographique G au moyen de la variable régionalisée. Le but du noyau $K(.,.;\lambda)$ est de pondérer les observations par rapport à un point de référence de telle sorte que les observations proches de la référence reçoivent plus de poids, tandis que celles éloignées en reçoivent moins. Le dénominateur en (2.4) est un facteur de normalisation qui assure que $\widehat{\gamma}(\mathbf{x},\mathbf{y};\lambda)$ soit sans biais lorsque l'espérance de la différence au carré entre les observations est spatialement constante. Si plusieurs réalisations sont disponibles, nous pouvons simplement prendre la moyenne de (2.4) sur les différentes réalisations comme estimateur.

S'agissant du noyau $K(.,.;\lambda)$, son choix est moins crucial que celui de son paramètre d'échelle λ ; un choix raisonnable de $K(.,.;\lambda)$ conduit généralement à des résultats raisonnables [Wand and Jones, 1995]. Il existe de nombreux noyaux dont les plus classiques sont :

1. *noyau uniforme* $K(\mathbf{x},\mathbf{y};\lambda) \propto \mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|\leq\lambda}$;
2. *noyau triangulaire* $K(\mathbf{x},\mathbf{y};\lambda) \propto (\lambda - \|\mathbf{x}-\mathbf{y}\|)\mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|\leq\lambda}$;
3. *noyau quadratique* $K(\mathbf{x},\mathbf{y};\lambda) \propto (\lambda^2 - \|\mathbf{x}-\mathbf{y}\|^2)\mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|\leq\lambda}$;
4. *noyau gaussien* $K(\mathbf{x},\mathbf{y};\lambda) \propto \exp(-\frac{1}{2\lambda^2}\|\mathbf{x}-\mathbf{y}\|^2)$.

La figure 2.1 présente différents noyaux et montre comment ils répartissent le poids d'un point cible sur une région. Dans ce travail, notre choix s'est porté sur le noyau quadratique (noyau d'Epanechnikov) qui est un noyau stationnaire et isotrope, à support compact, présentant des propriétés d'optimalité dans le cadre de l'estimation de la densité [Wand and Jones, 1995]. En effet, le coût de calcul de (2.4) est considérablement réduit en utilisant un noyau à support compact, car il réduit le nombre de termes à calculer. La sélection du paramètre de largeur de fenêtre est abordée à la section 2.3.4.



(a) Uniforme (b) Triangulaire

(c) Quadratique (d) Gaussien

FIGURE 2.1 – Modèle de déformation d'espace : exemples de noyaux.

2.3.2 Construction de l'espace déformé

La construction de l'espace déformé peut être vue comme un problème de réallocation d'une configuration de points dans un nouvel espace euclidien de dimension donnée. L'approche de positionnement multidimensionnel non-métrique (en anglais NMDS, pour *Non-metric MultiDimensional Scaling*) [Kruskal, 1964a] fournit une solution à ce problème. Le but étant de trouver une représentation de points dans un espace de dimension fixée telle que l'ordre des distances euclidiennes entre les points correspond autant que possible à l'ordre des dissimilarités observées entre points.

Points supports

Jusqu'à présent, l'ajustement du modèle de déformation d'espace basé sur la procédure NMDS telle que présentée par Sampson and Guttorp [1992] est un problème numérique difficile dont la dimension est proportionnelle au nombre d'observations. Un échantillon de taille n requiert une matrice de dissimilarité $(n \times n)$ à stocker et une matrice de coordonnées $(n \times q)$ à estimer. La recherche de l'espace déformé sur la base de cette matrice de dissimilarité nécessite un temps de calcul considérable même lorsque n est relativement modeste. Pour réduire la charge de calcul, nous pouvons éviter de transformer directement tous les points de données. En effet, les dissimilarités spatiales calculées pour des paires de points très proches peuvent être redondantes, car fortement corrélées. L'idée consiste alors à obtenir l'espace déformé en utilisant seulement un ensemble réduit de $m \ll n$ points

représentatifs dénommés points supports de l'espace géographique G. Ce sont ces derniers qui seront transformés par la procédure NMDS. L'interpolation des points supports de l'espace G et les estimations de leur déformation dans l'espace D produit une estimation $\widehat{f}(.)$ de la fonction de déformation. Ensuite, l'emplacement de tous les points de données dans l'espace déformé est obtenu via $\widehat{f}(.)$. Les points supports peuvent être choisis comme les nœuds d'une grille couvrant le domaine d'intérêt G ou comme un sous-ensemble des points de données. Ils permettent ainsi de réduire le temps de calcul et de robustifier les résultats de la procédure NMDS. En effet, la densité d'échantillonnage qui peut varier à travers le domaine, peut être représentée par une distribution non-uniforme des points supports.

Matrice de dissimilarité

Considérons un ensemble de m points supports (en dimension p) $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]^T$ de l'espace géographique G. L'estimateur $\widehat{\gamma}(.,.;\lambda)$ défini en 2.4 et calculé en chaque paire de points supports permet de construire une matrice de dissimilarité symétrique $(m \times m)$ $\widehat{\mathbf{\Gamma}}_\lambda = [\widehat{\gamma}_{ij}(\lambda)]$, avec $\widehat{\gamma}_{ij}(\lambda) = \widehat{\gamma}(\mathbf{x}_i, \mathbf{x}_j; \lambda)$. La spécification d'une matrice de dissimilarité est requise pour la procédure NMDS. Cette dernière est généralement appliquée en l'absence de coordonnées euclidiennes des points ayant produits les dissimilarités. Dans notre contexte, les points de l'espace d'origine G sont déjà repérés par leurs coordonnées euclidiennes. D'où la nécessité de prendre en compte ce contexte spatial pour assurer la cohérence dans le déplacement des points à transformer.

De plus, l'estimateur $\widehat{\gamma}(.,.;\lambda)$ reflète la dissimilarité spatiale localement, c'est-à-dire pour les paires proches. En effet, cet estimateur peut être fiable pour les courtes et moyennes distances, comme dans le contexte stationnaire, mais très imprécis quoique bien défini pour les grandes distances. Il est alors souhaitable de pénaliser l'importance accordée aux grandes distances par rapport aux courtes distances dans la recherche de la déformation. De ce fait, étant donnée la matrice de distance associée aux points supports $\mathbf{X}$, nous construisons une matrice de dissimilarité composite $(m \times m)$ $\mathbf{\Delta}_{(\lambda,\omega)} = [\delta_{ij}(\lambda, \omega)]$ définie comme suit :

$$\mathbf{\Delta}_{(\lambda,\omega)} = \omega \tilde{\mathbf{\Gamma}}_\lambda + (1 - \omega)\tilde{\mathbf{D}}, \tag{2.5}$$

$$\text{où } \tilde{\mathbf{\Gamma}}_\lambda = \left[\frac{\widehat{\gamma}_{ij}(\lambda) - \min(\widehat{\gamma}_{ij}(\lambda))}{\max(\widehat{\gamma}_{ij}(\lambda)) - \min(\widehat{\gamma}_{ij}(\lambda))}\right],\ \tilde{\boldsymbol{D}} = \left[\frac{d_{ij} - \min(d_{ij})}{\max(d_{ij}) - \min(d_{ij})}\right],\ \omega \in [0, 1].$$

La matrice de dissimilarité composite $\mathbf{\Delta}_{(\lambda,\omega)}$ est une combinaison linéaire d'une matrice de dissimilarité et d'une matrice de distance normalisées par l'étendue de leurs valeurs. L'idée de construire une mesure de dissimilarité spatiale hybride qui tient compte à la fois des dissimilarités observées dans la variable régionalisée et des proximités spatiales permet de réduire le risque que la fonction de déformation se replie, c'est-à-dire qu'elle

soit non bijective. Ainsi, le paramètre de mélange ω contrôle le non-repliement ; sa valeur $\omega = 0$, conduisant au cadre stationnaire. Le réglage du paramètre ω est abordé à la section 2.3.4.

Modèle NMDS

Étant donnée la matrice symétrique $(m \times m)$ $\mathbf{\Delta}_{(\lambda,\omega)} = [\delta_{ij}(\lambda,\omega)]$ des dissimilarités entre l'ensemble des m points supports (en dimension p) $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]^T$, l'objectif est de représenter $\mathbf{X}$ comme une configuration de m points (en dimension $q = p$) $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_m]^T$ telle que les relations suivantes soient satisfaites autant que possible :

$$\phi(\delta_{ij}(\lambda,\omega)) \approx h_{ij}(\mathbf{U}), \tag{2.6}$$

où $h_{ij}(\mathbf{U}) = \|\mathbf{u}_i - \mathbf{u}_j\|$; $\phi(.)$ est une fonction monotone qui conserve l'ordre des dissimilarités : $\delta_{ij}(\lambda,\omega) \leq \delta_{kl}(\lambda,\omega) \Longrightarrow \phi(\delta_{ij}(\lambda,\omega)) \leq \phi(\delta_{kl}(\lambda,\omega))$.

Autrement dit, étant données les $\frac{1}{2}m(m-1)$ dissimilarités (ignorant la diagonale de $\mathbf{\Delta}_{(\lambda,\omega)}$ ne contenant que des zéros), on recherche une configuration de points supports $\mathbf{U}$ dans un espace de dimension fixée telle que l'ordre des distances de la configuration soit en accord avec l'ordre des dissimilarités. La configuration des points supports $\mathbf{U}$ dans l'espace déformé D est déterminée de telle manière que les distances entre points dans l'espace D minimisent la fonction de perte (*stress*) suivante :

$$S_{(\lambda,\omega)}(\mathbf{U}) = \min_{\phi} \left[\sum_{i<j} \frac{p_{ij}(\lambda)[h_{ij}(\mathbf{U}) - \phi(\delta_{ij}(\lambda,\omega))]^2}{\sum_{i<j} p_{ij}(\lambda) h_{ij}^2(\mathbf{U})} \right]^{\frac{1}{2}}, \tag{2.7}$$

où la minimisation est faite sur l'ensemble des fonctions monotones croissantes $\phi(.)$; $p_{ij}(\lambda) = \sum_{k,l=1}^{n} K_\lambda \left((\mathbf{x}_i, \mathbf{x}_j), (\mathbf{s}_k, \mathbf{s}_l) \right) / \|\mathbf{x}_i - \mathbf{x}_j\|$; $K_\lambda(.,.)$ est le noyau utilisé dans le calcul de l'estimateur à noyau du variogramme non-stationnaire en (2.4).

$S_{(\lambda,\omega)}(\mathbf{U})$ dénote le *stress* de la configuration fixée $\mathbf{U}$. L'estimateur $\{\phi(\delta_{ij}(\lambda,\omega))\}$ représente la régression isotonique par moindres carrés pondérés des $\{h_{ij}(\mathbf{U})\}$ sur les $\{\delta_{ij}(\lambda,\omega)\}$. Les $\{p_{ij}(\lambda)\}$ sont des poids positifs utilisés pour pondérer la contribution des éléments de la matrice de dissimilarité $\mathbf{\Delta}_{(\lambda,\omega)}$ dans le calcul et la minimisation du *stress*. Nous introduisons un tel système de poids pour prendre en compte le fait que l'information locale disponible dans le voisinage de points supports peut fluctuer d'une paire de points supports à l'autre. Ainsi, les dissimilarités les plus fiables ont plus d'impact dans la fonction de perte que les moins fiables. Ces poids permettent également de mettre l'accent sur les dissimilarités aux courtes distances. Ces dernières ayant ainsi plus de poids dans la fonction de perte que celles aux grandes distances.

Le *stress* défini en (2.7) fournit une mesure du degré d'adéquation entre l'ordre des distances et celui des dissimilarités à transformer. Il est invariant par translation, rotation ou mise à l'échelle (dilatation ou contraction uniforme) de la configuration. Ainsi, la solution NMDS n'est connue qu'à une de ces transformations près.

Le problème défini en (2.7) peut être résolu par l'algorithme itératif de Shepard-Kruskal [Kruskal, 1964a]. Globalement, la méthode est la suivante : on part d'une configuration initiale $\mathbf{U}^{(0)}$, celle des points supports $\mathbf{X}$. On cherche alors les $\phi(\delta_{ij}(\lambda,\omega))$ telle que $\sum_{i<j} p_{ij}(\lambda)[h_{ij}(\mathbf{U}^{(0)}) - \phi(\delta_{ij}(\lambda,\omega))]^2$ soit minimum. Ce problème admet une solution unique : la régression isotonique [Kruskal, 1964b]. La valeur du *stress* est ainsi déduite. On modifie ensuite la configuration au moyen de petits déplacements des points selon une méthode de gradient pour diminuer le *stress*. On repasse ensuite à la phase de régression isotonique, ainsi de suite jusqu'à la convergence. Ces deux étapes sont décrites ci-dessous :

Régression isotonique

La régression isotonique est la première étape dans l'ajustement du modèle NMDS. Elle désigne un ensemble de méthodes non-paramétriques répondant spécifiquement au problème de l'ajustement d'une fonction monotone aux données. Elle ne s'appuie pas sur des estimateurs conçus pour un contexte habituel de régression. Étant donnée une configuration $\mathbf{U}^{(r)}$, on souhaite résoudre le problème de minimisation suivant :

$$\arg\min_{\phi} \sum_{i<j} p_{ij}(\lambda)[h_{ij}(\mathbf{U}^{(r)}) - \phi(\delta_{ij}(\lambda,\omega))]^2, \tag{2.8}$$

sous la contrainte de monotonie : $\delta_{ij}(\lambda,\omega) \leq \delta_{kl}(\lambda,\omega) \Longrightarrow \phi(\delta_{ij}(\lambda,\omega)) \leq \phi(\delta_{kl}(\lambda,\omega))$.

Pour plus de commodité, renommons les dissimilarités $\{\delta_{ij}(\lambda,\omega)\}$ par $\{\delta_t(\lambda,\omega), t = 1,\ldots,T\}$ avec $T = \frac{1}{2}m(m-1)$ et supposons qu'elles soient ordonnées dans l'ordre strictement croissant (*no ties*) $\delta_1(\lambda,\omega) < \delta_2(\lambda,\omega) <, \ldots, \delta_T(\lambda,\omega)$. Renommons aussi les distances $\{h_{ij}(\mathbf{U}^{(r)})\}$ comme $\{h_t(\mathbf{U}^{(r)}), t = 1,\ldots,T\}$ où $h_t(\mathbf{U}^{(r)})$ correspond à la dissimilarité $\delta_t(\lambda,\omega)$. De même pour les poids $\{p_{ij}(\lambda)\}$ qui deviennent $\{p_t(\lambda)\}$.

La détermination des $\{\phi(\delta_t(\lambda,\omega))\}$ est habituellement obtenue par l'algorithme PAV(*Pool Adjacent Violators*)[Kruskal, 1964b]. Cet algorithme décrit en 2.3.1 repose sur le principe consistant à "amalgamer" les blocs violant la monotonie. L'algorithme divise les dissimilarités en blocs, et à chaque étape de l'algorithme un de ces blocs devient actif. Ici, un bloc est considéré comme un ensemble de dissimilarités consécutives qui doivent toutes être égales pour assurer la monotonie. Une solution consiste alors à moyenner les valeurs à l'intérieur du bloc actif.

Algorithme 2.3.1 *Pool Adjacent Violators*

1. si la séquence initiale vérifie $h_1(\mathbf{U}^{(r)}) \leq \ldots \leq h_T(\mathbf{U}^{(r)})$, alors elle est aussi la séquence finale et pour $t = 1, \ldots, T$, prendre $\hat{\phi}(\delta_t(\lambda, \omega)) = h_t(\mathbf{U}^{(r)})$;
2. sinon,
   — considérer n'importe quelle paire de valeurs successives violant la contrainte de monotonie (*adjacent violators*), c'est-à-dire sélectionner un indice t tel que : $h_t(\mathbf{U}^{(r)}) > h_{t+1}(\mathbf{U}^{(r)})$;
   — regrouper (*to pool*) dans la séquence précédente, ces deux valeurs en un seul bloc. Autrement dit, substituer les couples $(h_t(\mathbf{U}^{(r)}), p_t(\lambda))$ et $(h_{t+1}(\mathbf{U}^{(r)}), p_{t+1}(\lambda))$ par le couple moyen :

$$\left(\frac{p_t(\lambda) h_t(\mathbf{U}^{(r)}) + p_{t+1}(\lambda) h_{t+1}(\mathbf{U}^{(r)})}{p_t(\lambda) + p_{t+1}(\lambda)}, p_t(\lambda) + p_{t+1}(\lambda) \right) ;$$

3. itérer les deux sous-étapes précédentes en considérant la séquence actualisée jusqu'à obtenir une séquence isotonique.
4. estimer $\hat{\phi}(\delta_t(\lambda, \omega))$ pour $t = 1, \ldots, T$ par la valeur finale associée au bloc dont elle fait partie.

Il est important de noter que dans cet algorithme, les dissimilarités n'interviennent que par leur rang. Ainsi, la mise en œuvre de l'algorithme suppose de les ordonner au préalable. Lorsque $\delta_{ij}(\lambda, \omega) = \delta_{kl}(\lambda, \omega)$, l'approche d'usage (*treatment of ties*) consiste à relâcher la contrainte de monotonie, ceci en n'imposant aucune contrainte d'ordre entre $\phi(\delta_{ij}(\lambda, \omega))$ et $\phi(\delta_{kl}(\lambda, \omega))$.

Méthode du gradient

Connaissant les $\hat{\phi}(\delta_{ij}(\lambda, \omega))$, la seconde étape de la procédure NMDS est la recherche d'une configuration de points minimisant le *stress*. Kruskal [1964a] propose d'en rechercher par une application de la méthode du gradient. D'autres méthodes peuvent être retenues. La méthode du gradient déplace la configuration dans une direction déterminée par les dérivées partielles de $S_{(\lambda,\omega)}$ par rapport à $\mathbf{U}$. Ainsi, étant donnée la configuration $\mathbf{U}^{(\beta)}$ à la β-ième itération de la méthode, une configuration mise à jour à l'itération suivante est donnée par :

$$\mathbf{U}^{(\beta+1)} = \mathbf{U}^{(\beta)} - \eta_{\beta+1} \mathbf{G}^{(\beta)}, \tag{2.9}$$

où $\eta_{\beta+1}$ est la taille du pas à la $(\beta + 1)$-ième itération et

$$\mathbf{G}^{(\beta)} = \frac{\partial S_{(\lambda,\omega)}}{\partial \mathbf{U}}(\mathbf{U}^{(\beta)}) / \left| \frac{\partial S_{(\lambda,\omega)}}{\partial \mathbf{U}}(\mathbf{U}^{(\beta)}) \right| \tag{2.10}$$

est le gradient normalisé calculé en $\mathbf{U}^{(\beta)}$. La formule explicite de $\mathbf{G}^{(\beta)}$ est obtenue en utilisant l'expression des dérivées partielles s'écrivant comme suit [Kruskal, 1964b] :

$$\frac{\partial S_{(\lambda,\omega)}}{\partial \mathbf{u}_{kl}} = S_{(\lambda,\omega)} \sum_{i<j} (\mathbb{1}_{\{i=k\}} - \mathbb{1}_{\{j=k\}}) p_{ij}(\lambda) \left[\frac{h_{ij} - \hat{h}_{ij}}{S^{\star}_{(\lambda,\omega)}} - \frac{h_{ij}}{T^{\star}_{\lambda}} \right] \frac{|\mathbf{u}_{il} - \mathbf{u}_{jl}|^{p-1}}{h_{ij}^{p-1}} \text{signe}(\mathbf{u}_{il} - \mathbf{u}_{jl}), \tag{2.11}$$

où $S^{\star}_{(\lambda,\omega)} = \sum_{i<j} p_{ij}(\lambda)[\phi(\delta_{ij}(\lambda,\omega)) - h_{ij}(\mathbf{U})]^2$; $T^{\star}_{\lambda} = \sum_{i<j} p_{ij}(\lambda) h^2_{ij}(\mathbf{U})$; $\hat{h}_{ij} = \hat{\phi}(\delta_{ij}(\lambda,\omega))$.

Dans cette procédure basée sur le gradient, il n'existe aucune garantie que l'algorithme converge vers un minimum global. En effet, il peut arriver que l'algorithme converge vers un minimum local. Ainsi, il est d'usage d'essayer différentes configurations initiales pour vérifier la convergence de l'algorithme. Si la même solution est obtenue à partir de l'application répétée de l'algorithme, alors la solution commune est probablement un minimum global ; sinon on peut prendre comme solution celle donnant le plus petit *stress*. La taille du pas peut être changée à chaque itération afin d'accélérer l'algorithme.

La procédure NMDS est un double processus d'optimisation. D'abord, la transformation monotone optimale des dissimilarités doit être trouvée (régression isotonique). Ensuite, les points doivent être disposés de façon optimale, de sorte que le rang des distances inter-points corresponde autant que possible à celui des dissimilarités (méthode de gradient). La solution finale sera obtenue en utilisant en alternance les deux algorithmes précédents (régression isotonique, méthode du gradient). On s'arrêtera soit lorsqu'il n'est plus possible d'améliorer la solution, soit lorsque le *stress* sera suffisamment petit.

L'expérience de Kruskal avec différents types de données réelles et simulées l'a conduit à évaluer l'ajustement global de toute solution NMDS par différents niveaux de valeurs du *stress* tels que présentés dans le tableau 2.1 [Kruskal, 1964a]. L'évaluation donnée par le tableau 2.1 doit être considérée comme une indication de la qualité d'ajustement d'une solution NMDS. Pour plus de détails sur l'approche NMDS, voir par exemple : Borg et al. [2012], Borg and Groenen [2005] et Cox and Cox [2000].

Stress	Qualité d'ajustement
0,20	Faible
0,10	Passable
0,05	Bien
0,025	Excellent
0,00	Parfait

TABLEAU 2.1 – Modèle de déformation d'espace : évaluation du *stress*.

En somme, la construction de l'espace déformé se résume en l'algorithme itératif 2.3.2 suivant :

Algorithme 2.3.2 Construction de l'espace déformé

1. définir une configuration de points supports $\mathbf{X}$;
2. calculer les dissimilarités $\{\delta_{ij}(\lambda, \omega)\}$ entre toutes les paires de points de $\mathbf{X}$;
3. classer les $\frac{1}{2}m(m-1)$ dissimilarités $\{\delta_{ij}(\lambda, \omega)\}$ par ordre croissant ;
4. choisir comme configuration initiale $\mathbf{U}^{(0)}$ celle définie par les coordonnées des points supports $\mathbf{X}$;
5. calculer l'ensemble des distances $\{h_{ij}(\mathbf{U}^{(0)})\}$ entre toutes les paires de points dans la configuration initiale ;
6. effectuer une régression isotonique par moindres carrés pondérés des $\{h_{ij}(\mathbf{U}^{(0)})\}$ sur $\{\delta_{ij}(\lambda, \omega)\}$ et prendre les valeurs ajustées $\{\hat{h}_{ij}(\mathbf{U}^{(0)})\}$. Calculer la valeur initiale du *stress* ;
7. réviser la configuration de façon à produire une configuration $\mathbf{U}^{(1)}$ qui fournit un nouvel ensemble de distances $\{h_{ij}(\mathbf{U}^{(1)})\}$ qui soit le plus étroitement lié aux disparités $\{\hat{h}_{ij}(\mathbf{U}^{(0)})\}$ générées à l'étape 6 ; ceci peut se faire via un algorithme itératif de descente de gradient qui va minimiser le *stress* ;
8. répéter les étapes 6 et 7 jusqu'à ce que la configuration courante produise une valeur minimale du *stress*, de sorte que d'autres améliorations dans le *stress* ne peuvent avoir lieu en reconfigurant encore les points.

2.3.3 Estimation des paramètres fonctionnels

On s'intéresse à l'estimation des paramètres fonctionnels $f(.)$ et $\gamma_0(.)$ qui gouvernent le modèle de second ordre (2.3). Partant de la déformation estimée aux points supports, la transformation des points supports est alors étendue à une fonction régulière $\widehat{f}(.)$ de l'espace entier G vers l'espace D, en utilisant une représentation basée sur les fonctions de base radiales de type spline plaque mince. En utilisant la déformation estimée $\widehat{f}(.)$, un modèle de variogramme stationnaire et isotrope $\gamma_0(.)$ est ajusté au variogramme expérimental dans l'espace déformé D. Enfin, la fonction de déformation estimée et le modèle de variogramme stationnaire et isotrope estimé sont utilisés pour calculer la structure de dépendance spatiale entre deux emplacements de l'espace géographique G.

Fonction de déformation

On souhaiterait avoir l'image de tout point $\mathbf{x} \in G$ dans l'espace déformé D, de même que dans G l'image réciproque de tout point $\mathbf{u} \in D$. En particulier, une représentation de tous les points de données dans l'espace D est requise.

Étant donnée la configuration des points supports $\mathbf{X}$ et l'estimation de leur déformation $\mathbf{U}$, on cherche une fonction q-variée $f : \mathbb{R}^p \to \mathbb{R}^q$ qui satisfait la condition d'interpolation $f(\mathbf{X}) = \mathbf{U}$. Bien qu'un grand nombre d'options existent, nous suivons Sampson and Guttorp [1992] dans l'utilisation de la classe des fonctions de base radiales de type spline plaque mince. Les splines plaques minces constituent l'une des approches les plus couramment utilisées pour l'interpolation et sont également appréciées en raison de la solution analytique de leurs paramètres [Lloyd, 2011]. Elles permettent d'obtenir une surface lissée la moins déformée possible passant par tous les points de contrôles. En dimension 1, ce sont les splines cubiques. Spécifiquement, l'estimateur spline plaque mince de $f(.)$ prend la forme suivante [Dryden and Mardia, 1998] :

$$\widehat{f}(\mathbf{x}) = (\widehat{f}_1(\mathbf{x}), \ldots, \widehat{f}_q(\mathbf{x}))^T = \mathbf{c} + \mathbf{A}\mathbf{x} + \mathbf{V}^T\boldsymbol{\sigma}(\mathbf{x}), \tag{2.12}$$

avec les matrices $\mathbf{c}(q \times 1), \mathbf{A}(q \times p), \mathbf{V}(m \times q)$ et $\boldsymbol{\sigma}(\mathbf{x}) = (\sigma(\mathbf{x} - \mathbf{x}_1), \ldots, \sigma(\mathbf{x} - \mathbf{x}_m))^T$, où $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ sont les points supports vus comme les centres de la fonction de base radiale $\sigma(\mathbf{h}) = \|\mathbf{h}\|^2 \log(\|\mathbf{h}\|)\mathbb{1}_{\|\mathbf{h}\|>0}$.

Les paramètres $\{\mathbf{c}, \mathbf{A}, \mathbf{V}\}$ dans l'expression (2.12) sont inconnus et doivent donc être estimés. Ceci se fait par la résolution du système d'équations $\widehat{f}(\mathbf{X}) = \mathbf{U}$ sous les contraintes $\mathbf{1}^T\mathbf{V} = 0$ et $\mathbf{X}^T\mathbf{V} = 0$ [Dryden and Mardia, 1998]. Ainsi, le système d'équations pour déterminer $\widehat{f}(.)$ s'écrit :

$$\boldsymbol{\Sigma} \begin{bmatrix} \mathbf{V} \\ \mathbf{c}^T \\ \mathbf{A}^T \end{bmatrix} = \begin{bmatrix} \mathbf{U} \\ 0 \\ 0 \end{bmatrix}, \tag{2.13}$$

où $\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{S} & \mathbf{1} & \mathbf{X} \\ \mathbf{1}^T & 0 & 0 \\ \mathbf{X}^T & 0 & 0 \end{bmatrix}$, avec $\mathbf{S}$ la matrice $(m \times m)$ de terme général $\mathbf{S}_{ij} = \sigma(\mathbf{x}_i - \mathbf{x}_j)$.

La matrice $\boldsymbol{\Sigma}$ est symétrique et définie positive sous réserve de l'existence de l'inverse de $\mathbf{S}$. Dans ce cas, son inverse existe et :

$$\begin{bmatrix} \mathbf{V} \\ \mathbf{c}^T \\ \mathbf{A}^T \end{bmatrix} = \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{U} \\ 0 \\ 0 \end{bmatrix}. \tag{2.14}$$

Remarque 2.5. L'interpolation par splines plaques minces est équivalente formellement au krigeage intrinsèque (FAI-k) basé sur la covariance généralisée spline : $K(\mathbf{h}) = \|\mathbf{h}\|^2 \log(\|\mathbf{h}\|)$ [Chilès and Delfiner, 2012]. Contrairement à Sampson and Guttorp [1992] qui travaillent avec une spline de lissage (leurs permettant ainsi de contrôler le repliement de la déformation), nous travaillons ici avec une spline d'interpolation.

Variogramme transformé

La performance de l'approche par déformation spatiale dépend aussi de sa capacité à bien ajuster un modèle de variogramme stationnaire et isotrope dans l'espace déformé. Ainsi, la modélisation et l'estimation de la structure de dépendance spatiale qui caractérise le comportement de la variable d'intérêt est une étape clé. Jusqu'à présent, et comme l'ont souligné Meiring et al. [1997], Perrin and Monestiez [1998] et Iovleff and Perrin [2004], aucun choix rationnel n'a encore été développé pour le modèle de variogramme stationnaire et isotrope dans le modèle de déformation spatiale. Ce choix est souvent effectué par inspection visuelle.

Ici, l'estimation de $\gamma_0(.)$ dans l'espace déformé D est effectuée en utilisant un mélange de modèles de base (effet de pépite, gaussien, exponentiel, sphérique,...), ce qui offre une plus grande flexibilité pour s'adapter à la structure observée. Nous avons :

$$\gamma_0(\|\mathbf{h}\|) = \sum_{k=1}^{K} c_k \gamma_k(\|\mathbf{h}\|), \quad \forall \mathbf{h} \in \mathbb{R}^p, \tag{2.15}$$

où $\gamma_1(.), \ldots, \gamma_K(.)$ sont des structures stationnaires, isotropes et valides dans $\mathbb{R}^p$; $c_1, \ldots, c_K$ sont des nombres réels positifs.

Pour estimer $\gamma_0(.)$, le modèle de variogramme stationnaire et isotrope dans l'espace déformé D, nous avons recours à une méthode robuste récemment développée par Desassis and Renard [2012]. Elle consiste à trouver automatiquement un modèle pour ajuster le variogramme expérimental. À partir d'une combinaison linéaire de certaines structures de base autorisées, un algorithme numérique itératif est utilisé pour estimer un modèle parcimonieux qui minimise une distance pondérée entre le modèle et le variogramme expérimental.

Plus précisément, étant donné le variogramme expérimental $\{\hat{\gamma}_0(\|\mathbf{h}_j\|), \mathbf{h}_j \in \mathbb{R}^p, j = 1, \ldots, J\}$ et une famille de structures de base paramétriques (normalisées) $\{\gamma_1^{(\theta_1)}, \ldots, \gamma_K^{(\theta_K)}\}$, l'objectif est de trouver une combinaison linéaire $\gamma_0^{(\mathbf{\Psi})}(\|\mathbf{h}\|) = \sum_{k=1}^{K} c_k \gamma_k^{(\theta_k)}(\|\mathbf{h}\|)$ à coefficients positifs tel que :

$$S(\mathbf{\Psi}) = \frac{1}{2} \sum_{j=1}^{J} \omega_j \Big(\gamma_0^{(\mathbf{\Psi})}(\|\mathbf{h}_j\|) - \hat{\gamma}_0(\|\mathbf{h}_j\|) \Big)^2, \tag{2.16}$$

soit minimal pour le vecteur de paramètres $\Psi = (\theta_1, \ldots, \theta_K, c_1, \ldots, c_K)$; avec $\{\omega_j, j = 1, \ldots, J\}$ un ensemble de poids, par exemple $\omega_j = N_j / \|\mathbf{h}_j\|$, où N_j est le nombre de paires utilisées dans le calcul de $\hat{\gamma}_0(\|\mathbf{h}_j\|)$.

2.3.4 Réglages des hyper-paramètres

L'approche proposée repose principalement sur les deux hyper-paramètres (λ, ω) entrant dans le calcul de la matrice de dissimilarité composite $\mathbf{\Delta}_{(\lambda,\omega)}$ définie en (2.5). La démarche adoptée pour choisir ces hyper-paramètres est guidée par les données (*data driven method*). Nous développons et combinons deux critères de validation croisée pour sélectionner leur valeur optimale.

La première forme de validation croisée est basée sur $\widehat{\gamma}(.,.;\lambda)$. Elle consiste à calculer le critère de l'erreur de validation croisée du variogramme sur une grille de valeurs du paramètre λ. Ceci permet d'extraire un ensemble réduit de valeurs aux scores les plus faibles :

$$CV(\lambda) = \frac{1}{n^2} \sum_{i,j=1}^{n} \left(\widehat{\gamma}_{(-i,-j)}(\mathbf{s}_i, \mathbf{s}_j; \lambda) - \gamma_{ij}^{\star} \right)^2, \tag{2.17}$$

où $\widehat{\gamma}_{(-i,-j)}(\mathbf{s}_i, \mathbf{s}_j; \lambda)$ est l'estimateur du variogramme non-stationnaire aux emplacements $\mathbf{s}_i$ et $\mathbf{s}_j$ calculé selon la formule (2.4) avec toutes les observations sauf $\{Z(\mathbf{s}_i), Z(\mathbf{s}_j)\}$; $\gamma_{ij}^{\star} = \frac{1}{2}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$ est un estimateur sans biais de $\gamma(\mathbf{s}_i, \mathbf{s}_j)$.

Puisque l'estimation de la structure de dépendance spatiale est rarement un but en soi, mais une étape intermédiaire avant le krigeage, il paraît approprié de choisir les hyper-paramètres qui minimisent l'erreur quadratique moyenne (MSE, *Mean Square Error*) en validation croisée. Plus explicitement, pour chaque paire de valeurs des hyper-paramètres, on calcule le score de validation croisée suivant :

$$MSE(\lambda, \omega) = \frac{1}{n} \sum_{i=1}^{n} \left(Z(\mathbf{s}_i) - \widehat{Z}_{-i}(\mathbf{s}_i; \lambda, \omega) \right)^2, \tag{2.18}$$

où $\widehat{Z}_{-i}(\mathbf{s}_i; \lambda, \omega)$ représente le krigeage à l'emplacement $\mathbf{s}_i$ en utilisant toutes les observations exceptées $\{Z(\mathbf{s}_i)\}$. La méthode de prédiction est décrite en détail à la section 2.4.1.

La première forme de validation croisée (2.17) permet d'obtenir une grille réduite de valeurs du paramètre de fenêtre λ. Celle-ci est ensuite utilisée dans la deuxième forme de validation croisée (2.18) pour déterminer de manière efficace la valeur optimale de (λ, ω). Ainsi, la première étape permet de réduire le temps de calcul de la deuxième. Cette approche de sélection d'hyper-paramètres est illustrée par un exemple à la section 2.5.2.

Il est important de souligner que la méthode proposée dépend par ailleurs du nombre de points supports. Le nombre de points supports est un compromis entre le temps de calcul et la précision de la déformation. Comme nous le verrons dans les exemples simulés du présent chapitre et dans les exemples réels du chapitre 4, il suffira d'un nombre de points supports assez raisonnable pour obtenir de bons résultats.

2.4 Prédiction

L'une des finalités de la modélisation et de l'estimation de la structure de dépendance spatiale est d'interpoler spatialement les données et/ou d'effectuer des simulations conditionnelles. Dans cette section, une description du krigeage ainsi que des simulations conditionnelles, basés sur le modèle non-stationnaire de déformation spatiale (2.3) est donnée.

2.4.1 Krigeage

L'estimation du modèle de déformation spatiale (2.3) peut être utilisée dans les équations de krigeage pour fournir des prédictions spatiales et des variances d'erreur plus fiables que celles basées sur une structure de variogramme stationnaire inadéquate.

Étant donné le vecteur $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$ aux sites d'observation $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset G$, la prédiction de la valeur inconnue de Z à l'emplacement non-observé $\mathbf{s}_0 \in G$ est donnée par l'estimateur de krigeage ordinaire :

$$\widehat{Z}(\mathbf{s}_0) = \widehat{Y}(\mathbf{u}_0) = \sum_{i=1}^{n} \alpha_i(\mathbf{u}_0) Y(\mathbf{u}_i) = \boldsymbol{\alpha}^T \mathbf{Y}, \tag{2.19}$$

qui minimise l'erreur quadratique moyenne $\mathbb{E}(\widehat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0))^2$ sous la contrainte : $\mathbf{1}^T \boldsymbol{\alpha} = 1$. $\mathbf{u}_i = \widehat{f}(\mathbf{s}_i), Y(\mathbf{u}_i) = Z(\mathbf{s}_i),\ i = 0, \ldots, n$ représentent les points transformés.

Le problème de prédiction de la fonction aléatoire non-stationnaire Z est alors transposé à la fonction aléatoire stationnaire et isotrope Y. Ainsi, les poids de krigeage $\boldsymbol{\alpha} = [\alpha_i(\mathbf{u}_0)]$ sont donnés par le système classique de krigeage ordinaire, dans le contexte stationnaire [Chilès and Delfiner, 2012] :

$$\begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ v \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma} \\ 1 \end{bmatrix}, \tag{2.20}$$

où $\boldsymbol{\Gamma} = [\gamma_0(\|\mathbf{u}_i - \mathbf{u}_j\|)]$; $\boldsymbol{\gamma} = [\gamma_0(\|\mathbf{u}_0 - \mathbf{u}_j\|)]$; v est le multiplicateur de Lagrange. La résolution du système conduit à :

$$\boldsymbol{\alpha} = \boldsymbol{\Gamma}^{-1} \left(\boldsymbol{\gamma} + \mathbf{1} \frac{1 - \mathbf{1}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}}{\mathbf{1}^T \boldsymbol{\Gamma}^{-1} \mathbf{1}} \right). \tag{2.21}$$

L'erreur quadratique moyenne minimisée (variance de krigeage) est donnée par :

$$\sigma^2(\mathbf{s}_0) = \boldsymbol{\alpha}^T \boldsymbol{\gamma} + v = \boldsymbol{\gamma}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma} - \frac{(\mathbf{1}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma} - 1)^2}{\mathbf{1}^T \boldsymbol{\Gamma}^{-1} \mathbf{1}}. \tag{2.22}$$

Ainsi, en utilisant l'approche par déformation d'espace pour kriger la variable régionalisée sur une grille de points cibles, on peut procéder comme suit :

Algorithme 2.4.1 Krigeage

1. obtenir l'image de la grille de points cibles et des points de données via la déformation estimée $\widehat{f}(.)$;
2. kriger la grille de points cibles transformés, en utilisant le variogramme stationnaire et isotrope estimé $\widehat{\gamma}_0(.)$ et les points de données transformés ;
3. obtenir le krigeage sur la grille d'origine de points cibles par simple correspondance.

2.4.2 Simulation conditionnelle

Il existe un certain nombre de situations où les techniques classiques de prédiction telles que le krigeage remplissent mal leurs objectifs. Il en est ainsi lorsqu'il s'agit d'intégrer des données, de prédire des quantités telles que des extremas ou des quantiles. Une approche intéressante pour former des estimateurs cohérents de n'importe quelle quantité d'intérêt ou pour quantifier les incertitudes est l'utilisation de simulations conditionnelles. Une simulation conditionnelle est une réalisation d'un modèle stochastique qui respecte l'ensemble des données disponibles. Pour y arriver, on peut construire une simulation non-conditionnelle puis la conditionner aux données à l'aide de l'algorithme classique qui consiste à ajouter au krigeage la simulation de l'erreur de krigeage [Chilès and Delfiner, 2012].

L'objectif ici est de simuler une réalisation $\{z(\mathbf{s}), \mathbf{s} \in G\}$ d'une fonction aléatoire gaussienne Z de moyenne constante m et de modèle de second ordre : $\gamma(\mathbf{x}, \mathbf{y}) = \gamma_0(\|f(\mathbf{x}) - f(\mathbf{y})\|)$, relativement aux données $\{Z(\mathbf{s}_i) = z_i, i = 1, \ldots, n\}$.

Effectuer une simulation conditionnelle de la fonction aléatoire non-stationnaire gaussienne Z définie sur le domaine G revient à le faire pour la fonction aléatoire gaussienne Y définie sur le domaine D et de variogramme stationnaire isotrope $\gamma_0(.)$. Ceci peut être réalisé, à partir d'une simulation non-conditionnelle W de même moyenne et variogramme que la fonction aléatoire stationnaire gaussienne Y. Pour cette dernière, des techniques de simulations classiques sont disponibles [Lantuéjoul, 2002] : méthode spectrale, méthode des bandes tournantes, ... Par conséquent, un algorithme de simulation conditionnelle de la fonction aléatoire non-stationnaire Z est le suivant :

Algorithme 2.4.2 Simulation conditionnelle

1. calculer les valeurs krigées $y^\star(\mathbf{u}) = m + \sum_{i=1}^{n} \alpha_i(\mathbf{u})[y(\mathbf{u}_i) - m]$ pour chaque $\mathbf{u} = f(\mathbf{s}), \mathbf{s} \in G$;
2. simuler w une réalisation de la fonction aléatoire gaussienne W de moyenne constante m et de variogramme stationnaire et isotrope $\gamma_0(.)$ dans le domaine D et aux points conditionnant transformés. Poser $\{w(\mathbf{u}), \mathbf{u} = f(\mathbf{s}), \mathbf{s} \in G\}$ et $\{w_i = w(\mathbf{u}_i), \mathbf{u}_i = f(\mathbf{s}_i)\}$ les valeurs simulées ;
3. calculer les valeurs krigées $w^\star(\mathbf{u}) = m + \sum_{i=1}^{n} \alpha_i(\mathbf{u})[w_i - m]$ pour chaque $\mathbf{u} = f(\mathbf{s}), \mathbf{s} \in G$;
4. retourner $z(\mathbf{s}) = y^\star(\mathbf{u}) + w(\mathbf{u}) - w^\star(\mathbf{u}), \mathbf{u} = f(\mathbf{s}), \mathbf{s} \in G$.

2.5 Illustration

Nous illustrons de quelle manière la méthodologie d'inférence proposée permet de déterminer la déformation sous-jacente et améliore ainsi le schéma d'interpolation par rapport à l'approche stationnaire classique. Deux exemples de données simulées en 1D ($p = 1$) et 2D ($p = 2$) sont utilisés. L'application à des ensembles de données réelles est présentée au chapitre 4.

2.5.1 Exemple 1D

On commence avec un exemple en une dimension pour lequel les résultats peuvent être facilement visualisés. On simule en 1000 points irrégulièrement espacés d'un domaine $G = [0, 1]$, une fonction aléatoire gaussienne Z de modèle de second ordre : $\gamma(x, y) = \gamma_0(|f(x) - f(y)|)$, où $\gamma_0(.)$ est un variogrammme exponentiel stationnaire isotrope, de paramètre de portée $0,125$ et $f(x) = x^4$. Notre objectif est d'estimer la fonction de déformation sous-jacente $f(.)$ à partir des 1000 points.

La déformation $f(.)$ laisse le segment G globalement inchangé, mais contracte le côté gauche du segment tandis que le côté droit est étiré. Ainsi, les points du segment qui sont situés près de l'origine sont fortement corrélés avec leurs voisins et ceux situés près de l'extrémité sont légèrement corrélés avec leurs voisins. Un ensemble de 125 points supports régulièrement espacés sur G est utilisé pour construire la matrice de dissimilarité.

Les figures 2.2a et 2.2b montrent une réalisation de la fonction aléatoire non-stationnaire Z dans l'espace géographique G et l'espace déformé estimé D. Comme on peut le voir, la fonction aléatoire non-stationnaire est transformée en stationnaire par simple contraction et étirement du segment G. Les zones de fortes corrélations ont été contractées, tandis que celle de faibles corrélations ont été étirées. Les valeurs optimales des hyper-paramètres sont $\lambda = 0,83$ et $\omega = 0,65$.

Les figures 2.2c et 2.2d présentent la vraie déformation et celle estimée par l'approche. Nous pouvons constater que la déformation estimée et la vraie sont assez proches. En terme de structure de dépendance spatiale, les variogrammes non-stationnaire et transformé estimés sont présentés aux figures 2.3a et 2.3c. Là aussi, on peut observer la manière dont la déformation opère (contraction et étirement) sur le domaine d'intérêt. Comme le montrent les figures 2.3b et 2.3d la structure de dépendance spatiale non-stationnaire estimée est assez proche de la vraie. Ce simple exemple illustre la capacité de l'approche proposée à retrouver la déformation sous-jacente.

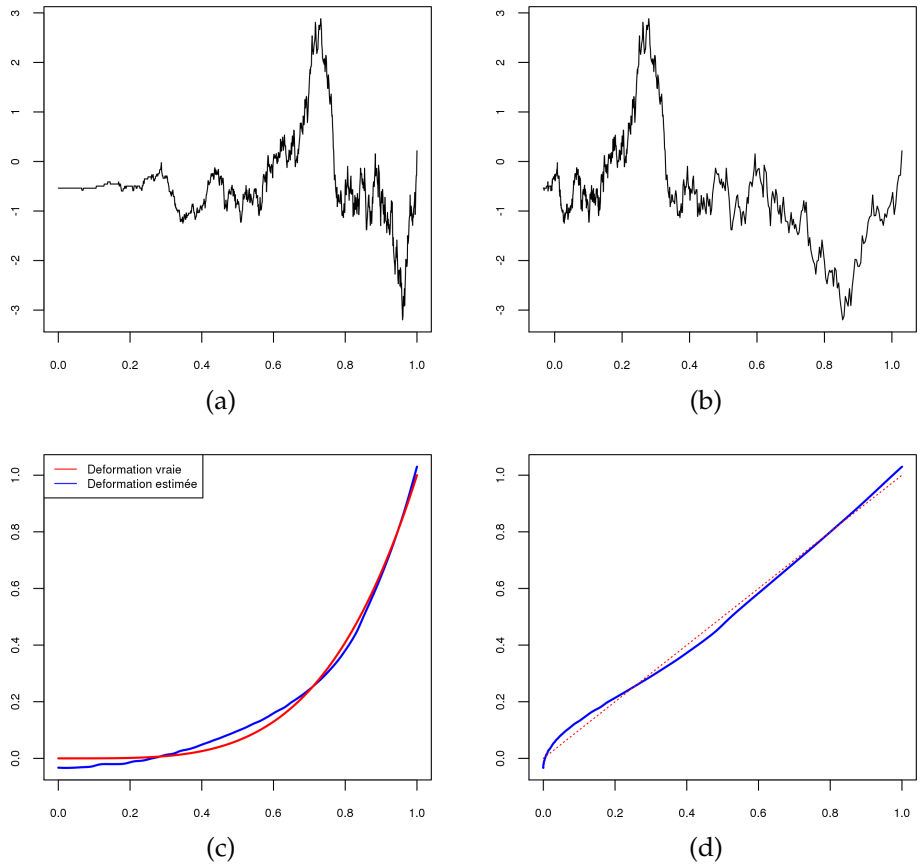


FIGURE 2.2 – Méthode de déformation d'espace : réalisation d'une fonction aléatoire non-stationnaire dans l'espace géographique (a) et l'espace déformé estimé (b) ; (c) déformation vraie et son estimation ; (d) points déformés vrais *versus* points déformés estimés.

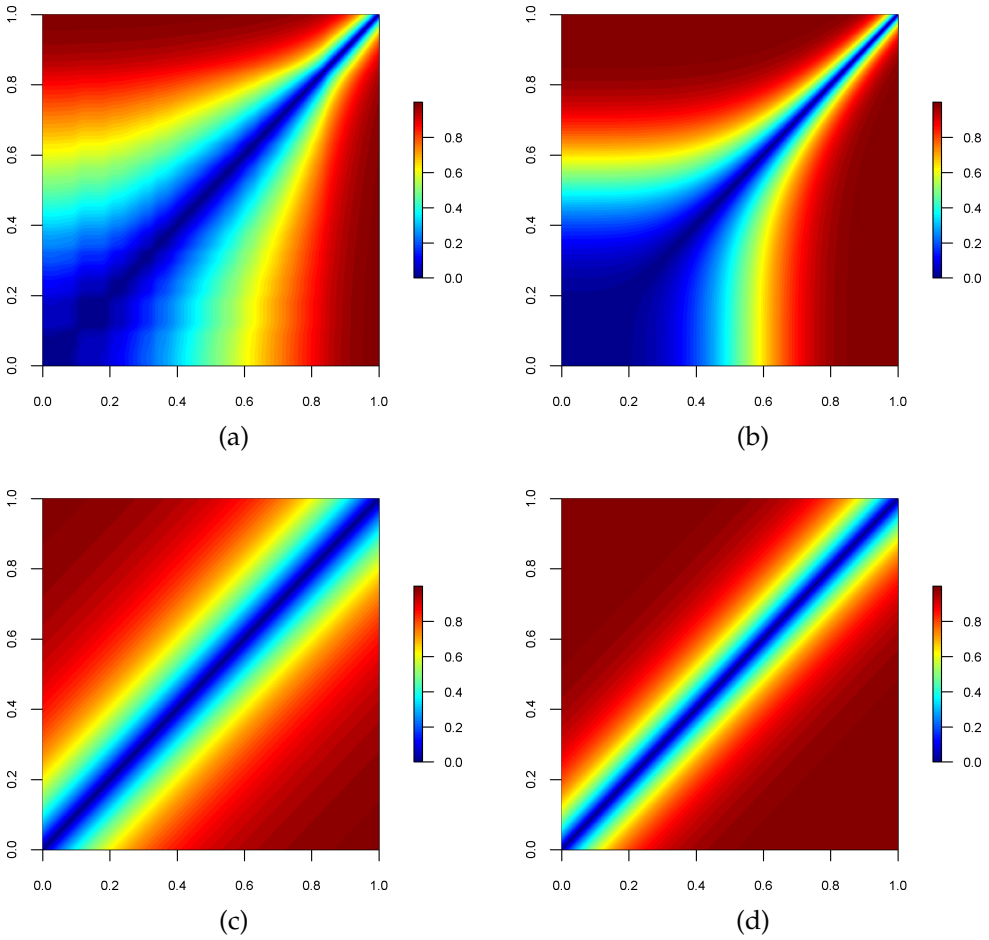


FIGURE 2.3 – Méthode de déformation d'espace : (a, b) variogrammes non-stationnaires estimé $\hat{\gamma}(.,.)$ et vrai $\gamma(.,.)$; (c, d) variogrammes transformés estimé $\hat{\gamma}_0(.)$ et vrai $\gamma_0(.)$.

2.5.2 Exemple 2D

On considère maintenant une fonction aléatoire gaussienne standardisée Z définie sur le domaine $G = [0,1]^2$, de variogramme : $\gamma(\mathbf{x},\mathbf{y}) = \gamma_0(\|f(\mathbf{x}) - f(\mathbf{y})\|)$, où $f(\mathbf{x}) = \mathbf{o} + (\mathbf{x}-\mathbf{o})\|\mathbf{x}-\mathbf{o}\|$ est une fonction radiale de centre $\mathbf{o} = (0,5;0,5)$ et $\gamma_0(.)$ un variogramme cubique stationnaire et isotrope, de portée $0,05$.

On simule Z en 200×200 points d'une grille régulière de G. Les points du domaine géographique qui sont situés près du centre sont fortement corrélés avec leurs voisins tandis que les points qui sont éloignés du centre sont peu corrélés avec leurs voisins (figure 2.4a). 2249 points sont échantillonnés et séparés en deux échantillons : un d'apprentissage de 1225 points et un de validation de 1024 points. Une grille régulière (13×13) est utilisée comme configuration de points supports (voir figure 2.4b).

La figure 2.4d montre l'espace déformé estimé qui est assez proche du vrai espace déformé (figure 2.4c). Notre méthode d'estimation étire le domaine dans les régions de corrélation spatiale relativement faible (aux extrémités), tandis qu'elle contracte le domaine dans les régions de corrélation spatiale relativement élevée (au centre), de sorte qu'un variogramme stationnaire isotrope puisse modéliser la structure de dépendance spatiale dans l'espace déformé.

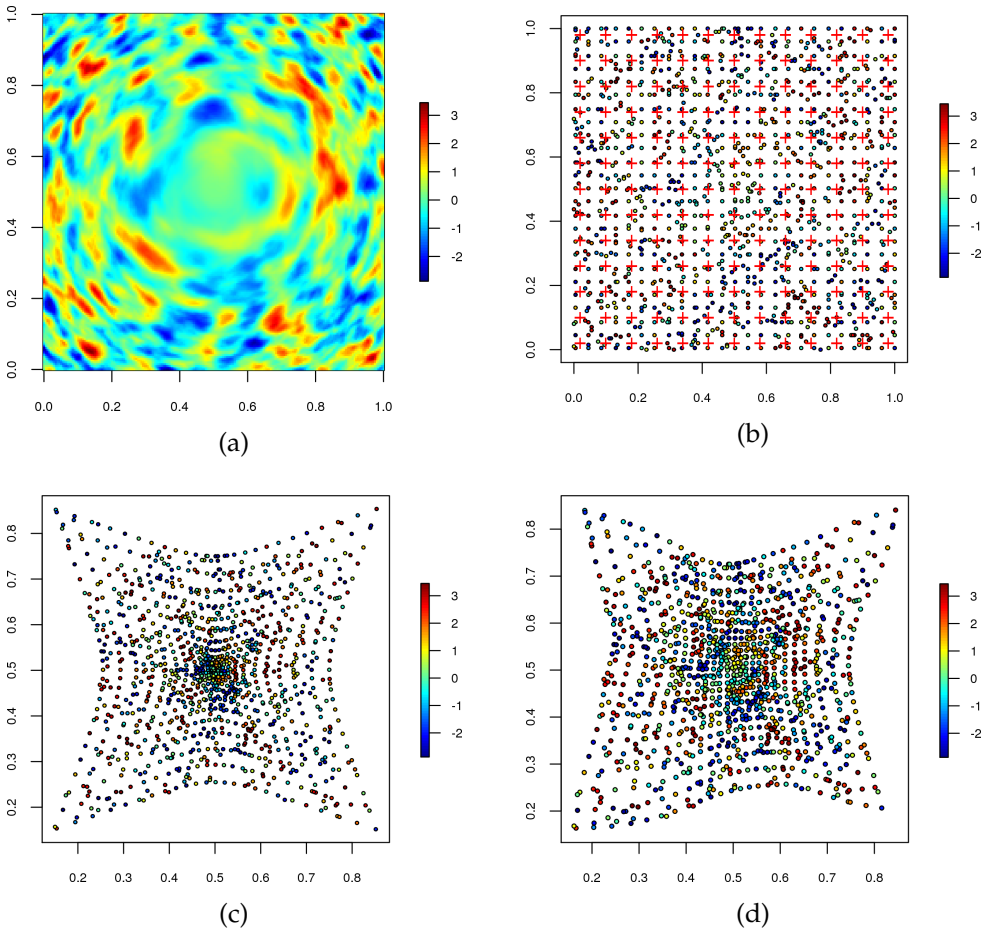


FIGURE 2.4 – Méthode de déformation d'espace : (a) réalisation d'une fonction aléatoire non-stationnaire par déformation d'espace ; (b) points de données et points supports (croix rouge) ; points de données dans l'espace déformé vrai (c) et l'espace déformé estimé (d).

La figure 2.5a présente la fonction score en validation croisée $CV(\lambda)$ définie en (2.17). Nous observons que les valeurs modérées à grande de la largeur de la fenêtre λ tendent à donner des scores faibles. La figure 2.5b montre la seconde fonction de score de validation croisée $MSE(\lambda, \omega)$ défini en (2.18). Suivant la procédure de sélection des hyper-paramètres décrite à la section 2.3.4, la valeur optimale des hyper-paramètres est obtenue pour $\lambda = 0,65$ and $\omega = 0,725$. Ces valeurs optimales sont corroborées en validation externe. En effet, la figure 2.5d présente la fonction de score $MSE(\lambda, \omega)$ calculée cette fois sur l'échantillon de validation. Sur cet échantillon, les valeurs optimales sont très proches des

valeurs précédentes. Ainsi, la validation croisée produit un biais faible dans la sélection des hyper-paramètres. En observant les figures 2.5b et 2.5d, on constate que les fonctions de score en validation croisée et externe sont assez similaires. Ceci peut s'expliquer du fait de la corrélation existante entre les données d'apprentissage et de validation. Le diagramme de Shepard de la solution NMDS (espace déformé) est présenté à la figure 2.5c. Il mesure le degré de concordance entre les dissimilarités observées et les distances ajustées lors de l'étape de régression isotonique. La valeur finale du *stress* correspondant à l'espace déformé vaut près de 8%, soit une qualité d'ajustement assez bien selon le tableau 2.1.

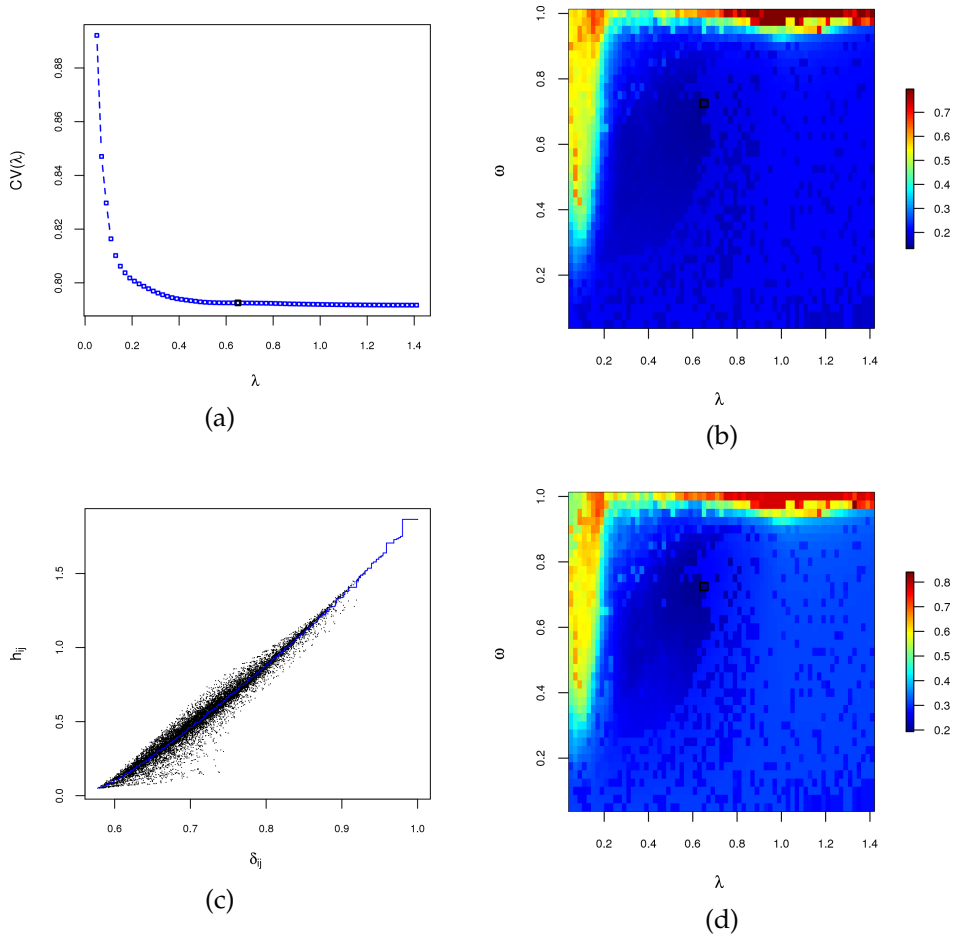


FIGURE 2.5 – Méthode de déformation d'espace : (a) fonction score $CV(\lambda)$ en validation croisée ; fonction score $MSE(\lambda, \omega)$ en validation croisée (b) et externe (d) ; (c) diagramme de Shepard de la solution NMDS.

Une visualisation du variogramme en quelques points pour les modèles stationnaire estimé M1, non-stationnaire estimé M2 et non-stationnaire de référence M3 (vraie déformation) est présentée aux figures 2.6 et 2.7. Le changement de la structure de dépendance spatiale non-stationnaire d'un point à un autre s'observe clairement, notamment l'anisotropie.

Le krigeage et l'écart type de krigeage associés aux modèles M1, M2 et M3 sont présentées à la figure 2.8. Les différences entre cartes d'écart type de krigeage associées à chacun des modèles sont visiblement marquées. Bien que le krigeage soit moins sensible à une mauvaise spécification du modèle, la variance d'erreur y est par contre plus sensible. Il apparaît que la méthode stationnaire donne une version trop simpliste de l'écart type de krigeage, car ne prenant pas en compte les variations de variabilité dans le champ des données. La méthode par déformation d'espace fournit une variance d'erreur faible dans les zones de forte corrélation spatiale et une variance d'erreur forte dans les zones de faible corrélation spatiale.

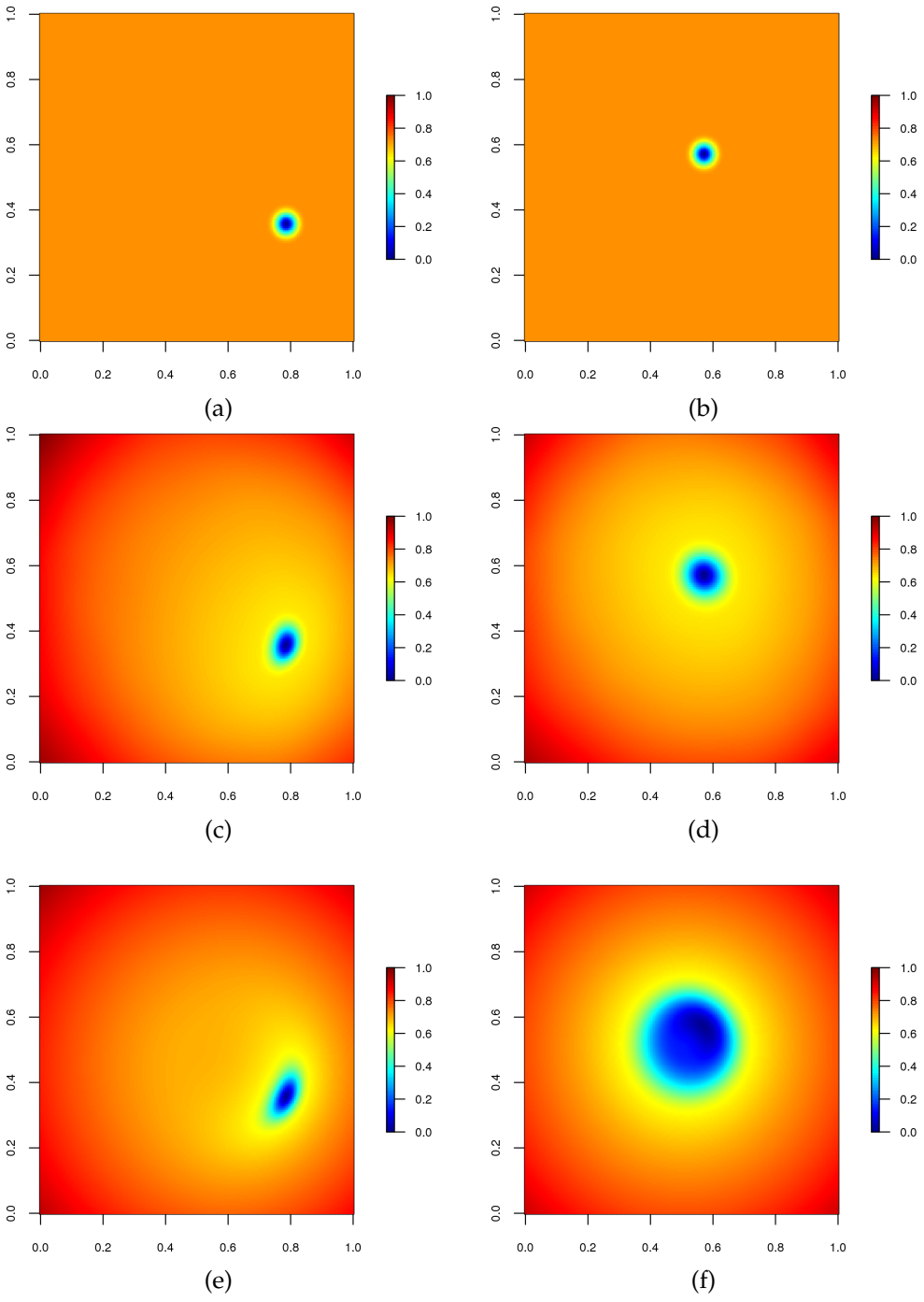


FIGURE 2.6 – Méthode de déformation d'espace : variogramme d'un point avec tous les autres points pour les modèles M1 (a, b), M2 (c, d) et M3 (e, f).

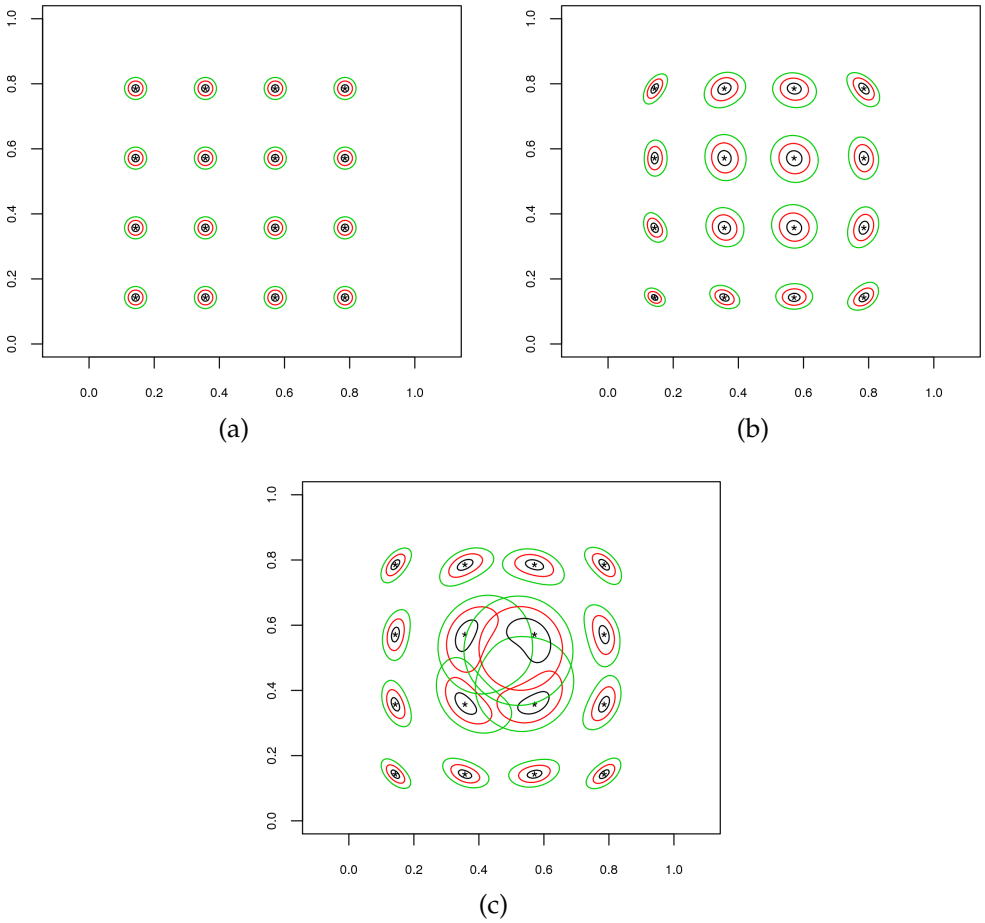


FIGURE 2.7 – Méthode de déformation d'espace : contours du variogramme en quelques points pour les modèles M1 (a), M2 (b) et M3 (c). Les isolignes correspondent aux valeurs : $0,1$ (noir), $0,3$ (rouge) et $0,5$ (vert).

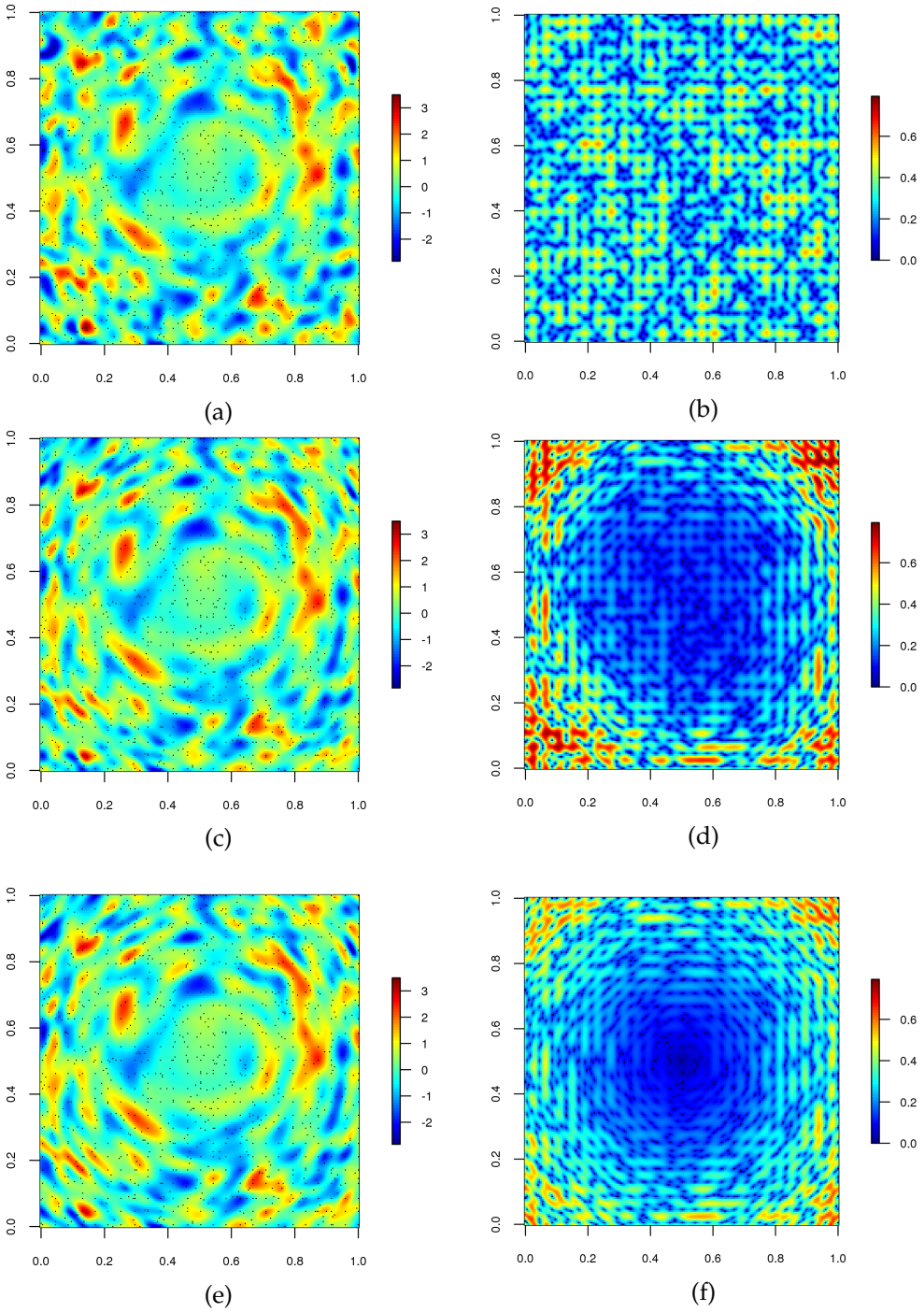


FIGURE 2.8 – Méthode de déformation d'espace : (a, c, e) résultats d'interpolation pour les modèles M1, M2 et M3 ; (b, d, f) cartes d'écart type de krigeage correspondantes.

Pour évaluer la performance prédictive de l'approche proposée, la variable régionalisée est prédite en 1024 points de données de validation (validation externe). Le tableau 2.2 fournit les performances comparatives entre les modèles M1, M2 et M3. Quelques mesures de score de prédiction bien connues sont utilisées, à savoir l'erreur absolue moyenne (*Mean Absolute Error*, MAE), la racine carrée de l'erreur quadratique moyenne (*Root Mean Square Error*, RMSE), l'erreur quadratique moyenne normalisée (*Normalized Mean Square Error*, NMSE), le score logarithmique (*Logarithmic Score*, LogS) et l'indice continu de probabilité ordonnée (*Continued Ranked Probability Score*, CRPS). Pour le RMSE, LogS et CRPS plus ils sont petits meilleurs ils sont. Pour MAE, plus il est proche de 0, meilleur il est ; pour le NMSE, plus il proche de 1 meilleur il est. La formulation de ces métriques est présentée en annexe B.

Le tableau 2.2 résume les résultats des statistiques de performance prédictive calculés sur l'échantillon de validation. On peut constater que le modèle stationnaire M1 est moins bon que les deux autres modèles non-stationnaires M2 et M3, en termes par exemple de RMSE, LogS et CRPS. Le coût de non-utilisation du modèle non-stationnaire est substantiel. En effet, le modèle stationnaire estimé est 18% moins bon que le modèle non-stationnaire estimé, en termes de RMSE.

	Stationnaire (M1)	Non-stationnaire (M2)	Référence (M3)
MAE	0,28	0,24	0,23
RMSE	0,44	0,37	0,35
NMSE	2,03	1,17	1,08
LogS	1118	-69	-92
CRPS	0,35	0,29	0,28

TABLEAU 2.2 – Méthode de déformation d'espace : validation externe sur un échantillon de 1024 observations.

2.6 Conclusion

Dans ce chapitre, nous avons abordé le problème de l'inférence du modèle de déformation d'espace dans le cadre d'une réalisation unique. Jusqu'ici, cette approche n'a été appliquée avec succès sur des données réelles que dans le contexte de réalisations multiples. La méthodologie d'inférence proposée est capable d'approcher convenablement la déformation sous-jacente, comme illustré sur les deux exemples simulés. La comparaison des performances prédictives de l'approche stationnaire avec celle proposée sur un exemple simulé est en faveur de cette dernière. En particulier, elle donne une estimation plus réaliste de la variance de l'erreur d'estimation et est ainsi en accord avec la structure

sous-jacente des données. Cependant, travailler avec des données simulées est probablement la situation la plus favorable. Nous aurons l'occasion d'appliquer la méthode proposée sur divers jeux de données réelles et d'effectuer le même type de comparaisons, au chapitre 4. Le chapitre suivant est consacré au développement d'une nouvelle approche de modélisation et d'inférence non-stationnaire basée sur l'idée de convolution.

Chapitre 3

Modèle de convolution généralisé et inférence

"Imagination is more important than knowledge : for knowledge is limited to what we know and understand while imagination embraces the entire world and all that ever will be known and understood." A. Einstein

Sommaire

3.1 Introduction **91**
3.2 Définition du modèle **94**
3.3 Classes de covariance non-stationnaire **95**
3.4 Inférence statistique **98**
3.5 Prédiction **105**
3.6 Illustration **107**
3.7 Conclusion **113**

3.1 Introduction

Comme nous l'avons vu au chapitre 1, l'approche par convolution est une autre manière constructive d'introduction de la non-stationnarité d'ordre 2 [Higdon, 1998, Higdon et al., 1999]. Elle consiste à prendre un bruit blanc spatial, puis à le moyenner en utilisant des poids variant spatialement pour ainsi obtenir une fonction aléatoire non-stationnaire. Par cette façon de faire, la structure spatiale de second ordre résultante est non-stationnaire. Le variogramme (ou la covariance) est ainsi autorisé(e) à varier d'une certaine manière entre les différents emplacements du domaine d'intérêt. Higdon et al. [1999] utilisent comme poids un noyau gaussien variant spatialement pour induire la structure de covariance non-stationnaire. Cette dernière a alors une expression analytique mais est indéfiniment différentiable, ce qui peut ne pas être souhaitable pour la modélisation de phénomènes réels [Stein, 1999]. Zhu and Wu [2010] travaillent avec une famille de noyaux

de Bessel modifiés variant spatialement pour produire une covariance non-stationnaire. Cette dernière possède des caractéristiques de régularité locale semblables à celles de la classe de covariances stationnaires de type Matérn. Une limite de cette approche est que la covariance non-stationnaire n'a pas d'expression analytique. Par conséquent, elle ne peut être évaluée en général que par intégration numérique. En outre, cette approche ne prend pas en compte l'anisotropie géométrique locale variable.

Par ailleurs, des expressions explicites de classe de covariances non-stationnaires inspirées de l'approche par convolution ont été introduites par Paciorek [2003] et Paciorek and Schervish [2004, 2006]. Celles-ci ont été étendues par Stephenson et al. [2005], Porcu et al. [2009] et Mateu et al. [2010]. Elles incluent une version non-stationnaire de la covariance stationnaire de Matérn. Cependant, ces classes de covariances non-stationnaires ne dérivent pas directement d'un modèle de fonctions aléatoires comme les représentations en convolution ou spectrale, ce qui ne facilite pas leur compréhension et leur interprétation. Il serait utile et intéressant d'avoir une approche constructive pour les fonctions aléatoires admettant de telles covariances non-stationnaires.

Aussi, l'inférence des paramètres de ces classes de covariances non-stationnaires reste un problème crucial. Paciorek and Schervish [2006] listent les difficultés et suggèrent quelques pistes d'inférence dont celle par fenêtre glissante basée sur le variogramme local ou la vraisemblance locale. Anderes and Stein [2011] mentionnent deux problèmes généralement rencontrés avec la méthode de fenêtre glissante. Le premier concerne le domaine de validité d'une approximation stationnaire qui peut être trop petit pour contenir suffisamment de données locales pour estimer fiablement la structure de dépendance spatiale. Le second est le risque d'avoir des estimations de paramètres non lisses spatialement, conduisant ainsi à des discontinuités sur la carte de krigeage, ce qui n'est pas souhaitable dans de nombreux cas. En utilisant la version non-stationnaire de la covariance stationnaire de Matérn, Anderes and Stein [2011] développent une estimation des paramètres qui régissent la dépendance spatiale d'une fonction aléatoire localement stationnaire par vraisemblance locale pondérée. Cette approche atténue lentement le poids de l'influence des observations lointaines et autorise un échantillonnage irrégulier. Les limites de cette approche sont la charge de calcul liée à l'inversion d'une matrice de covariance à chaque emplacement pour l'estimation des paramètres et l'hypothèse de distribution gaussienne de la fonction aléatoire non-stationnaire qui permet d'effectuer les calculs analytiques.

Dans ce chapitre, nous nous intéressons à la modélisation et à l'inférence des fonctions aléatoires non-stationnaires de second ordre, en développant une approche constructive suivant l'esprit des modèles de convolution. Nous présentons un nouveau modèle pour les fonctions aléatoires non-stationnaires comme la convolution d'une mesure aléatoire orthogonale avec une fonction de pondération stochastique variant spatialement. Il s'agit

là d'une extension de l'approche classique de convolution de fonctions aléatoires, basée sur une fonction de pondération jusqu'ici déterministe. Par cette façon de procéder, la classe de covariances non-stationnaires résultante est un mélange de modèles. Elle est très générale, flexible, et permet pour des choix spécifiques de la famille de fonctions de pondération stochastiques, de retrouver les expressions analytiques de classes de covariances non-stationnaires connues de la littérature. Ces dernières montrent localement un comportement stationnaire et leurs paramètres sont autorisés à varier spatialement, donnant ainsi des variances, portées, anisotropies géométriques et régularités locales. Ceci permet de regrouper naturellement les paramètres de covariance locale stationnaire dans une covariance globale non-stationnaire valide. Nous établissons ainsi une connexion directe entre les classes explicites de covariances non-stationnaires existantes et les modèles de convolution. Cette construction présente certaines similitudes avec le modèle de moyenne mobile à fonction de pondération aléatoire introduit par Matérn [1986] qui permet de générer des familles de covariances stationnaires comme celle de Matérn et de Cauchy.

Par ailleurs, nous développons une procédure d'inférence des paramètres de ces classes explicites de covariances non-stationnaires. L'estimation des paramètres se fait dans le contexte de réalisation unique et de stationnarité locale, par une approche étape par étape. Tout d'abord, nous construisons un estimateur non-paramétrique à noyau du variogramme local. Ensuite, il est utilisé dans une procédure de moindres carrés locaux pondérés pour estimer les paramètres en un ensemble réduit de points représentatifs appelés points supports. Enfin, une méthode de lissage par noyau est utilisée pour interpoler les paramètres en tout emplacement d'intérêt (points d'observation et de prédiction). Notre procédure d'estimation ne repose pas sur une hypothèse en loi (*free distribution*) et ne nécessite aucune inversion de matrice ou calcul de déterminant.

Le présent chapitre s'articule comme suit : à la section 3.2, le modèle de convolution généralisé est défini à travers ses ingrédients de base et principales propriétés. Dans la section 3.3, nous montrons comment le modèle proposé permet de construire des classes explicites de covariances non-stationnaires. En section 3.4, l'inférence statistique des paramètres de ces classes explicites de covariances non-stationnaires est détaillée. Nous abordons la prédiction spatiale et la simulation conditionnelle en section 3.5. Un jeu de données synthétiques est utilisé dans la section 3.6 pour illustrer l'approche. Enfin, la section 3.7 conclut le chapitre.

3.2 Définition du modèle

Soit $Z = \{Z(\mathbf{x}) : \mathbf{x} \in G \subseteq \mathbb{R}^p, p \geq 1\}$ une fonction aléatoire définie sur un domaine d'intérêt fixe et continu G de l'espace euclidien $\mathbb{R}^p$. La fonction aléatoire Z est définie comme suit :

$$Z(\mathbf{x}) = \int_{\mathbb{R}^p} f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))W(d\mathbf{u}), \quad \forall \mathbf{x} \in G, \tag{3.1}$$

où $W(.)$ est une mesure aléatoire orthogonale [1] sur $\mathbb{R}^p$ avec $\mathbb{E}(W(d\mathbf{u})) = 0$, $\mathbb{E}\left(W(d\mathbf{u})W(d\mathbf{v})\right) = \lambda\delta_{\mathbf{u}}(d\mathbf{v})d\mathbf{u}$, $\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^p$; $\delta_{\mathbf{u}}(.)$ est la mesure de Dirac en $\mathbf{u}$ et $\lambda \in \mathbb{R}_+^*$. $\{T(\mathbf{u}), \mathbf{u} \in \mathbb{R}^p\}$ est une famille de variables aléatoires indépendantes et identiquement distribuées, et indépendantes de $W(.)$, prenant ses valeurs dans un sous-ensemble $\mathcal{T}$ de $\mathbb{R}$ selon une mesure de probabilité $\mu(.)$. $\{\mathbf{u} \mapsto f_{\mathbf{x}}(\mathbf{u}; t), \mathbf{x} \in G \subseteq \mathbb{R}^p\}$ est une famille de fonctions de carré intégrable sur $\mathbb{R}^p$, pour tout $t \in \mathcal{T}$. Pour $\mathbf{u}$ fixé, $f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))$ est une variable aléatoire dont les deux premiers moments sont supposés finis et intégrables sur $\mathbb{R}^p$.

Sous ces hypothèses, la convolution stochastique définie en (3.1) est bien définie dans L^2 : $\mathbb{V}(Z(\mathbf{x})) = \lambda \int_{\mathbb{R}^p} \mathbb{E}\{f_{\mathbf{x}}^2(\mathbf{u}; T(\mathbf{u}))\}d\mathbf{u} < +\infty$, $\forall \mathbf{x} \in G$. Pour se fixer les idées, voici deux exemples de mesures aléatoires orthogonales pouvant être utilisées pour simuler la fonction aléatoire Z [Chilès and Delfiner, 2012] :

1. soit un processus de Poisson sur $\mathbb{R}^p$ homogène d'intensité λ. Soit $N(A)$ le nombre de points tombant dans l'ensemble mesurable $A \subset \mathbb{R}^p$. La fonction $W(.)$ définie par $W(A) = N(A) - \lambda\nu(A)$ est une mesure aléatoire orthogonale, $\nu(.)$ étant la mesure de Lebesgue sur $\mathbb{R}^p$;
2. le bruit blanc gaussien sur $\mathbb{R}^p$ est une mesure aléatoire $W(.)$ telle que : $W(A)$ suit une loi normale de moyenne 0 et de variance $\lambda\nu(A)$; lorsque A et B sont disjoints $W(A)$ et $W(B)$ sont indépendants et $W(A \cup B) = W(A) + W(B)$.

Proposition 3.1. *Sous le modèle spécifié en (3.1), nous avons les propriétés relatives aux deux premiers moments suivantes :*

$$\mathbb{E}(Z(\mathbf{x})) = 0, \quad \forall \mathbf{x} \in G, \tag{3.2}$$

$$\mathbb{C}ov(Z(\mathbf{x}), Z(\mathbf{y})) = \int_{\mathcal{T}} C(\mathbf{x}, \mathbf{y}; t)M(dt), \quad \forall (\mathbf{x}, \mathbf{y}) \in G \times G, \tag{3.3}$$

$$\text{où } C(\mathbf{x}, \mathbf{y}; t) = \int_{\mathbb{R}^p} f_{\mathbf{x}}(\mathbf{u}; t) f_{\mathbf{y}}(\mathbf{u}; t)d\mathbf{u} \quad \text{et} \quad M(dt) = \lambda\mu(dt). \tag{3.4}$$

On obtient ainsi une classe de covariances non-stationnaires très générale et flexible. Celle-ci est définie en toute paire d'emplacements d'intérêt et est valide sur $\mathbb{R}^p \times \mathbb{R}^p$

1. Une mesure aléatoire orthogonale sur $\mathbb{R}^p$ est une fonction $W : \mathcal{B}(\mathbb{R}^p) \to L^2$ telle que :
   a. $W(\sum_{k=1}^{+\infty} A_k) = \sum_{k=1}^{+\infty} W(A_k)$ pour toute famille d'éléments deux à deux disjoints de $\mathcal{B}(\mathbb{R}^p)$;
   b. Il existe une mesure $\nu(.)$ sur $\mathcal{B}(\mathbb{R}^p)$ telle que pour tout $A, B \in \mathcal{B}(\mathbb{R}^p)$, $\mathbb{E}(W(A)W(B)) = \nu(A \cap B)$.

(définie positive). Par conséquent, la fonction aléatoire Z est non-stationnaire à travers son moment d'ordre 2. La preuve de la proposition 3.1 est présentée en annexe C.

Remarque 3.2. La classe de covariances non-stationnaires obtenue en (3.3) est un mélange de convolutions. Elle permet de générer différentes formes de régularité à l'origine, y compris les faibles degrés de régularité. En effet, l'opération de randomisation par la mesure positive $M(.)$ produit des modèles qui peuvent être moins réguliers que le modèle de base induit par $f_{\mathbf{x}}(.;t)$, mais jamais plus réguliers que lui.

3.3 Classes de covariance non-stationnaire

Soient $\mathbf{\Sigma} : \mathbb{R}^p \to \mathcal{M}_p(\mathbb{R}), \mathbf{x} \mapsto \mathbf{\Sigma}_{\mathbf{x}}$ une application de $\mathbb{R}^p$ vers $\mathcal{M}_p(\mathbb{R})$ l'ensemble des matrices carrées réelles $(p \times p)$ définies positives ; $R^S(.)$ une fonction de corrélation stationnaire isotrope, définie positive sur $\mathbb{R}^p$, pour tout $p \in \mathbb{N}^\star$.

Posons : $\phi_{\mathbf{xy}} = |\mathbf{\Sigma}_{\mathbf{x}}|^{\frac{1}{4}} |\mathbf{\Sigma}_{\mathbf{y}}|^{\frac{1}{4}} \left| \frac{\mathbf{\Sigma}_{\mathbf{x}} + \mathbf{\Sigma}_{\mathbf{y}}}{2} \right|^{-\frac{1}{2}}$ et $Q_{\mathbf{xy}}(\mathbf{h}) = \mathbf{h}^T \left(\frac{\mathbf{\Sigma}_{\mathbf{x}} + \mathbf{\Sigma}_{\mathbf{y}}}{2} \right)^{-1} \mathbf{h}$, $\forall \mathbf{h} \in \mathbb{R}^p$.

Considérons les familles de fonctions $\{g(.;t)\}$ et $\{k_{\mathbf{x}}(.;t), \mathbf{x} \in \mathbb{R}^p\}$ pour tout $t \in \mathcal{T}$ telle que $\forall \mathbf{x} \in \mathbb{R}^p, g(\mathbf{x};.) \in L^2(M)$ et $k_{\mathbf{x}}(.;t)$ est la densité gaussienne multivariée centrée en $\mathbf{x}$, de matrice de covariance $\frac{t^2}{4}\mathbf{\Sigma}_{\mathbf{x}}$.

Les propositions et corollaires qui vont suivre sont démontrés en annexe C.

Proposition 3.3. *Si $f_{\mathbf{x}}(.;t) = g(\mathbf{x};t)k_{\mathbf{x}}(.;t)$, alors la classe de covariances non-stationnaires générée via* (3.3) *s'écrit :*

$$C^{NS}(\mathbf{x},\mathbf{y}) = \pi^{-\frac{p}{2}} \left| \frac{\mathbf{\Sigma}_{\mathbf{x}} + \mathbf{\Sigma}_{\mathbf{y}}}{2} \right|^{-\frac{1}{2}} \int_0^{+\infty} g(\mathbf{x};t) g(\mathbf{y};t) t^{-p} \exp\left(-\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2} \right) M(dt). \quad (3.5)$$

La proposition 3.3 fournit une classe de covariance non-stationnaire très générale et similaire à celle proposée par Stein [2005]. Les corollaires qui vont suivre montrent que pour des choix appropriés de la fonction $g(.;t)$ et de la mesure $M(.)$, on obtient entre autres les classes explicites de covariances non-stationnaires de la littérature. Nous établissons ainsi une connexion directe entre ces dernières et le modèle de convolution (3.1).

Corollaire 3.4. *Si $g(\mathbf{x};t) \propto 1$ et $M(dt) = t^p \xi(dt)$, pour $\xi(.)$ une mesure positive finie sur $\mathbb{R}_+$ telle que : $R^S(\tau) = \int_0^{+\infty} \exp\left(-\frac{\tau^2}{t^2} \right) \xi(dt), \tau \geq 0$, alors la classe de fonctions de corrélation non-stationnaire générée via* (3.5) *s'écrit :*

$$R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} R^S\left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})} \right). \quad (3.6)$$

Nous retrouvons ainsi la classe explicite de fonctions de corrélation non-stationnaire introduite par Paciorek [2003] et Paciorek and Schervish [2004, 2006]. L'intuition sous-jacente à cette classe est qu'à chaque emplacement $\mathbf{x}$ est assignée une matrice de noyau gaussien local $\mathbf{\Sigma_x}$. La corrélation entre deux points cibles $\mathbf{x}$ et $\mathbf{y}$ est alors calculée en moyennant les deux matrices de noyaux locaux aux emplacements $\mathbf{x}$ et $\mathbf{y}$. De cette façon, les caractéristiques locales aux deux emplacements influencent la corrélation modélisée aux points cibles correspondants. Elles permettent ainsi de prendre en compte la non-stationnarité. Celle-ci est donnée par l'application $\mathbf{\Sigma}(.)$ qui permet de modéliser l'anisotropie (azimut, échelle) de la fonction de corrélation. La matrice de noyau $\mathbf{\Sigma_x}$ en chaque point $\mathbf{x}$ est interprétée comme une matrice d'anisotropie géométrique localement variable. Elle contrôle le comportement anisotrope de la fonction aléatoire au voisinage de chaque point $\mathbf{x}$.

Remarque 3.5. La mesure positive $\xi(.)$ telle que définie dans le corollaire 3.4 est la mesure associée à la représentation de Schoenberg de la fonction de corrélation $R^S(.)$, fonction continue, définie positive et radiale sur $\mathbb{R}^p$ pour tout $p \in \mathbb{N}^\star$ [Schoenberg, 1938].

Remarque 3.6. Une approche basée sur la classe de modèles non-stationnaires définie en (3.6) ne permet pas l'utilisation de modèles de covariances stationnaires isotropes qui ne sont pas valides dans toutes les dimensions. Ainsi, on perd les covariances stationnaires isotropes à support compact telles que le modèle sphérique (valide dans $\mathbb{R}^p,\ p \leq 3$).

Exemple 3.7. Pour des choix spécifiques de la mesure positive $M(.)$ dans le corollaire 3.4, nous obtenons certaines fonctions de corrélations explicites non-stationnaires :

1. $M(dt) = t^p \delta_a(t) \mathbb{1}_{[0,+\infty[}(t)dt, a > 0,\ R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} \exp\left(-\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{a^2}\right)$;
2. $M(dt) = \frac{t^p}{a\sqrt{\pi}} \exp\left(-\frac{t^2}{4a^2}\right) \mathbb{1}_{[0,+\infty[}(t)dt, a > 0,\ R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} \exp\left(-\frac{\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}}{a}\right)$;
3. $M(dt) = 2t^{p+1} h(t^2) \mathbb{1}_{[0,+\infty[}(t)dt, h(.)$ est la densité d'une loi Gamma $\mathcal{G}a(\nu, 1/4a^2), \nu, a > 0, R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}}{a}\right)^\nu K_\nu \left(\frac{\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}}{a}\right)$, où $K_\nu(.)$ est la fonction de Bessel modifiée de deuxième espèce d'ordre ν [Gradshteyn and Ryzhik, 2007] ;
4. $M(dt) = 2t^{p+1} h(t^2) \mathbb{1}_{[0,+\infty[}(t)dt, h(.)$ est la densité d'une loi inverse Gamma $\mathcal{IG}(\alpha, a^2), \alpha, a > 0,\ R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} \left(1 + \frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{a^2}\right)^{-\alpha}$.

L'exemple 3.7 nous montre que des choix appropriés de la mesure positive $M(.)$ produisent des versions non-stationnaires de fonctions de corrélations stationnaires bien connues : gaussien, exponentiel, Matérn et Cauchy. Ces exemples se déduisent également des corollaires 3.8 et 3.9 suivants :

Corollaire 3.8. *Si* $g(\mathbf{x};t) \propto t^{\nu(\mathbf{x})}, \nu(\mathbf{x}) > 0, \forall \mathbf{x} \in \mathbb{R}^p$ *et* $M(dt) = 2t^{p-1}h(t^2)\mathbb{1}_{[0,+\infty[}(t)dt$, *avec* $h(.)$ *la densité d'une loi Gamma* $\mathcal{G}a(1, 1/4a^2), a > 0$, *alors la classe de fonctions de corrélation non-stationnaires générée via* (3.5) *s'écrit :*

$$R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} \frac{2^{1-\nu(\mathbf{x},\mathbf{y})}}{\sqrt{\Gamma(\nu(\mathbf{x}))\Gamma(\nu(\mathbf{y}))}} M_{\nu(\mathbf{x},\mathbf{y})}\left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}\right), \tag{3.7}$$

où $M_{\nu(\mathbf{x},\mathbf{y})}(\tau) = (\tau/a)^{\nu(\mathbf{x},\mathbf{y})} K_{\nu(\mathbf{x},\mathbf{y})}(\tau/a)$, $\tau \geq 0$; $K_{\nu(\mathbf{x},\mathbf{y})}(.)$ *étant la fonction de Bessel modifiée de deuxième espèce d'ordre* $\nu(\mathbf{x},\mathbf{y}) = (\nu(\mathbf{x}) + \nu(\mathbf{y}))/2$.

Le corollaire 3.8 permet de retrouver la version non-stationnaire de la fonction de corrélation stationnaire de Matérn, introduite par Stein [2005]. Cette classe est assez flexible, car elle autorise à la fois les paramètres d'anisotropie géométrique et de régularité de varier spatialement. Pour la fonction $\nu(.,.)$ constante, on retrouve bien la fonction de corrélation non-stationnaire de Matérn de paramètre de régularité constant présentée à l'exemple 3.7. Toutefois, la fonction $g(;t)$ et la mesure $M(.)$ permettant d'obtenir cette dernière diffèrent de celles des corollaires 3.4 et 3.8.

Corollaire 3.9. *Si* $g(\mathbf{x};t) \propto t^{-\alpha(\mathbf{x})}, \alpha(\mathbf{x}) > 0, \forall \mathbf{x} \in \mathbb{R}^p$ *et* $M(dt) = 2t^{p+3}h(t^2)\mathbb{1}_{[0,+\infty[}(t)dt$, *avec* $h(.)$ *la densité d'une loi inverse Gamma* $\mathcal{IG}(1, a^2), a > 0$, *alors la classe de fonctions de corrélation non-stationnaire générée via* (3.5) *s'écrit :*

$$R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} \frac{\Gamma(\alpha(\mathbf{x},\mathbf{y}))}{\sqrt{\Gamma(\alpha(\mathbf{x}))\Gamma(\alpha(\mathbf{y}))}} \left(1 + \frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{a^2}\right)^{-\alpha(\mathbf{x},\mathbf{y})}, \tag{3.8}$$

où $\alpha(\mathbf{x},\mathbf{y}) = (\alpha(\mathbf{x}) + \alpha(\mathbf{y}))/2$.

Le corollaire 3.9 donne la version non-stationnaire de la fonction de corrélation stationnaire de Cauchy, introduite par Stein [2005]. Les paramètres d'anisotropie géométrique locale et de longue portée sont autorisés à varier spatialement. Pour le paramètre de longue portée $\alpha(.,.)$ constant, on retombe sur la fonction de corrélation non-stationnaire de Cauchy présentée à l'exemple 3.7. Cependant, la fonction $g(;t)$ et la mesure $M(.)$ permettant d'obtenir cette dernière ne sont pas les mêmes que celles des corollaires 3.4 et 3.9.

Il est très simple de construire des covariances non-stationnaires explicites à partir des corrélations non-stationnaires explicites, ceci en incorporant une fonction d'écart-type. La covariance non-stationnaire explicite se définit comme suit :

$$C^{NS}(\mathbf{x},\mathbf{y}) = \sigma(\mathbf{x})\sigma(\mathbf{y})R^{NS}(\mathbf{x},\mathbf{y}), \tag{3.9}$$

où $\sigma(.)$ est la fonction d'écart-type et $R^{NS}(.,.)$ est la fonction de corrélation non-stationnaire explicite.

Remarque 3.10. Les covariances non-stationnaires explicites définies en (3.9) ont comme point de départ des covariances stationnaires. Elles ont la propriété désirable d'inclure les covariances stationnaires comme cas particulier.

Remarque 3.11. Il est parfois utile de décrire la structure spatiale de second ordre via le variogramme. Ainsi, la classe de variogrammes non-stationnaires correspondant à (3.9) s'écrit comme suit :

$$\gamma^{NS}(\mathbf{x},\mathbf{y}) = \frac{1}{2}\sigma^2(\mathbf{x}) + \frac{1}{2}\sigma^2(\mathbf{y}) - C^{NS}(\mathbf{x},\mathbf{y}). \tag{3.10}$$

Si $C^{NS}(\mathbf{x},\mathbf{y}) \to 0$ lorsque $\|\mathbf{x}-\mathbf{y}\| \to +\infty$, alors pour les grandes valeurs de $\|\mathbf{x}-\mathbf{y}\|$,

$$\gamma^{NS}(\mathbf{x},\mathbf{y}) \approx \frac{1}{2}\left(\sigma^2(\mathbf{x}) + \sigma^2(\mathbf{y})\right).$$

3.4 Inférence statistique

Nous nous intéressons maintenant à l'estimation des paramètres spatialement variables qui régissent les classes explicites de modèles de second ordre non-stationnaires présentées en section 3.3. Bien que la procédure d'inférence proposée s'applique naturellement à toutes ces classes, nous nous focalisons sur l'une d'entre elles, celle définie en (3.6). Cette dernière est suffisamment flexible pour de nombreuses applications. La méthodologie d'estimation proposée nous permet de traiter de façon simultanée d'autres types de non-stationnarités : moyenne et variance. Ainsi, nous considérons un modèle étendu construit à partir du modèle spécifié en (3.1), incluant les non-stationnarités en moyenne et variance. La procédure d'estimation basée sur l'hypothèse de stationnarité locale est effectuée en utilisant un schéma d'estimation en trois étapes. Tout d'abord, sous l'hypothèse de quasi-stationnarité un estimateur non-paramétrique à noyau du variogramme local est construit. Il est ensuite utilisé dans une procédure de moindres carrés locaux pondérés pour estimer les paramètres spatialement variables en un ensemble réduit de points représentatifs (points supports). Par la suite, une méthode de lissage par noyau est utilisée pour interpoler les paramètres en tout emplacement de la région d'intérêt.

3.4.1 Modélisation

Soit la fonction aléatoire $Y = \{Y(\mathbf{x}) : \mathbf{x} \in G \subseteq \mathbb{R}^p, p \geq 1\}$ définie comme suit :

$$Y(\mathbf{x}) = m(\mathbf{x}) + \sigma(\mathbf{x})Z(\mathbf{x}),\ \forall \mathbf{x} \in G, \tag{3.11}$$

où $m : \mathbb{R}^p \to \mathbb{R}$ est une fonction fixée inconnue ; $\sigma : \mathbb{R}^p \to \mathbb{R}^+$ est une fonction positive fixée inconnue ; Z est une fonction aléatoire de moyenne nulle, de variance unité et de

fonction de corrélation non-stationnaire définie en (3.6) :

$$R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} R^S\left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}\right).$$

Ainsi, la fonction aléatoire Z porte la structure de second ordre de la fonction aléatoire Y. La formulation du modèle défini en (3.11) conduit aux moments d'ordre 1 et 2 suivants :

$$\mathbb{E}(Y(\mathbf{x})) = m(\mathbf{x}), \tag{3.12}$$

$$\mathbb{C}\mathrm{ov}(Y(\mathbf{x}),Y(\mathbf{y})) = \sigma(\mathbf{x})\sigma(\mathbf{y})R^{NS}(\mathbf{x},\mathbf{y}) \equiv C^{NS}(\mathbf{x},\mathbf{y}). \tag{3.13}$$

Des expressions (3.12) et (3.13), il apparaît que la non-stationnarité de la fonction aléatoire Y est entièrement caractérisée par les paramètres $\sigma(.)$, $\mathbf{\Sigma}(.)$ et $m(.)$, définis en tout emplacement du domaine d'intérêt.

Soit $\mathbf{Y} = (Y(\mathbf{s}_1),\ldots,Y(\mathbf{s}_n))^T$ le $(n \times 1)$ vecteur d'observations issu de la fonction aléatoire Y aux emplacements $\{\mathbf{s}_1,\ldots,\mathbf{s}_n\} \subset G \subseteq \mathbb{R}^p$. Étant donnée $\mathbf{Y}$, l'objectif est d'inférer les fonctions d'écart-type $\sigma(.)$, de corrélation non-stationnaire déterminée par $\mathbf{\Sigma}(.)$ et de moyenne $m(.)$. Ensuite, ces dernières sont utilisées pour prédire la valeur de la fonction aléatoire Y aux emplacements non-observés. Dans la suite de l'exposé et sans perte de généralité, nous considérons l'espace de dimension 2 ($p = 2$).

3.4.2 Estimation des paramètres

L'inférence statistique des paramètres $\sigma(.)$, $\mathbf{\Sigma}(.)$ et $m(.)$ repose sur l'hypothèse de quasi-stationnarité ou stationnarité locale telle qu'introduite par Matheron [1971].

Quasi-stationnarité

Une fonction aléatoire $Y = \{Y(\mathbf{x}) : \mathbf{x} \in G \subseteq \mathbb{R}^p, p \geq 1\}$ sera dite localement stationnaire si elle possède une espérance $m(\mathbf{x})$ et une covariance $C^{NS}(\mathbf{x},\mathbf{y})$ telles que :

1. $m(\mathbf{x})$ est une fonction très régulière et variant lentement dans l'espace à l'échelle de l'information disponible ; plus précisément, $m(\mathbf{x})$ peut être considérée comme constante sur un voisinage de $\mathbf{x}$;

2. il existe une fonction de trois arguments $C^S(\mathbf{h};\mathbf{x},\mathbf{y})$ telle que l'on ait $C^{NS}(\mathbf{x},\mathbf{y}) = C^S(\mathbf{x}-\mathbf{y};\mathbf{x},\mathbf{y})$ et que à $\mathbf{h}$ fixé, $C^S(\mathbf{h};\mathbf{x},\mathbf{y})$ soit une fonction régulière et variant lentement (dans le même sens que 1.) des arguments $\mathbf{x}$ et $\mathbf{y}$. Autrement dit, pour des points d'implantation $\mathbf{x}$ et $\mathbf{y}$ pas trop éloignés l'un de l'autre, $C^S(\mathbf{h};\mathbf{x},\mathbf{y})$ ne dépend que de $\mathbf{h}$, et tout se passe comme si la covariance $C^{NS}(.,.)$ était stationnaire.

L'idée intuitive sous-jacente à cette définition est que si une fonction aléatoire est localement stationnaire, alors en tout emplacement $\mathbf{x}_0 \in G$, il existe un voisinage $\mathcal{V}_{\mathbf{x}_0}$ où la fonction aléatoire peut être approximée par une fonction aléatoire stationnaire. Ainsi, $\forall(\mathbf{x},\mathbf{y}) \in \mathcal{V}_{\mathbf{x}_0} \times \mathcal{V}_{\mathbf{x}_0}$, $m(\mathbf{x}) \approx m(\mathbf{y}) \approx m(\mathbf{x}_0)$ et $C^{NS}(\mathbf{x},\mathbf{y}) \approx C^S(\mathbf{x}-\mathbf{y};\mathbf{x}_0) = C^S(\mathbf{h};\mathbf{x}_0)$, $\|\mathbf{h}\| \leq b$; où $C^S(.)$ est une covariance stationnaire et la limite b représente le rayon du voisinage de quasi-stationnarité $\mathcal{V}_{\mathbf{x}_0}$. De cette manière, les paramètres sont supposés être des fonctions très régulières, variant lentement d'une extrémité du domaine à une autre ; l'espérance de la fonction aléatoire étant approximativement constante à l'intérieur du voisinage de quasi-stationnarité ; la covariance locale résultante en tout emplacement $\mathbf{x}_0 \in G$ s'écrivant comme suit :

$$C^S(\mathbf{h};\mathbf{x}_0) = \sigma^2(\mathbf{x}_0) R^S\left(\sqrt{\mathbf{h}^T \boldsymbol{\Sigma}_{\mathbf{x}_0}^{-1} \mathbf{h}}\right),\ \|\mathbf{h}\| \leq b. \tag{3.14}$$

En termes plus pratique, on peut définir des voisinages glissants $\mathcal{V}_{\mathbf{x}_0} = \{\mathbf{x} \in G,\ \|\mathbf{x} - \mathbf{x}_0\| \leq b\}$ à l'intérieur desquels l'espérance et la covariance peuvent être considérées comme stationnaires, et où l'information disponible est suffisante pour en faire l'inférence.

Remarque 3.12. L'hypothèse de quasi-stationnarité correspond à un compromis entre les distances d'homogénéité du phénomène étudié et la densité d'information disponible. En effet, il est toujours possible pour atteindre la stationnarité de réduire considérablement la taille b des voisinages de quasi-stationnarité. Mais alors la plupart de ces voisinages ne contiendront presque pas de données ; il ne sera plus possible de faire l'inférence des moments quasi-stationnaires dans ces voisinages.

Paramétrisation de l'anisotropie

Localement la structure de dépendance spatiale non-stationnaire définie en (3.13) est réduite à une structure stationnaire anisotrope (3.14). La fonction $\boldsymbol{\Sigma}(.)$ est paramétrée de telle façon qu'elle soit toujours définie positive et capture l'anisotropie géométrique locale variable. Ceci est obtenu en utilisant une décomposition spectrale. En tout emplacement $\mathbf{x}_0 \in G$, on a : $\boldsymbol{\Sigma}_{\mathbf{x}_0} = \boldsymbol{\Psi}_{\mathbf{x}_0} \boldsymbol{\Lambda}_{\mathbf{x}_0} \boldsymbol{\Psi}_{\mathbf{x}_0}^T$, où $\boldsymbol{\Lambda}_{\mathbf{x}_0}$ est la matrice diagonale des valeurs propres et $\boldsymbol{\Psi}_{\mathbf{x}_0}$ la matrice des vecteurs propres. Plus précisément :

$$\boldsymbol{\Lambda}_{\mathbf{x}_0} = \begin{pmatrix} \lambda_1^2(\mathbf{x}_0) & 0 \\ 0 & \lambda_2^2(\mathbf{x}_0) \end{pmatrix},\ \boldsymbol{\Psi}_{\mathbf{x}_0} = \begin{pmatrix} \cos\{\psi(\mathbf{x}_0)\} & \sin\{\psi(\mathbf{x}_0)\} \\ -\sin\{\psi(\mathbf{x}_0)\} & \cos\{\psi(\mathbf{x}_0)\} \end{pmatrix},\ \lambda_1(\mathbf{x}_0), \lambda_2(\mathbf{x}_0) > 0, \psi(\mathbf{x}_0) \in [0, \pi[.$$

Le paramètre d'échelle de la fonction de corrélation $R^S(.)$ est fixé à l'unité pour éviter tout problème de sur-paramétrisation. La fonction d'anisotropie $\boldsymbol{\Sigma}(.)$ est caractérisée par les fonctions $\lambda_1(.)$, $\lambda_2(.)$ et $\psi(.)$. En chaque emplacement, les racines carrées des valeurs propres contrôlent les portées locales et la matrice de vecteurs propres spécifie les orientations d'anisotropie géométrique locale. Ainsi, le vecteur de paramètres structuraux

est donné par $\boldsymbol{\theta}(.) = (\sigma(.), \lambda_1(.), \lambda_2(.), \psi(.))$.

Estimateur à noyau du variogramme local

Il est parfois plus commode de décrire la structure spatiale de second ordre via le variogramme. Sous l'hypothèse de stationnarité locale, nous définissons un estimateur non-paramétrique à noyau du variogramme local stationnaire $\gamma(\mathbf{h}; \mathbf{x}_0) = \sigma^2(\mathbf{x}_0) - C^S(\mathbf{h}; \mathbf{x}_0)$, $\|\mathbf{h}\| \leq b$ en l'emplacement $\mathbf{x}_0 \in G$ et au vecteur d'inter-distance $\mathbf{h} \in \mathbb{R}^p$ comme suit :

$$\widehat{\gamma}_\epsilon(\mathbf{h}; \mathbf{x}_0) = \frac{\sum_{V(\mathbf{h})} K^\star_\epsilon(\mathbf{x}_0, \mathbf{s}_i) K^\star_\epsilon(\mathbf{x}_0, \mathbf{s}_j) [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2}{2 \sum_{V(\mathbf{h})} K^\star_\epsilon(\mathbf{x}_0, \mathbf{s}_i) K^\star_\epsilon(\mathbf{x}_0, \mathbf{s}_j)}, \ \|\mathbf{h}\| \leq b, \tag{3.15}$$

où la moyenne (3.15) est prise sur $V(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}$, l'ensemble de toutes les paires d'emplacements séparés par le vecteur $\mathbf{h}$; $K^\star_\epsilon(\mathbf{x}_0, \mathbf{s}_i) = K_\epsilon(\mathbf{x}_0, \mathbf{s}_i) / \sum_{l=1}^n K_\epsilon(\mathbf{x}_0, \mathbf{s}_l)$ sont les poids normalisés ; $K_\epsilon(.,.)$ est un noyau symétrique sur $\mathbb{R}^p \times \mathbb{R}^p$, de paramètre de largeur de fenêtre $\epsilon > 0$. Pour les données irrégulièrement espacées, il n'y a généralement pas assez de paires d'observations séparées exactement de $\mathbf{h}$. Alors $V(\mathbf{h})$ est habituellement modifié par $\{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j \in T(\mathbf{h})\}$, où $T(\mathbf{h})$ est une région de tolérance de $\mathbb{R}^p$ autour de $\mathbf{h}$.

L'estimateur du variogramme local en tout emplacement $\mathbf{x}_0 \in G$ est défini comme une moyenne locale pondérée des accroissements au carré de la variable régionalisée. Le noyau est utilisé pour atténuer en douceur les accroissements au carré (pour chaque vecteur d'inter-distance) en fonction de la distance des paires d'observations correspondantes à l'emplacement cible. On attribue à chaque paire de données un poids proportionnel au produit des poids individuels. Les paires d'observations proches du point cible $\mathbf{x}_0$ ont plus d'influence sur l'estimateur du variogamme local $\widehat{\gamma}_\epsilon(\mathbf{h}; \mathbf{x}_0)$ que celles qui en sont éloignées.

Remarque 3.13. Lorsque $K_\epsilon(\mathbf{x}, \mathbf{y}) \propto 1$, $\forall (\mathbf{x}, \mathbf{y}) \in G \times G$ et $b = D/2$; où D est le diamètre du domaine d'intérêt G, on retrouve l'estimateur classique de Matheron pour une structure stationnaire globale [Matheron, 1971]. Pour le noyau uniforme $K_\epsilon(\mathbf{x}, \mathbf{y}) \propto \mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|<\epsilon}, \forall (\mathbf{x}, \mathbf{y}) \in G \times G$ et $b = \epsilon$, on a l'estimateur par fenêtre glissante [Haas, 1990b].

Les exemples classiques de noyaux ont été présentés à la section 2.3.1 du chapitre 2. Pour le calcul de l'estimateur non-paramétrique (3.15), nous optons pour un noyau gaussien stationnaire isotrope : $K_\epsilon(\mathbf{x}, \mathbf{y}) \propto \exp(-\frac{1}{2\epsilon^2} \|\mathbf{x} - \mathbf{y}\|^2)$, $\forall (\mathbf{x}, \mathbf{y}) \in G \times G$. Ce dernier est à support non-compact et considère donc toutes les observations. Ainsi, l'estimateur du variogramme local ne se limite pas seulement à l'information locale, les points éloignés sont aussi considérés. Ceci permet d'éviter des artefacts dans l'estimation des paramètres causés par la seule utilisation d'observations proches de l'emplacement cible. Il permet également de réduire l'instabilité du variogramme local calculé dans les régions à faible

densité d'échantillonnage. De plus, il fournit des estimations lisses et est ainsi compatible avec l'hypothèse de quasi-stationnarité. S'agissant de la taille du voisinage de quasi-stationnarité b, elle est paramétrée en fonction de la largeur de fenêtre ϵ. Nous prenons $b = \sqrt{3}\epsilon$ de telle sorte que l'écart-type du noyau gaussien stationnaire isotrope correspond à celui du noyau stationnaire isotrope uniforme (à support compact). Un autre choix possible pour b est de prendre un quantile du noyau stationnaire isotrope gaussien (par exemple $b \approx 2\epsilon$) ou encore la moitié de sa largeur à mi-hauteur ($b = \sqrt{2\log(2)}\epsilon$).

Estimation brute des paramètres

On souhaite estimer le vecteur de paramètres structuraux $\boldsymbol{\theta}(.) = (\sigma(.), \lambda_1(.), \lambda_2(.), \psi(.))$ et le paramètre de moyenne $m(.)$ en tout emplacement du domaine d'intérêt.

L'estimation du vecteur de paramètres $\boldsymbol{\theta}(\mathbf{x}_0)$ qui caractérise le variogramme local stationnaire $\gamma(.;\mathbf{x}_0) \equiv \gamma(.;\boldsymbol{\theta}(\mathbf{x}_0))$ à un emplacement cible $\mathbf{x}_0 \in G$ est donnée via le problème de minimisation suivant :

$$\widehat{\boldsymbol{\theta}}(\mathbf{x}_0) = \underset{\boldsymbol{\theta}(\mathbf{x}_0)\in\Theta}{\arg\min} \|\mathbf{w}_\epsilon(\mathbf{x}_0) \odot (\boldsymbol{\gamma}(\boldsymbol{\theta}(\mathbf{x}_0)) - \widehat{\boldsymbol{\gamma}}_\epsilon(\mathbf{x}_0))\|, \tag{3.16}$$

où $\odot$ désigne le produit terme à terme ; $\boldsymbol{\gamma}(\boldsymbol{\theta}(\mathbf{x}_0)) = [\gamma(\mathbf{h}_1;\boldsymbol{\theta}(\mathbf{x}_0)), \ldots, \gamma(\mathbf{h}_J;\boldsymbol{\theta}(\mathbf{x}_0))]^T$; $\widehat{\boldsymbol{\gamma}}_\epsilon(\mathbf{x}_0) = [\widehat{\gamma}_\epsilon(\mathbf{h}_1;\mathbf{x}_0), \ldots, \widehat{\gamma}_\epsilon(\mathbf{h}_J;\mathbf{x}_0)]^T$; $\mathbf{w}_\epsilon(\mathbf{x}_0) = [w_\epsilon(\mathbf{h}_1;\mathbf{x}_0), \ldots, w_\epsilon(\mathbf{h}_J;\mathbf{x}_0)]^T$, $w_\epsilon(\mathbf{h};\mathbf{x}_0) = \left[(\sum_{V(\mathbf{h})} K^\star_\epsilon(\mathbf{x}_0,\mathbf{s}_i)K^\star_\epsilon(\mathbf{x}_0,\mathbf{s}_j))/\|\mathbf{h}\|\right]^{1/2}$; $\{\mathbf{h}_j \in \mathbb{R}^p, j = 1,\ldots,J\}$ sont des vecteurs d'inter-distances donnés ; $\boldsymbol{\theta}(\mathbf{x}_0) \in \Theta$ est le vecteur de paramètres inconnus et Θ l'espace ouvert des paramètres.

Remarque 3.14. L'estimation des paramètres structuraux $\sigma(.), \lambda_1(.), \lambda_2(.), \psi(.)$ dépend du paramètre de largeur de la fenêtre ϵ via le noyau $K_\epsilon(.)$. La largeur de la fenêtre contrôle la taille de validité d'une approximation stationnaire. Sa sélection est traitée à la section 3.4.3.

Remarque 3.15. L'estimation des paramètres structuraux $\sigma(.), \lambda_1(.), \lambda_2(.)$ et $\psi(.)$ ne nécessite pas au préalable l'estimation de la fonction de moyenne $m(.)$. Par ailleurs, aucun modèle n'est spécifié pour cette dernière.

Concernant l'estimation du paramètre de moyenne, $m(\mathbf{x}_0)$ en un emplacement fixé $\mathbf{x}_0 \in G$ est approximativement constant à l'intérieur du voisinage de stationnarité locale $\mathcal{V}_{\mathbf{x}_0} = \{\mathbf{x} \in G,\ \|\mathbf{x} - \mathbf{x}_0\| \leq b\}$. Ainsi, en utilisant l'estimation du vecteur de paramètres structuraux $\widehat{\boldsymbol{\theta}}(\mathbf{x}_0)$ obtenu en (3.16), la moyenne $m(\mathbf{x}_0)$ peut être estimée explicitement par le krigeage local stationnaire de la moyenne [Chilès and Delfiner, 2012]. Plus précisément nous avons :

$$\widehat{m}(\mathbf{x}_0) = \sum_{\mathbf{s}_i \in \mathcal{V}_{\mathbf{x}_0}} \alpha_i(\mathbf{x}_0) Y(\mathbf{s}_i), \tag{3.17}$$

où $\boldsymbol{\alpha}_{\mathbf{x}_0} = [\alpha_i(\mathbf{x}_0)]$ sont les poids de krigeage donnés par :

$$\boldsymbol{\alpha}_{\mathbf{x}_0} = \frac{\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}_0}^{-1}\mathbf{1}}{\mathbf{1}^T\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}_0}^{-1}\mathbf{1}}, \tag{3.18}$$

avec $\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}_0} = [\gamma(\mathbf{s}_i - \mathbf{s}_j; \widehat{\boldsymbol{\theta}}(\mathbf{x}_0))], (\mathbf{s}_i, \mathbf{s}_j) \in \mathcal{V}_{\mathbf{x}_0} \times \mathcal{V}_{\mathbf{x}_0}$.

Lissage des paramètres

À des fins de prédiction, il nécessaire de calculer les paramètres, $\sigma(.), \lambda_1(.), \lambda_2(.), \psi(.)$ et $m(.)$ aux sites d'observation et de prédiction. Dans la pratique, il n'est pas nécessaire de résoudre le problème de minimisation (3.16) en chaque emplacement cible. En effet, le faire ainsi est intensif en calcul et redondant pour les emplacements très proches, puisque les estimations en ces derniers sont fortement corrélées. Afin de réduire la charge de calcul, l'idée proposée consiste à obtenir les estimations de paramètres uniquement pour un ensemble réduit de $m \ll n$ points représentatifs (points supports) définis sur le domaine. Puis, en utilisant les estimations obtenues aux points supports, un lisseur à noyau est utilisé pour rendre disponible les estimations en n'importe quel emplacement d'intérêt. Les paramètres $\sigma(.), \lambda_1(.), \lambda_2(.), \psi(.)$ et $m(.)$ étant considérés comme lentement variable dans l'espace (quasi-stationnarité), l'estimateur à noyau de type Nadaraya-Watson paraît approprié, en plus d'être relativement simple. Cependant d'autres lisseurs peuvent aussi bien être utilisés (polynômes locaux, splines,...).

L'estimateur à noyau de Nadaraya-Watson de $\sigma(.)$ en tout emplacement $\mathbf{x}_0 \in G$ est donné par [Wand and Jones, 1995] :

$$\widetilde{\sigma}(\mathbf{x}_0) = \sum_{k=1}^{m} W_k(\mathbf{x}_0)\widehat{\sigma}(\mathbf{x}_k), \quad W_k(\mathbf{x}_0) = \frac{K(\frac{\mathbf{x}_0 - \mathbf{x}_k}{\delta})}{\sum_{k=1}^{m} K(\frac{\mathbf{x}_0 - \mathbf{x}_k}{\delta})}, \tag{3.19}$$

où $K(.)$ est un noyau sur $\mathbb{R}^p$; $\delta > 0$ est le paramètre de lissage ; $\{\widehat{\sigma}(\mathbf{x}_k), k = 1, \ldots, m\}$ sont les estimations brutes du paramètre $\sigma(.)$ aux points supports $\{\mathbf{x}_k, k = 1, \ldots, m\}$.

On définit de manière similaire les estimateurs à noyau $\widetilde{\lambda}_1(.)$, $\widetilde{\lambda}_2(.)$ et $\widetilde{m}(.)$ en tout emplacement $\mathbf{x}_0 \in G$. Pour le cas spécifique du paramètre d'orientation $\psi(.)$ appartenant à S^p et non à $\mathbb{R}^p$, l'estimateur à noyau en tout emplacement $\mathbf{x}_0 \in G$ est défini via la minimisation du critère :

$$\widetilde{\psi}(\mathbf{x}_0) = \underset{\psi(\mathbf{x}_0) \in S^p}{\arg\min} \sum_{k=1}^{m} W_k(\mathbf{x}_0) d^2(\psi(\mathbf{x}_0), \widehat{\psi}(\mathbf{x}_k)), \tag{3.20}$$

où $\{\widehat{\psi}(\mathbf{x}_k), k = 1, \ldots, m\}$ sont les estimations brutes du paramètre $\psi(.)$ aux points supports $\{\mathbf{x}_k, k = 1, \ldots, m\}$; S^p est l'hyper-sphère en dimension p, de centre l'origine et de rayon unité ; $d(\psi(\mathbf{x}_0), \widehat{\psi}(\mathbf{x}_k))$ est une mesure de distance entre deux orientations. Pour $p = 2$, on peut prendre : $d(\psi(\mathbf{x}_0), \widehat{\psi}(\mathbf{x}_k)) = \min(|\psi(\mathbf{x}_0) - \widehat{\psi}(\mathbf{x}_k)|, |\psi(\mathbf{x}_0) - \widehat{\psi}(\mathbf{x}_k) - \pi|, |\psi(\mathbf{x}_0) - \widehat{\psi}(\mathbf{x}_k) + \pi|)$.

Comme nous l'avons déjà souligné en section 2.3.1 du précédent chapitre, le choix du noyau est relativement peu important par rapport au choix de la valeur du paramètre de lissage. Nos paramètres étant supposés varier régulièrement et lentement à travers le domaine d'intérêt (stationnarité locale), nous choisissons $K(.)$ comme étant un noyau gaussien stationnaire isotrope. La sélection du paramètre de lissage δ est abordée à la section 3.4.3.

3.4.3 Choix des hyper-paramètres

Un point crucial dans l'inférence est la détermination du paramètre de largeur de fenêtre ϵ entrant dans le calcul de l'estimateur non-paramétrique à noyau du variogramme local stationnaire défini en (3.15). La taille du voisinage de stationnarité locale s'exprime en effet en termes du paramètre de largeur de fenêtre. Nous nous intéressons par ailleurs au choix du paramètre de lissage utilisé dans l'interpolation des estimations brutes de paramètres à la section 3.4.2.

Concernant le paramètre de largeur de fenêtre ϵ, il peut être approprié de le choisir subjectivement dans certaines applications, mais en général il est souhaitable de disposer de méthodes pour le choisir automatiquement à partir des données. Nous avons recours à la méthode fondée sur les données (*data driven method*) qu'est la validation croisée. Puisque l'estimation de la structure de dépendance spatiale est rarement un objectif en soi, mais une étape avant le krigeage, il paraît approprié de choisir la largeur de fenêtre qui donne la meilleure erreur quadratique moyenne (MSE, *Mean Square Error*) en validation croisée. Plus explicitement, nous minimisons relativement à ϵ le score de validation croisée défini comme suit [Zhang and Wang, 2010] :

$$MSE(\epsilon) = \frac{1}{n} \sum_{i=1}^{n} \left(Y(\mathbf{s}_i) - \widehat{Y}_{-i}(\mathbf{s}_i; \epsilon) \right)^2, \tag{3.21}$$

où $\widehat{Y}_{-i}(\mathbf{s}_i; \epsilon)$ représente le krigeage à l'emplacement $\mathbf{s}_i$ en utilisant toutes les observations à l'exception de $\{Y(\mathbf{s}_i)\}$. La méthode de prédiction est décrite en détail à la section 3.5.1.

Le choix du paramètre de lissage δ associé au paramètre $\sigma(.)$ est effectué en utilisant le critère de validation croisée défini comme suit [Wand and Jones, 1995] :

$$CV(\delta) = \frac{1}{m}\sum_{k=1}^{m}\left(\frac{\widehat{\sigma}(\mathbf{x}_k) - \widetilde{\sigma}(\mathbf{x}_k)}{1 - W_k(\mathbf{x}_k)}\right)^2, \tag{3.22}$$

où $\{\widehat{\sigma}(\mathbf{x}_k), \widetilde{\sigma}(\mathbf{x}_k), k = 1, \ldots, m\}$ sont respectivement les valeurs estimées brutes et lissées du paramètre $\sigma(.)$ aux points supports $\{\mathbf{x}_k, k = 1, \ldots, m\}$. La fonction de score de validation croisée $CV(\delta)$ est minimisée sur une grille de valeurs pour choisir la valeur optimale.

Théoriquement, la largeur de fenêtre associée à chaque paramètre peut être différente. Nos exemples numériques nous indiquent que se limiter à la même largeur de fenêtre pour tous les paramètres en vue de réduire la charge de calcul, dégrade très peu les performances prédictives. Cette remarque a déjà été soulignée par Zhu and Wu [2010].

3.5 Prédiction

La modélisation et l'estimation de la structure de dépendance spatiale a pour objectif premier la prédiction (krigeage, simulation conditionnelle, ...). Dans cette section, une description du krigeage et de la simulation conditionnelle, basés sur le modèle non-stationnaire de convolution (3.11) est présentée.

3.5.1 Krigeage

Le bénéfice souhaité en utilisant le modèle de covariance non-stationnaire explicite (3.13) est de pouvoir obtenir des prédictions spatiales et variances d'erreur d'estimation plus réalistes que celles basées sur une structure de covariance stationnaire inappropriée.

Soit $C^{NS}(.,.)$ la covariance non-stationnaire de la fonction aléatoire Y et $m(.)$ sa moyenne. Étant donné le vecteur $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))^T$ aux n sites d'observation $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset G$, la prédiction de la valeur inconnue de Y à l'emplacement non-échantillonné $\mathbf{s}_0 \in G$ est donnée par le prédicteur linéaire optimal :

$$\widehat{Y}(\mathbf{s}_0) = m(\mathbf{s}_0) + \sum_{i=1}^{n}\eta_i(\mathbf{s}_0)(Y(\mathbf{s}_i) - m(\mathbf{s}_i)). \tag{3.23}$$

Le vecteur de poids de krigeage $\boldsymbol{\eta} = [\eta_i(\mathbf{s}_0)]$ et la variance de krigeage associée $Q(\mathbf{s}_0)$ sont donnés par :

$$\boldsymbol{\eta} = \mathbf{C}^{-1}\mathbf{C}_0 \quad \text{et} \quad Q(\mathbf{s}_0) = \sigma^2(\mathbf{s}_0) - \mathbf{C}_0^T\mathbf{C}^{-1}\mathbf{C}_0. \tag{3.24}$$

où $\mathbf{C}_0 = [C^{NS}(\mathbf{s}_i, \mathbf{s}_0)]$; $\mathbf{C} = [C^{NS}(\mathbf{s}_i, \mathbf{s}_j)]$.

3.5.2 Simulation conditionnelle

L'estimateur obtenu par krigeage bien qu'optimal lisse la réalité (propriété du krigeage). Ainsi, les détails présents en réalité ne peuvent pas être reproduits à partir des seuls points de données. Or, dans certaines situations, il importe non pas d'obtenir la meilleure précision, mais de reproduire la variabilité spatiale. On souhaite ainsi disposer d'une réalisation qui présente le même degré de variabilité que le phénomène réel et qui coïncide avec les observations aux points de données. Pour ce faire, on a recours aux simulations conditionnelles. Une simulation conditionnelle a les mêmes caractéristiques structurales que le phénomène réel et passe par les points de données. Elle n'est pas la réalité qui reste inconnue, ni le meilleur estimateur de celle-ci, le meilleur estimateur linéaire au sens de l'erreur quadratique moyenne restant le krigeage. Toutefois, elle représente une variante plausible de la réalité.

Nous considérons la fonction aléatoire Y gaussienne, de moyenne $m(.)$ et de covariance $C^{NS}(.,.)$. On désire simuler en un grand nombre d'emplacements une fonction aléatoire gaussienne de même moyenne et covariance que Y, et telle que la réalisation passe par les valeurs observées $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$. Ceci peut se faire sur la base d'une simulation non-conditionnelle de la fonction aléatoire Y comme suit [Lantuéjoul, 2002] :

Algorithme 3.5.1 Simulation conditionnelle

1. réaliser une simulation non-conditionnelle $\{X(\mathbf{s}), \mathbf{s} \in G\}$ de la fonction aléatoire Y ;
2. effectuer un krigeage simple de $\{X(\mathbf{s}) - Y(\mathbf{s})\}$ à partir de ses valeurs prises aux points de données $\{\mathbf{s}_i, i = 1, \ldots, n\}$, à l'aide de $m(.)$ et $C^{NS}(.,.)$;
3. additionner la simulation non-conditionnelle et le résultat du krigeage.

Nous avons $Y(\mathbf{x}) = m(\mathbf{x}) + \sigma(\mathbf{x})Z(\mathbf{x}), \forall \mathbf{x} \in G$. Ainsi, pour simuler la fonction aléatoire gaussienne Y, il est suffisant de savoir comment simuler Z. Z est une fonction aléatoire gaussienne centrée réduite et de fonction de corrélation non-stationnaire $R^{NS}(.,.)$. La simulation de Z peut se faire en utilisant une version propagative de l'échantillonneur de Gibbs proposée par Lantuéjoul and Desassis [2012]. Cet algorithme permet de simuler un vecteur gaussien en un grand nombre d'implantations (comparativement aux algorithmes génériques disponibles tels que la méthode de Cholesky ou l'échantillonneur de Gibbs) et sans reposer sur une hypothèse markovienne (elle n'a pas besoin d'avoir une matrice de précision creuse). En effet, les algorithmes de simulation par décomposition de Cholesky ou échantillonneur de Gibbs ne peuvent pas être implémentés lorsqu'on souhaite simuler en un grand nombre de points, la matrice de covariance ne pouvant pas s'inverser numériquement. L'algorithme proposé par Lantuéjoul and Desassis [2012] ne nécessite ni inversion ni factorisation de la matrice de covariance. Par ailleurs, les méthodes de

simulation telle que la méthode spectrale ou celle des bandes tournantes ne sont pas adaptées au cas non-stationnaire ; la représentation qui sous-tend ces méthodes repose en effet sur l'hypothèse de stationnarité. Formellement l'algorithme proposé par Lantuéjoul and Desassis [2012] se décline comme suit :

Soit $\mathbf{Z} = (Z_a, a \in A \subset G)$ un vecteur gaussien standardisé, de matrice de covariance $\mathbf{C} = (C_{ab}, a, b \in A)$. L'idée sous-jacente est qu'au lieu de simuler le vecteur $\mathbf{Z}$, on peut simuler le vecteur $\mathbf{C}^{-1}\mathbf{Z}$. Ce dernier est bien un vecteur gaussien de matrice de covariance $\mathbf{C}^{-1}$. En remarquant que l'inverse de $\mathbf{C}^{-1}$ est précisément $\mathbf{C}$, l'échantillonneur de Gibbs est appliqué à $\mathbf{C}^{-1}\mathbf{Z}$ et $\mathbf{Z}$ est mis à jour en conséquence. On obtient ainsi, l'algorithme suivant :

Algorithme 3.5.2 Simulation par Gibbs propagatif

1. poser $z_A^c = \mathbf{0}$;
2. générer $a \sim \mathcal{U}(A)$;
3. générer $z_a^n \sim \mathcal{N}(0,1)$ et prendre $z_b^n = z_b^c + C_{ab}(z_a^n - z_a^c)$ pour chaque $b \neq a$;
4. prendre $z_A^c = z_A^n$ et aller en 2.

Ainsi, à chaque étape de l'algorithme, une valeur gaussienne est affectée à un pivot aléatoirement, puis propagée aux autres composantes. Cet algorithme possède en outre de bonnes propriétés de convergence (voir Lantuéjoul and Desassis [2012]).

3.6 Illustration

Considérons le modèle défini en (3.11), $Y(\mathbf{x}) = m(\mathbf{x}) + \sigma(\mathbf{x})Z(\mathbf{x})$, $\forall \mathbf{x} \in G$. On se focalise sur la non-stationnarité de la structure de dépendance spatiale. Ainsi, la fonction moyenne est fixée égale à zéro : $m(\mathbf{x}) = 0$, $\forall \mathbf{x} \in G$. La fonction d'écart-type $\sigma(.)$ est prise égale à $\sigma(\mathbf{x}) = \pi^{-\frac{1}{2}}|\mathbf{\Sigma_x}|^{-\frac{1}{4}}$, $\forall \mathbf{x} \in G$. On prend comme structure de corrélation non-stationnaire celle de Matérn de paramètre de régularité $\nu = 1$:

$$R^{NS}(\mathbf{x}, \mathbf{y}) = \phi_{\mathbf{xy}}\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}K_1\left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}\right), \forall (\mathbf{x}, \mathbf{y}) \in G \times G.$$

Les paramètres $\lambda_1(.), \lambda_2(.)$ et $\psi(.)$ de la fonction d'anisotropie $\mathbf{\Sigma}(.)$ sont définis comme suit : $\psi(\mathbf{x}) = \left|\tan^{-1}\left(\frac{x_2-0,5}{x_1-0,5}\right)\right|$, $\lambda_2(\mathbf{x}) = 0,05$ et $\frac{\lambda_1(\mathbf{x})}{\lambda_2(\mathbf{x})} = 2 + \left|\tan^{-1}\left(\frac{x_2-0,5}{x_1-0,5}\right)\right|$, où $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$.

On simule une fonction aléatoire gaussienne centrée, de variance $\sigma^2(.)$ et de corrélation $R^{NS}(.,.)$, en 200×200 points d'une grille régulière de $G = [0,1]^2$. À partir de cette réalisation, des échantillons d'apprentissage (400 points) et de validation (1024 points)

sont constitués. Les figures 3.1a et 3.1b présentent la réalisation obtenue et les données d'apprentissage extraites. Comme on peut le voir, la variabilité spatiale des caractéristiques structurales telles que l'anisotropie géométrie sont clairement visibles.

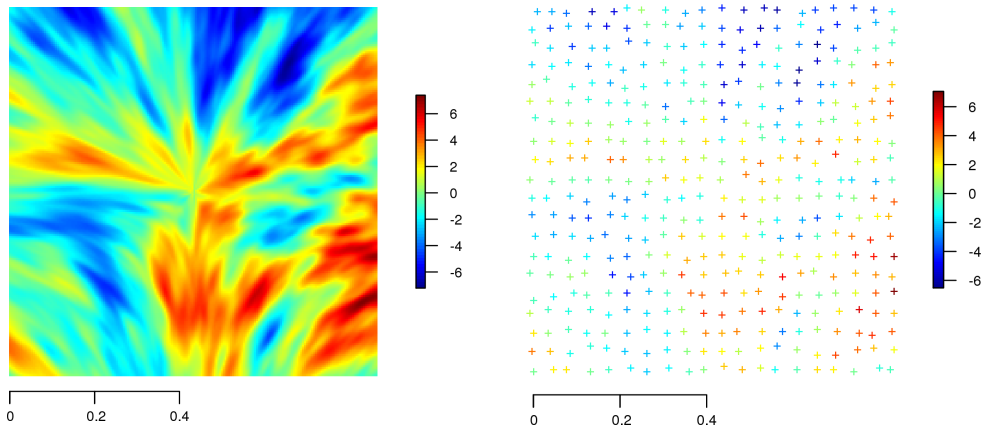


FIGURE 3.1 – Modèle de convolution : (a) données exhaustives ; (b) données d'apprentissage.

Les figures 3.2a et 3.2b montrent respectivement les estimations brutes des paramètres d'anisotropie et de variance en 400 points supports du domaine d'observation. Les matrices de covariance (définies positives) ont une interprétation géométrique attrayante : elles peuvent être identifiées uniquement par une ellipsoïde. Ainsi, la fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports est représentée par des ellipses comme le montre la figure 3.2a. Les ellipses ont été mises à échelle pour une bonne visualisation. Sur la figure 3.2a, on voit que la méthode proposée permet de bien retrouver les anisotropies locales variables, ces dernières ne pouvant pas être révélées par une approche stationnaire.

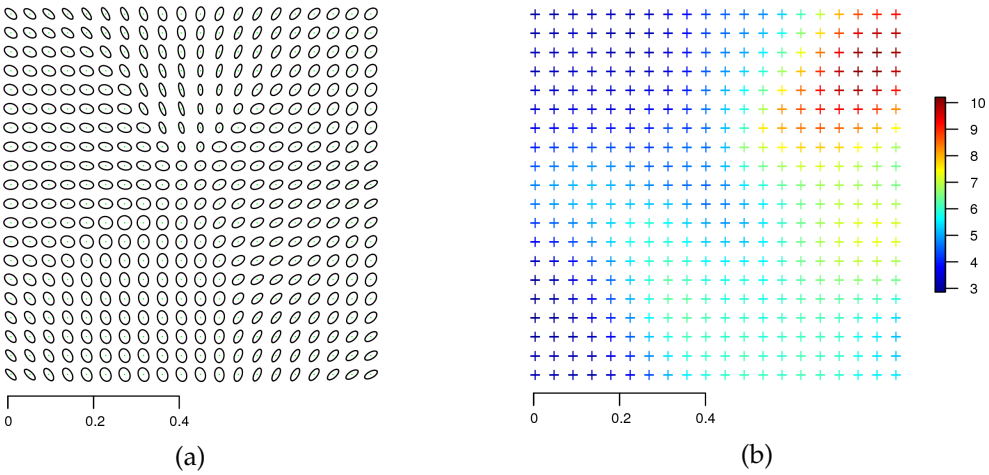
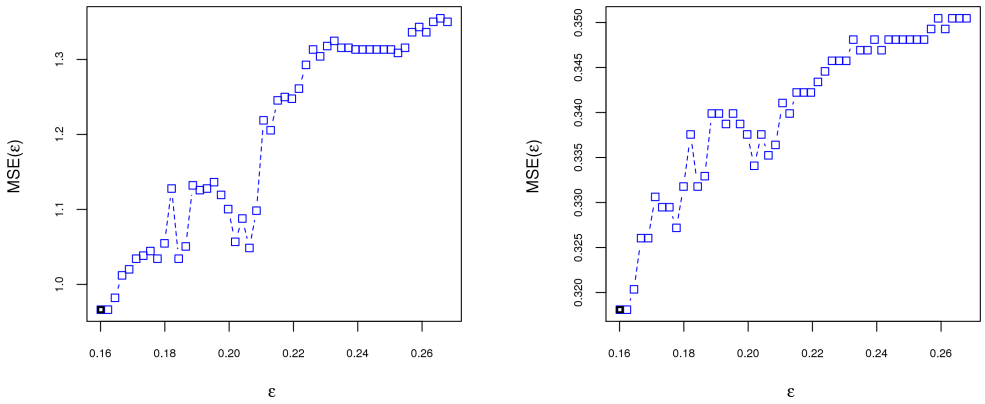


FIGURE 3.2 – Méthode de convolution : (a) fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports ; (b) fonction de variance estimée $\widehat{\sigma}^2(.)$ aux points supports.

La procédure de sélection de la largeur de fenêtre décrite en section 3.4.3 permet d'obtenir les fonctions de score en validation croisée et externe présentées aux figures 3.3a et 3.3b. En validation croisée, la largeur de fenêtre optimale associée à l'estimateur du variogramme local est obtenue pour $\epsilon = 0,16$. Cette valeur optimale est la même en validation externe. Ainsi, la validation croisée semble ne pas biaiser le choix de l'hyper-paramètre ϵ. Comme nous l'avons déjà souligné en section 2.5.2 (pour la méthode de déformation d'espace), ceci peut s'expliquer en raison de la corrélation existant entre les données d'apprentissage et de validation.



(a) (b)

FIGURE 3.3 – Méthode de convolution : fonction de score $MSE(\epsilon)$ en validation croisée (a) et externe (b).

Les figures 3.4a, 3.4b et 3.4c montrent les cartes des paramètres lissés (variance, ratio d'anisotropie[2] et azimut) sur le domaine d'intérêt. Suivant la procédure de sélection décrite en section 3.4.3, la valeur optimale du paramètre de lissage associée au noyau gaussien correspond à $\delta = 0,03$.

2. Le ratio d'anisotropie est défini comme le rapport de la petite portée sur la grande portée. Ainsi, plus il est faible, plus l'anisotropie est marquée.

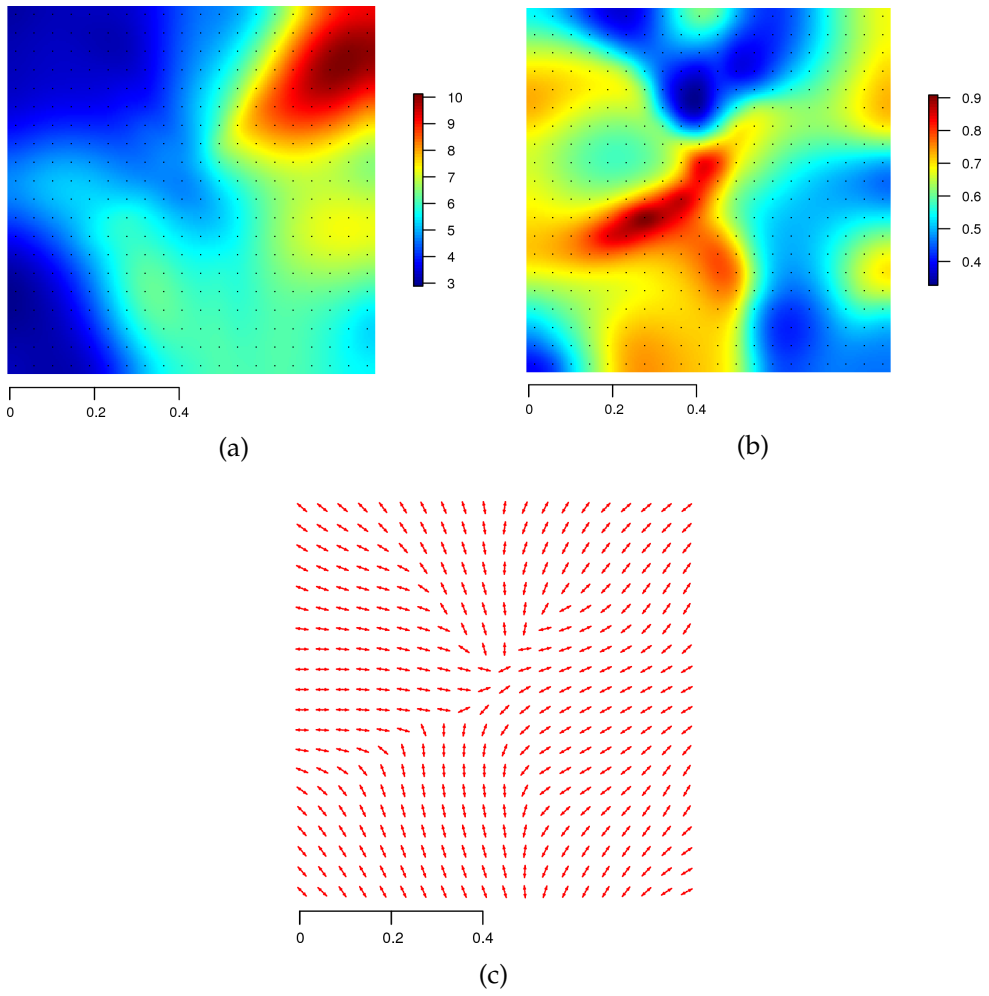


FIGURE 3.4 – Méthode de convolution : paramètres lissés sur le domaine des observations : (a) variance, (b) ratio d'anisotropie, (c) azimut.

Une représentation en quelques points de la structure de covariance pour les modèles stationnaire et non-stationnaire estimés est donnée aux figures 3.6a et 3.6b. Le modèle stationnaire estimé à partir des données est une covariance stationnaire isotrope de Matérn, de paramètre de régularité 1. Alors que la structure de covariance stationnaire isotrope estimée est la même en tous les points, celle non-stationnaire estimée change d'un emplacement à un autre.

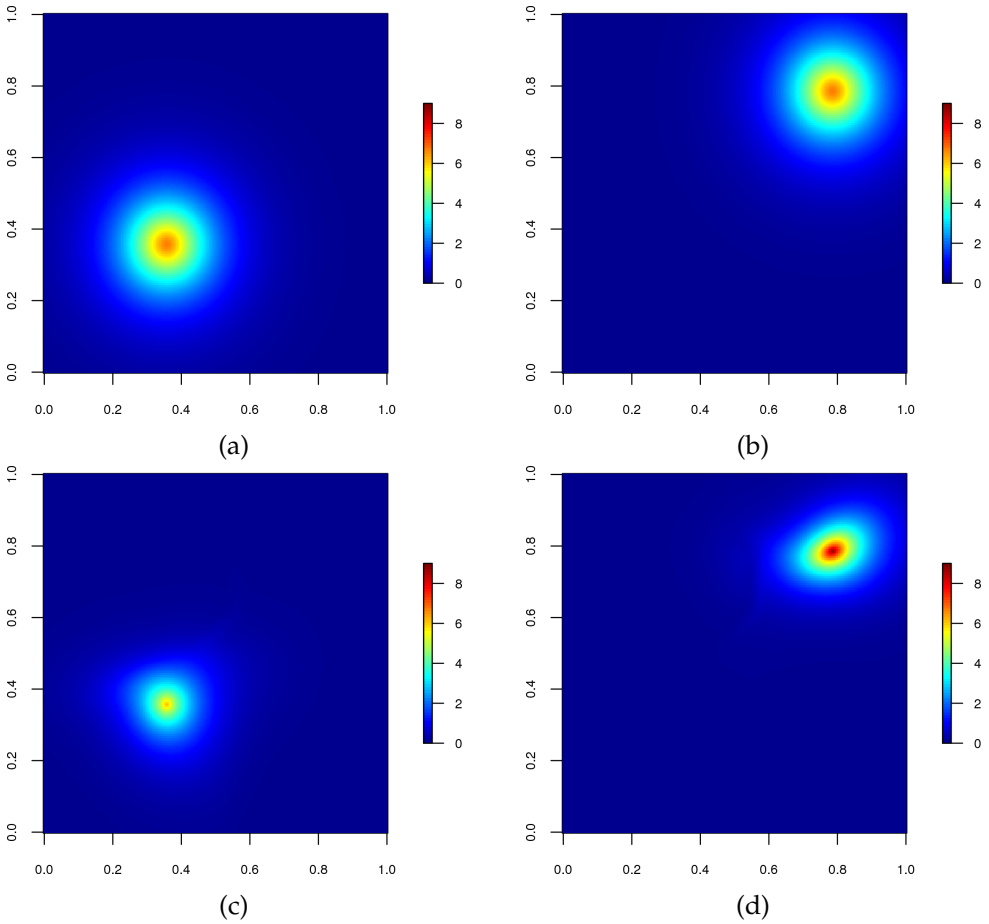


FIGURE 3.5 – Méthode de convolution : covariance d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d).

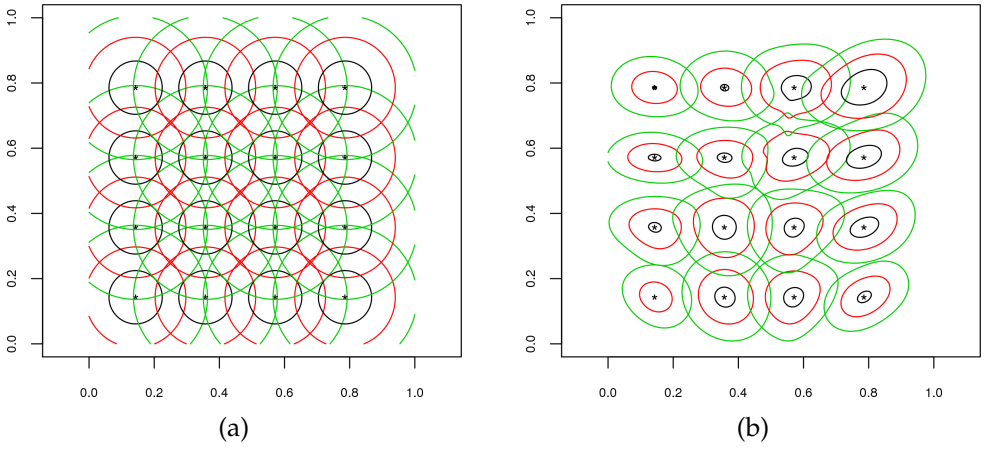


FIGURE 3.6 – Méthode de convolution : contours de la covariance en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 4 (noir), 2 (rouge) et 1 (vert).

Les résultats d'interpolation par krigeage ordinaire pour les modèles stationnaire et non-stationnaire estimés sont donnés à la figure 3.7. À première vue, les cartes krigées correspondant aux deux modèles semblent être proches (figures 3.7a et 3.7b). Toutefois, nous observons que l'approche non-stationnaire reproduit mieux certains détails qui échappent à l'approche stationnaire (grâce aux fortes anisotropies locales). Les différences entre cartes d'écart-type de krigeage associées à chacune des méthodes sont plus prononcées (figures 3.7c et 3.7d), l'approche non-stationnaire prenant en compte les variations de variabilité.

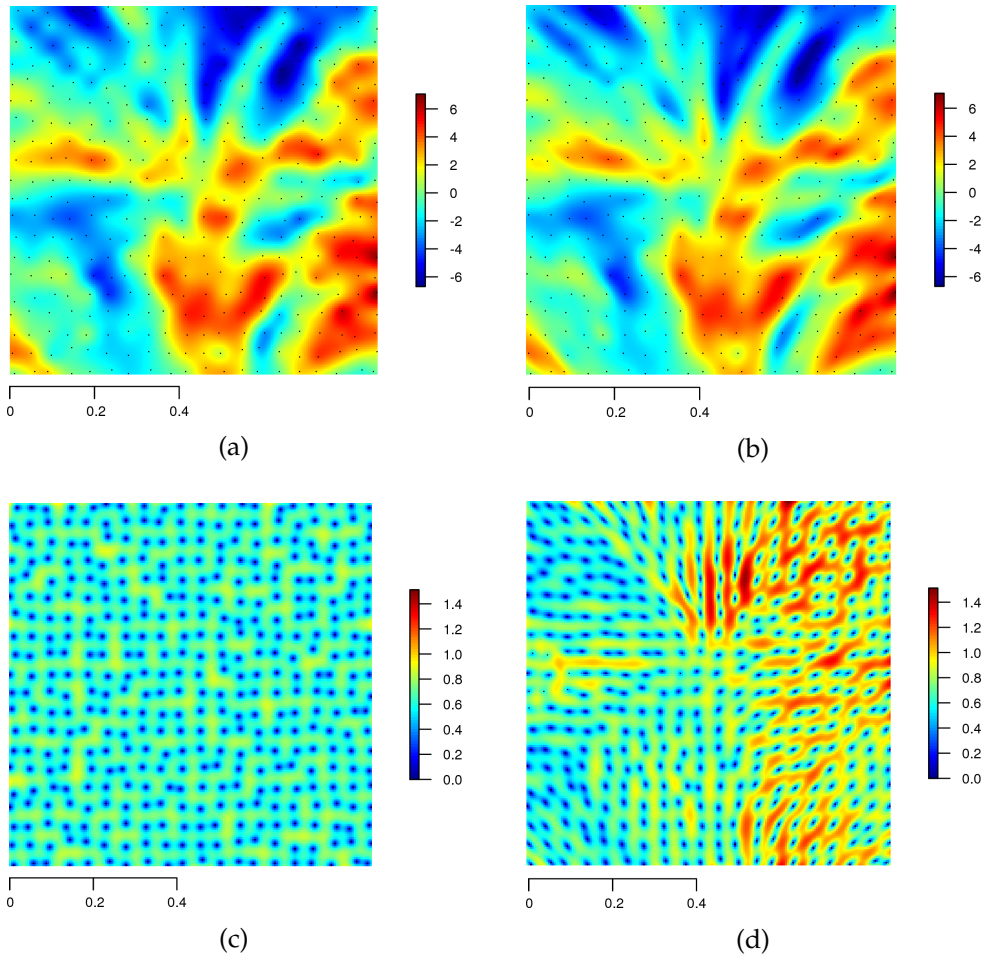


FIGURE 3.7 – Méthode de convolution : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; cartes d'écart-type de krigeage correspondantes (c, d).

Les résultats de la validation externe sur un échantillon de 1024 observations sont repris dans le tableau 3.1. Les scores de prédiction présentés sont définis en annexe B. Nous observons globalement que la méthode par convolution présente des meilleures performances que la stationnaire. Par exemple, en terme de RMSE (*Root Mean Square Error*), on note une amélioration globale de 10% par rapport à l'approche stationnaire.

	Stationnaire	Non-stationnaire
MAE	0,46	0,42
RMSE	0,62	0,56
NMSE	0,95	0,54
LogS	1758	1734
CRPS	0,60	0,58

TABLEAU 3.1 – Méthode de convolution : validation externe sur un échantillon de 1024 observations.

3.7 Conclusion

Nous avons présenté au cours de ce chapitre un nouveau modèle pour les fonctions aléatoires non-stationnaires de second ordre. Ce modèle généralise le modèle classique de convolution et permet en particulier de retrouver les classes explicites de covariances non-stationnaires existantes. Par ailleurs, nous nous sommes intéressés au problème de l'inférence de ces classes. La méthodologie d'inférence proposée basée sur l'hypothèse de quasi-stationnarité permet de traiter dans un cadre unique les différents types de non-stationnarité : moyenne, variance et covariance. Sur un exemple de données simulées, l'approche proposée donne de meilleurs résultats de prédiction qu'une approche stationnaire. Elle est capable de retrouver les principales caractéristiques structurales de la régionalisation, notamment l'anisotropie géométrique locale variable. Jusqu'à présent, les méthodes non-stationnaires proposées n'ont pas été éprouvées sur des données réelles. Le prochain chapitre sera ainsi consacré à la mise en œuvre des méthodes proposées sur des cas concrets.

Chapitre 4

Synthèse - trois exemples d'application

"Models are to be used, but not to be believed." H. Theil.

Sommaire

4.1 Introduction **115**
4.2 Application 1 : données de sol **116**
4.3 Application 2 : données de pluie **127**
4.4 Application 3 : données de topographie **138**
4.5 Conclusion **150**

4.1 Introduction

Les deux précédents chapitres avaient une portée essentiellement méthodologique. Nous avons développé des méthodes non-stationnaires et nous les avons illustrées sur des données synthétiques. Le présent chapitre a quant à lui une portée pratique. Il illustre la démarche des méthodes non-stationnaires proposées sur des ensembles de données réelles couvrant divers domaines : données de sol, de pluie et de topographie. L'objectif escompté est que l'estimation de la structure de dépendance spatiale via une méthode non-stationnaire se traduise par une amélioration de la prédiction spatiale, ainsi que de l'incertitude associée.

Les sections suivantes présentent pour chaque jeu de données les résultats de l'inférence de chacune des méthodes non-stationnaires proposées, ainsi que leur application aux problèmes de prédiction spatiale et de simulation conditionnelle. Par ailleurs, les performances prédictives en validation externe sont présentées et comparées avec celles de l'approche stationnaire. Une représentation graphique de la structure de dépendance spatiale est également donnée.

4.2 Application 1 : données de sol

On s'intéresse à un ensemble de données issu d'une campagne de mesures (gamma-ray) de concentration en potassium (en coup par seconde, cps) dans la région de Hunter Valley, Nouvelle-Galles du Sud, Australie [McBratney and Minasny, 2013]. On dispose d'un échantillon d'apprentissage (537 observations) qui sert à calibrer le modèle et d'un échantillon de validation (1000 observations) qui sert uniquement à évaluer les performances de prédiction du modèle.

Méthode de déformation d'espace

Les figures 4.1a et 4.1b montrent respectivement les points de données dans l'espace géographique et leur image dans l'espace déformé. Ce résultat est obtenu à partir d'une configuration de 125 points supports. On peut observer que la déformation contracte l'espace géographique dans la région nord-est tandis qu'elle l'étire dans les autres régions. Ainsi, la région nord-est correspond à une région où la variable présente une corrélation spatiale relativement élevée. Comme nous le verrons par la suite, cette façon qu'a la déformation d'opérer est en adéquation avec la carte de variance donnée par la méthode de convolution (figure 4.9b). En effet, la région contractée correspond à une région de faible variance, tandis que celles étirées correspondent à des régions de forte variance. Les figures 4.1c et 4.1d représentent respectivement le variogramme stationnaire isotrope dans les espaces géographique et déformé [1]. La modélisation non-stationnaire par déformation d'espace et celle stationnaire conduisent respectivement aux modèles suivants :

$$\widehat{\gamma}_0(\|\mathbf{h}\|) = 62 \times \mathrm{Exp}_{101}(\|\mathbf{h}\|) + 102 \times \mathrm{Sph}_{428}(\|\mathbf{h}\|), \tag{4.1}$$

$$\widehat{\gamma}_1(\|\mathbf{h}\|) = 46 \times \mathrm{Nug}(\|\mathbf{h}\|) + 21 \times \mathrm{Exp}_{187}(\|\mathbf{h}\|) + 89 \times \mathrm{Sph}_{234}(\|\mathbf{h}\|), \tag{4.2}$$

où $\widehat{\gamma}_0(.)$ est un variograme stationnaire isotrope gigogne (exponentiel et sphérique) de variance totale 164 cps^2 et $\widehat{\gamma}_1(.)$ est un variogramme stationnaire isotrope gigogne (effet pépite, exponentiel et sphérique) de variance totale 156 cps^2.

1. L'approche stationnaire n'a pas détecté d'anisotropie géométrique globale.

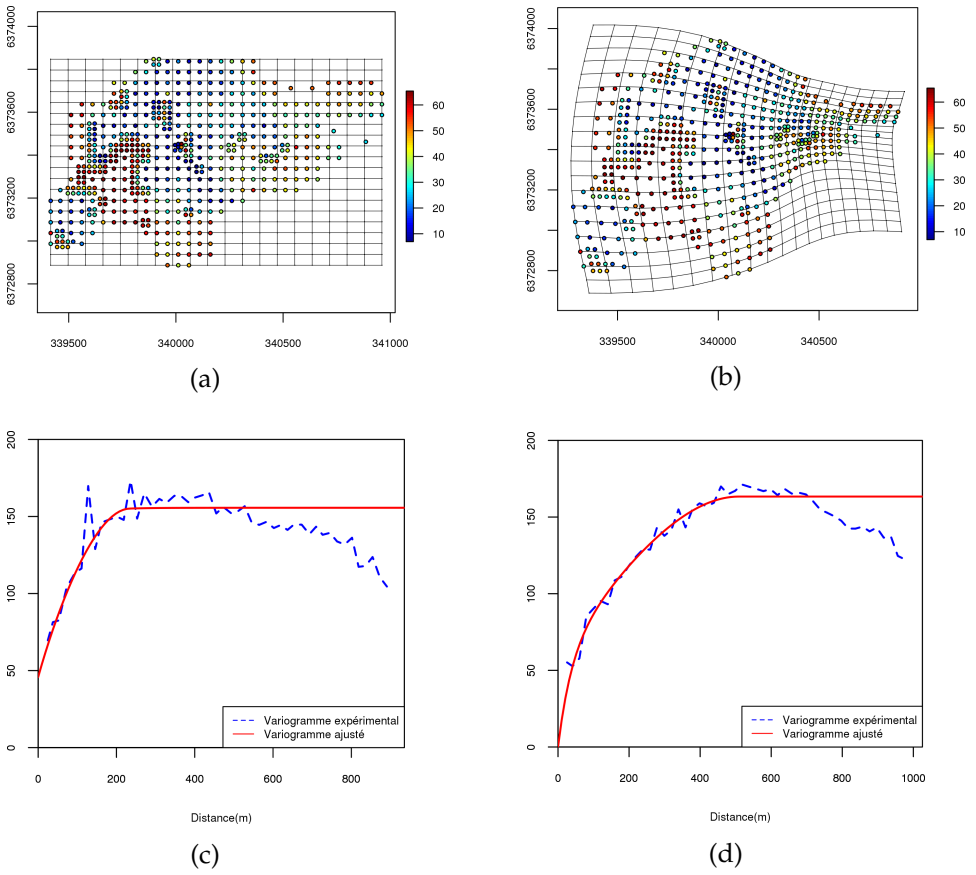


FIGURE 4.1 – Méthode de déformation d'espace : (a, c) données d'apprentissage dans l'espace géographique et modèle de variogramme stationnaire estimé ; (b, d) données d'apprentissage dans l'espace déformé et modèle de variogramme transformé estimé.

Les fonctions de score $CV(\lambda)$ et $MSE(\lambda, \omega)$ en validation croisée permettant de sélectionner les hyper-paramètres sont représentées aux figures 4.2a et 4.2b. Sur la figure 4.2a, on peut noter que les valeurs modérées et grandes de la largeur de la fenêtre λ tendent à donner des scores $CV(\lambda)$ faibles. Les valeurs optimales des hyper-paramètres du point de vue de $MSE(\lambda, \omega)$ sont $\lambda = 664$ m et $\omega = 0,275$. Ces valeurs sont corroborées par la validation externe. En effet, la figure 4.2d présente la fonction de score $MSE(\lambda, \omega)$ calculée plutôt sur l'échantillon de validation. Sur cet échantillon de validation, les valeurs optimales sont très proches des valeurs précédentes. En observant les figures 4.2b et 4.2d, on constate que les fonctions de score en validation croisée et externe sont assez similaires. Comme nous l'avions déjà souligné dans l'exemple simulé en section 2.5.2, la validation croisée semble ne pas produire de biais dans la sélection des hyper-paramètres. Ceci en raison de la corrélation existante entre les données d'apprentissage et de validation. Le diagramme de Shepard de la solution NMDS (espace déformé) est présenté à la figure 4.2c. La valeur finale du *stress* correspondant à l'espace déformé est de 6%, ce que nous pouvons considérer comme satisfaisant du point de vue du tableau 2.1.

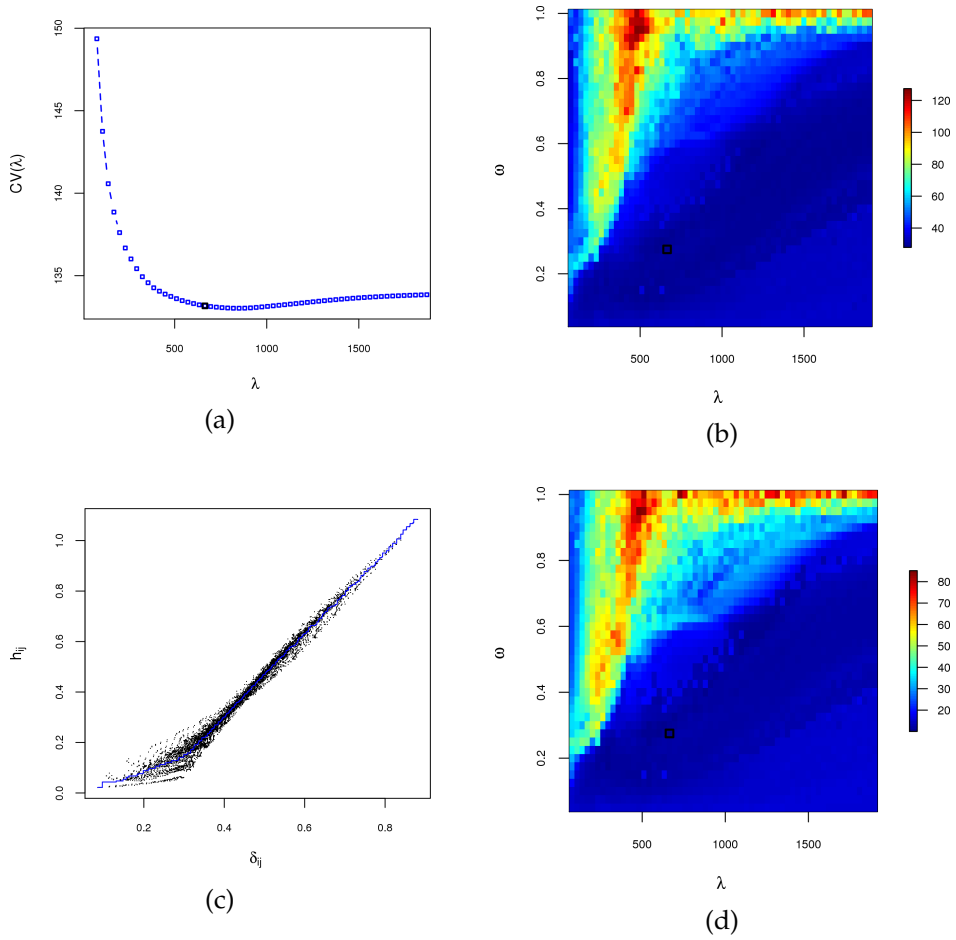


FIGURE 4.2 – Méthode de déformation d'espace : (a) fonction score $CV(\lambda)$ de validation croisée ; (c) diagramme de Shepard de la solution NMDS ; fonction score $MSE(\lambda, \omega)$ en validation croisée (b) et externe (d).

Une visualisation du variogramme en certains points pour les modèles stationnaire et non-stationnaire estimés est présentée aux figures 4.3 et 4.4. On peut ainsi voir comment la structure de dépendance spatiale non-stationnaire change de forme d'un endroit à un autre comparativement à son homologue stationnaire.

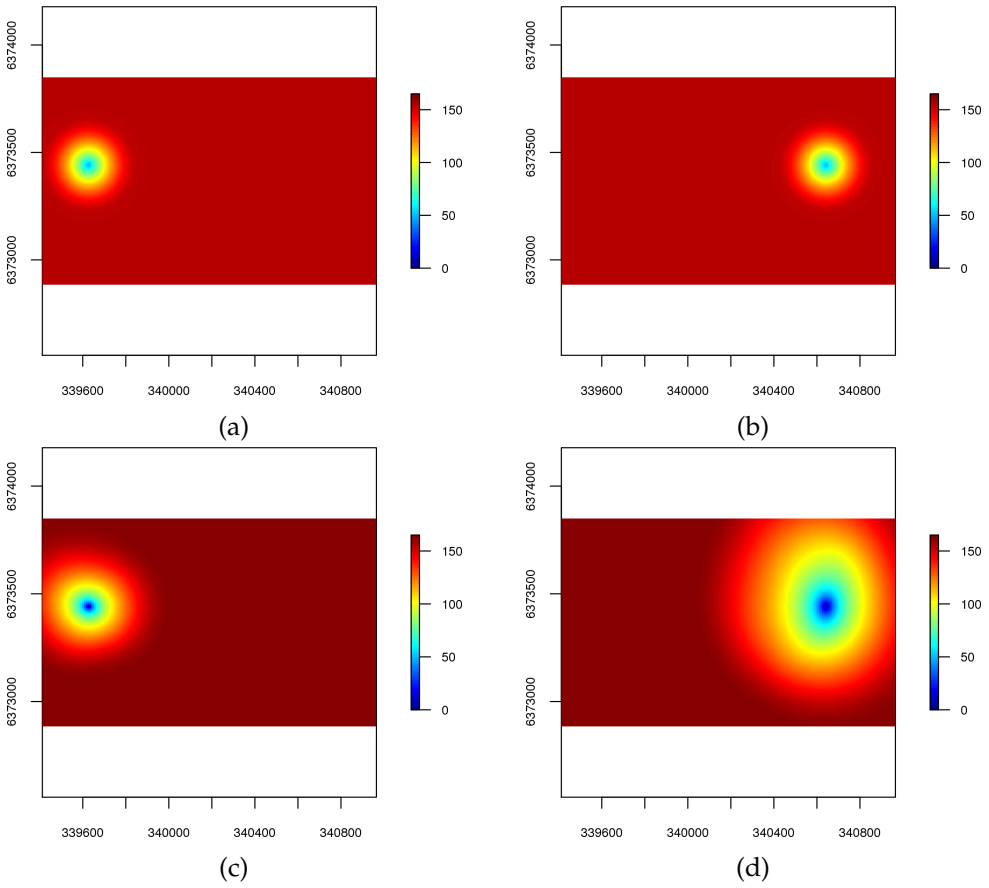


FIGURE 4.3 – Méthode de déformation d'espace : variogramme d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d).

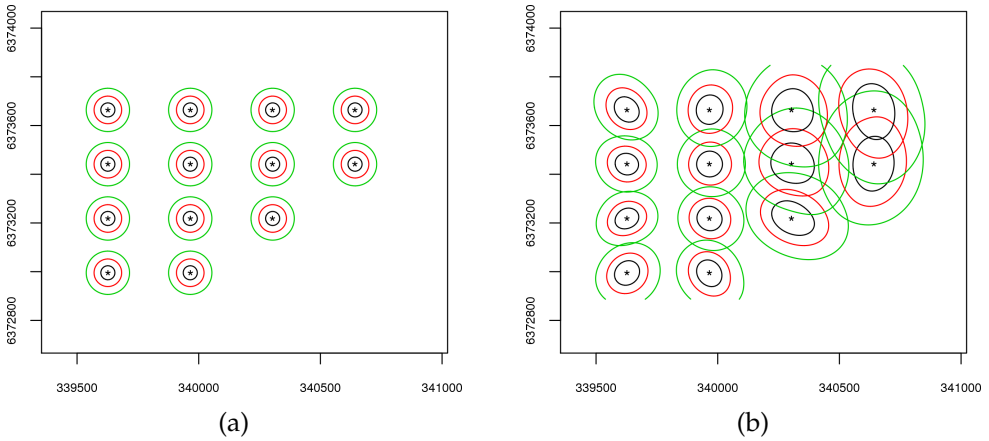


FIGURE 4.4 – Méthode de déformation d'espace : contours du variogramme en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 70 (noir), 90 (rouge) et 110 (vert).

Les prédictions par krigeage pour les modèles stationnaire et non-stationnaire estimés sont présentées à la figure 4.5. L'aspect général des cartes de valeurs prédites et d'écart-type de krigeage associées à chaque modèle diffère. En particulier, la méthode de déformation d'espace semble prendre en compte certaines caractéristiques locales de la régionalisation que l'approche stationnaire est incapable de récupérer. La figure 4.6 nous montre les simulations conditionnelles (sous hypothèse gaussienne) selon le modèle non-stationnaire estimé.

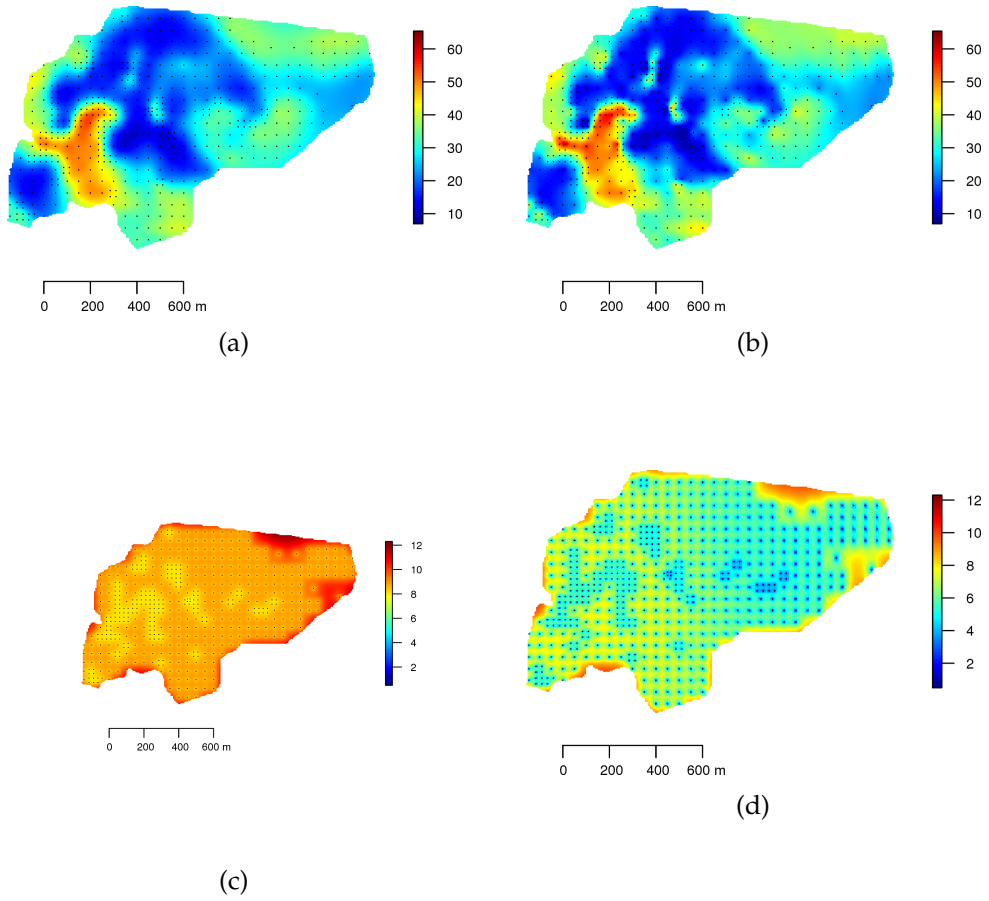
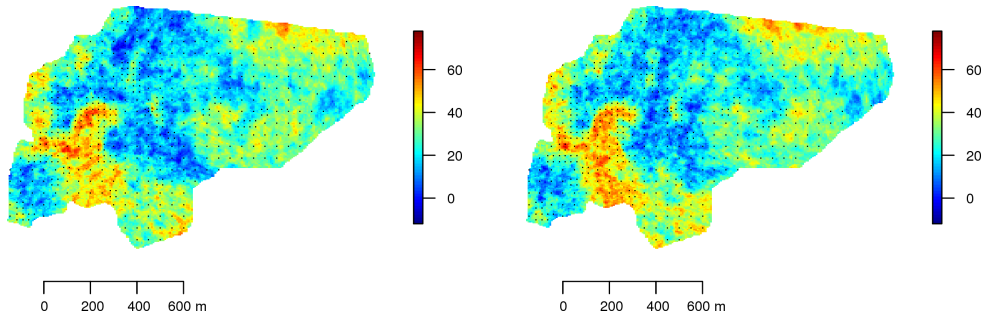
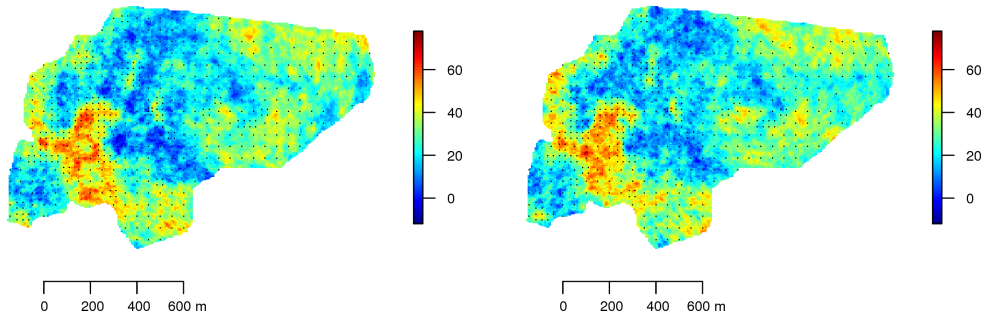


FIGURE 4.5 – Méthode de déformation d'espace : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; (c, d) cartes d'écart-type de krigeage associées.



(a) Simulation #1 (b) Simulation #2



(c) Simulation #3 (d) Simulation #4

FIGURE 4.6 – Méthode de déformation d'espace : simulations conditionnelles.

Méthode de convolution

Les estimations brutes des paramètres $m(.)$, $\sigma^2(.)$ et $\mathbf{\Sigma}(.)$ aux points supports sont présentées respectivement aux figures 4.7b, 4.7c et 4.7d. Elles sont basées sur une structure de covariance non-stationnaire exponentielle (exemple 3.7). Sur la base de ces estimations, la non-stationnarité dans les données est assez visible. Notamment sur la figure 4.7d où l'on peut voir clairement que l'azimut varie spatialement. Ces effets directionnels sont nettement visibles sur les données (figure 4.7a). Notons que l'approche stationnaire n'a pas détecté d'anisotropie géométrique globale.

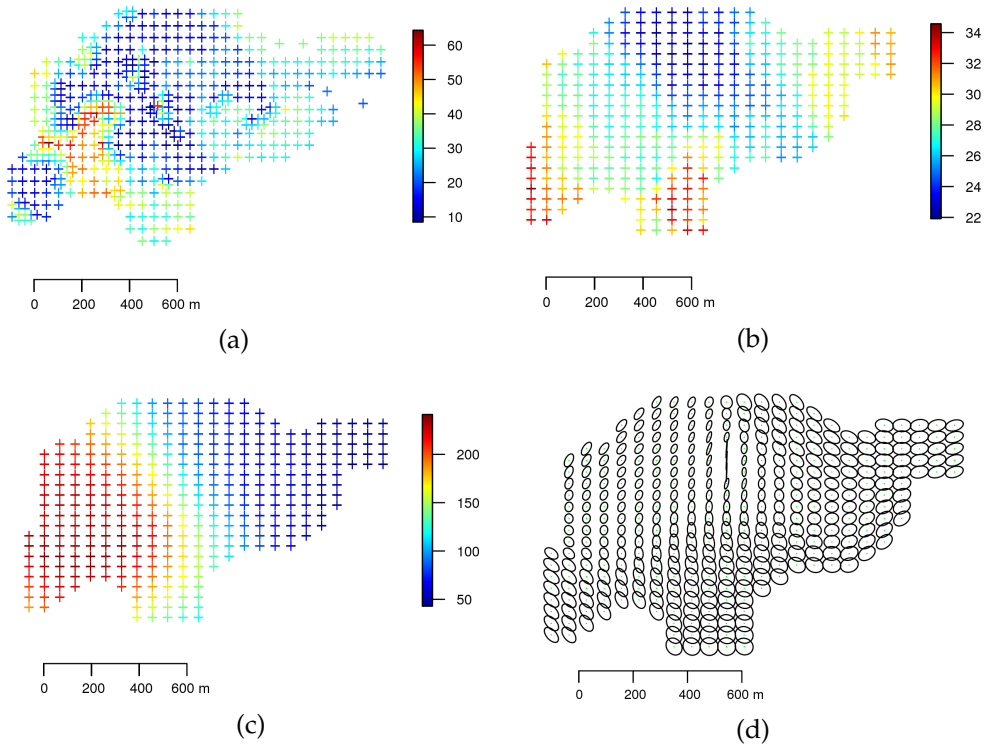
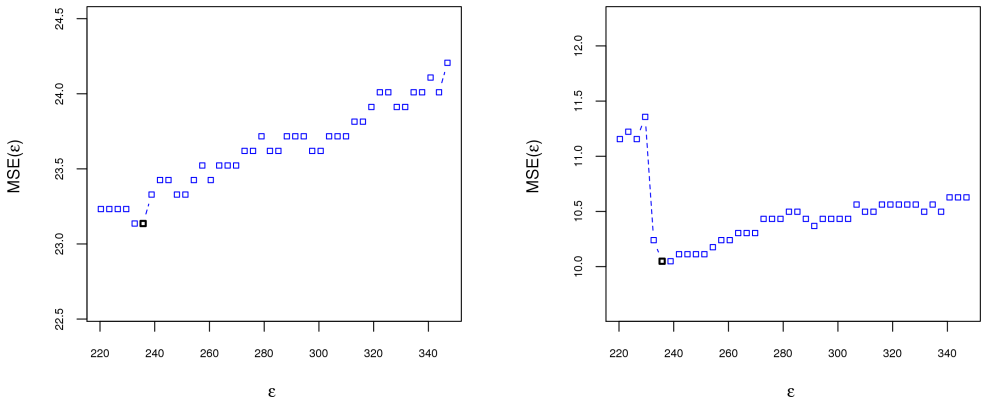


FIGURE 4.7 – Méthode de convolution : (a) données d'apprentissage ; (b) fonction moyenne estimée $\widehat{\mathrm{m}}(.)$ aux points supports ; (c) fonction variance estimée $\widehat{\sigma}^2(.)$ aux points supports ; (d) fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports.

La figure 4.8a montre la fonction de score $MSE(\epsilon)$ en validation croisée. Suivant la procédure de sélection présentée à la section 3.4.3, la largeur de fenêtre associée à l'estimateur du variogramme local vaut $\epsilon = 234$ m. Cette valeur est en accord avec la fonction de score $MSE(\epsilon)$ en validation externe présentée à la figure 4.8b. La figure 4.9 montre les cartes des paramètres lissés sur le domaine d'intérêt : moyenne, variance, ratio d'anisotropie [2] et azimut. Suivant la procédure de sélection décrite en section 3.4.3, la valeur optimale du paramètre de lissage associé au noyau gaussien correspond à $\delta = 30$ m.

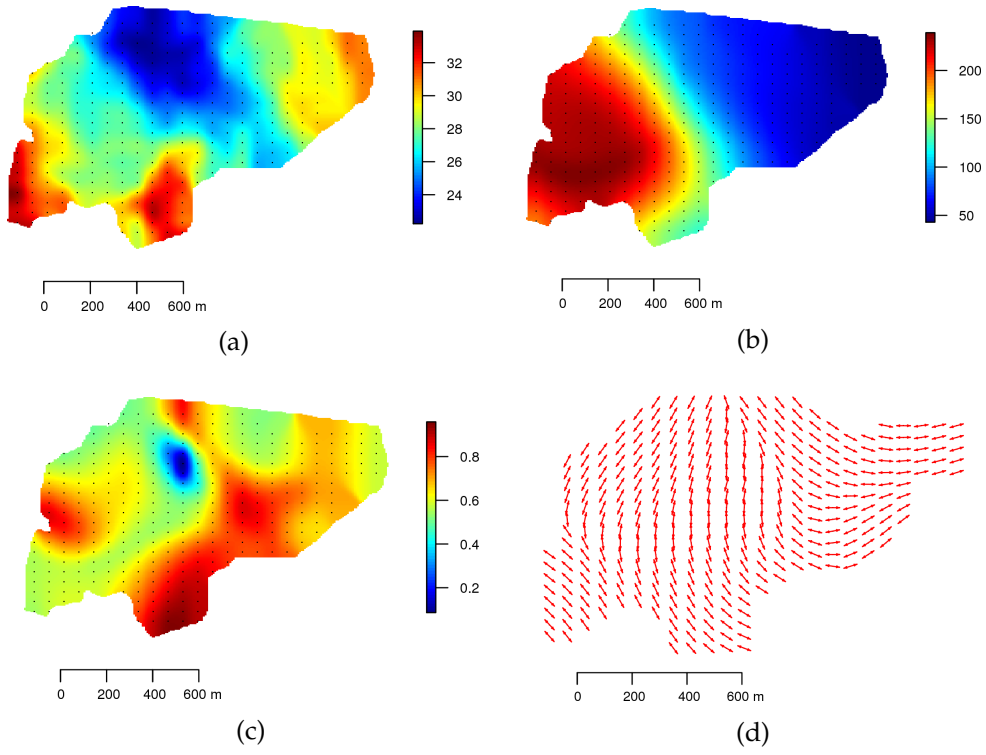
La méthode de convolution est en accord avec celle par déformation d'espace, en ce sens que la carte de variance produite à la figure 4.9b est cohérente avec l'action de la déformation présentée à la figure 4.1b. En effet, la région nord-est contractée correspond à une région de faible variance. Tandis que les autres régions étirées correspondent à des régions de forte variance.

2. Le ratio d'anisotropie est défini comme le rapport de la petite portée sur la grande portée. Ainsi, plus il est faible, plus l'anisotropie est marquée.



(a) (b)

FIGURE 4.8 – Méthode de convolution : fonction de score $MSE(\epsilon)$ en validation croisée (a) et externe (b).



(a) (b) (c) (d)

FIGURE 4.9 – Méthode de convolution : paramètres lissés sur le domaine des observations : (a) moyenne, (b) variance, (c) ratio d'anisotropie, (d) azimut.

Les figures 4.10 et 4.11 permettent de visualiser les structures de covariance non-stationnaire et stationnaire estimées en quelques emplacements. La variation de la structure de covariance non-stationnaire d'un site à un autre est assez visible.

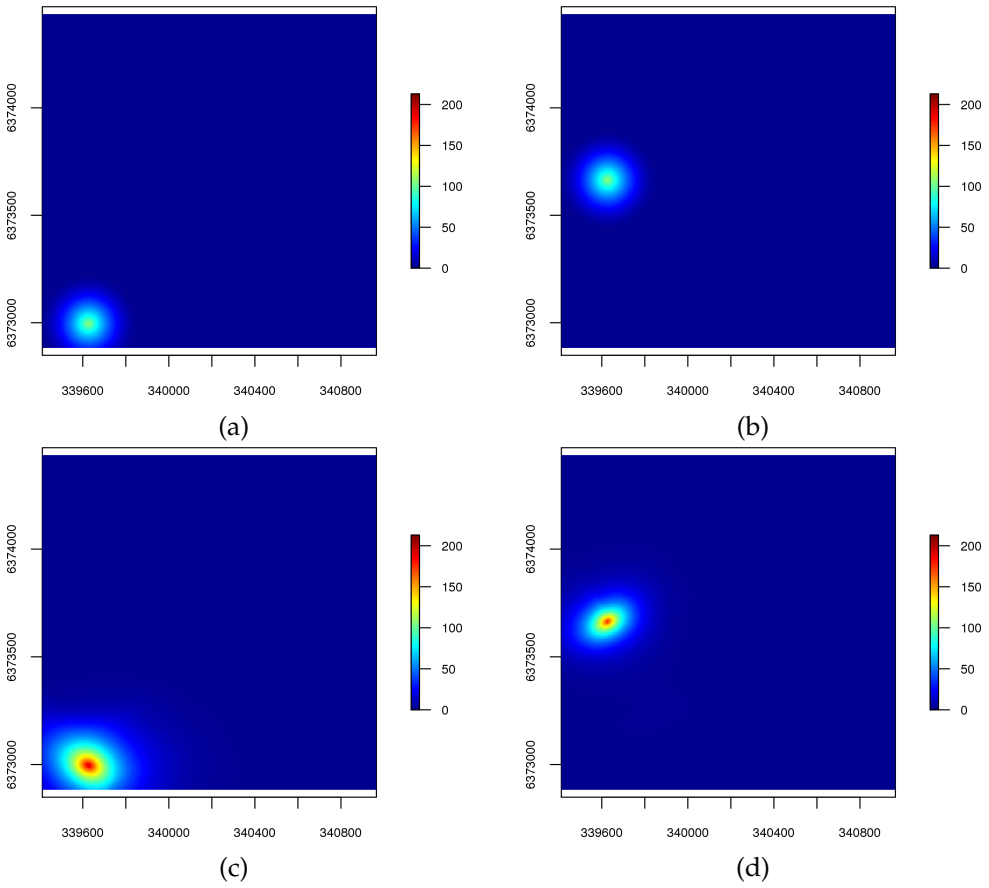


FIGURE 4.10 – Méthode de convolution : covariance d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d).

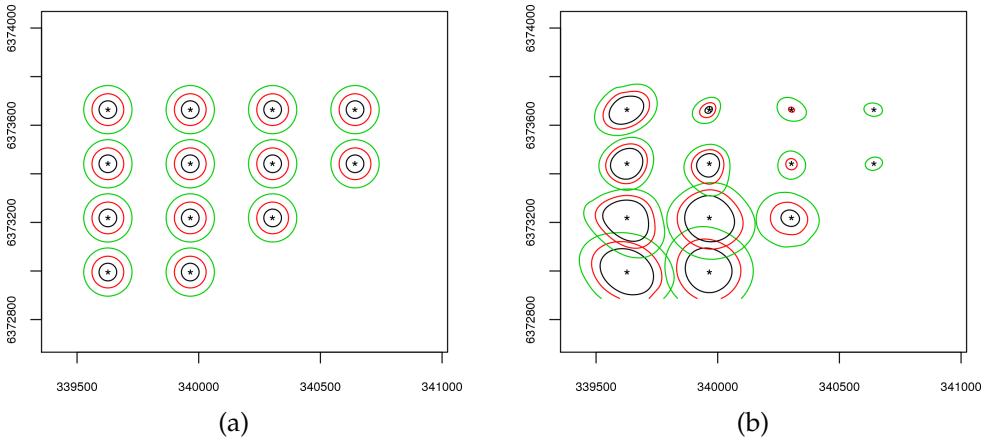


FIGURE 4.11 – Méthode de convolution : contours de la covariance en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 80 (noir), 60 (rouge) et 40 (vert).

Les résultats du krigeage pour les modèles stationnaire et non-stationnaire estimés sont présentés à la figure 4.12. Les cartes de krigeage et d'écart-type de krigeage sont différentes d'un modèle à l'autre. La carte de krigeage basée sur la méthode de convolution produit certains détails de la régionalisation absents sur la carte de krigeage en stationnaire. L'écart-type de krigeage donné par l'approche stationnaire se comporte de manière quasi-identique sur l'ensemble du domaine d'intérêt. Tandis qu'avec l'approche par convolution, l'écart-type de krigeage est en accord avec la structure sous-jacente des données. La figure 4.13 montrent quelques réalisations conditionnelles (dans le contexte gaussien) du modèle non-stationnaire estimé.

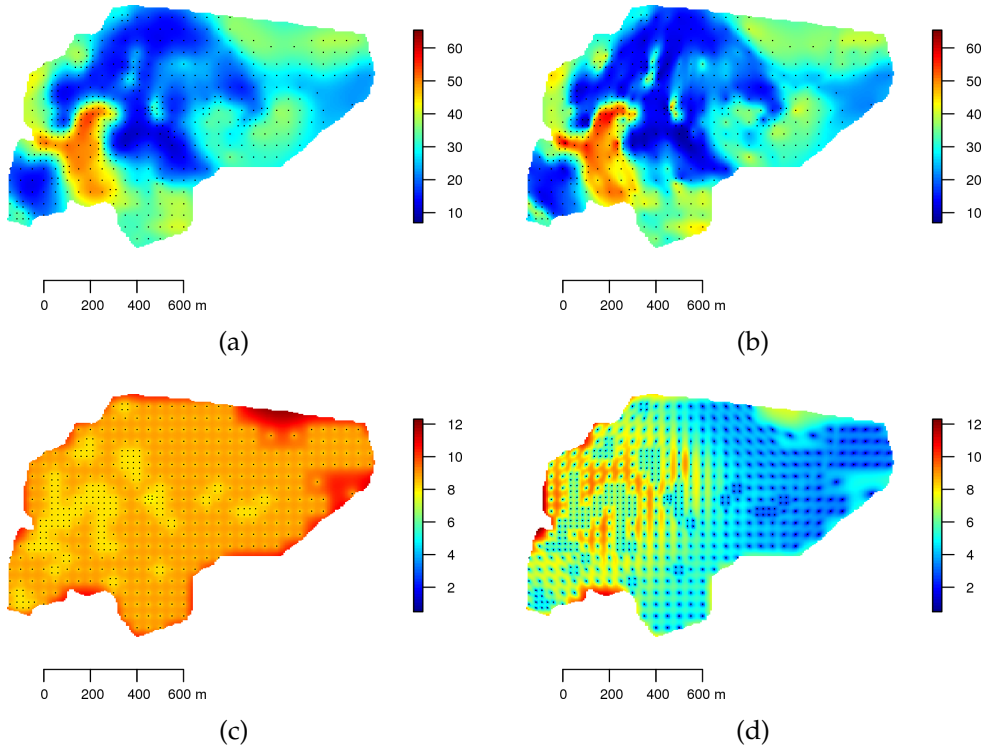
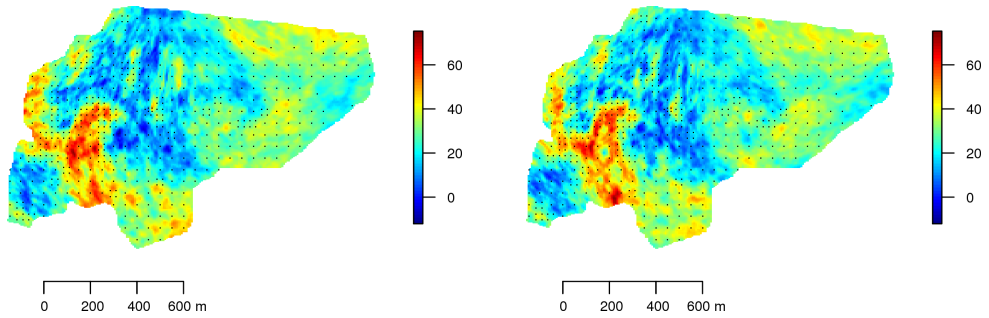
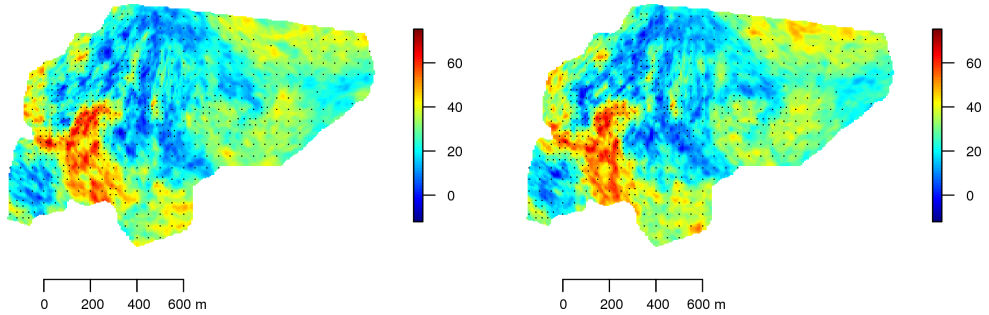


FIGURE 4.12 – Méthode de convolution : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; cartes d'écart-type de krigeage associées (c, d).



(a) Simulation #1 (b) Simulation #2



(c) Simulation #3 (d) Simulation #4

FIGURE 4.13 – Méthode de convolution : simulations conditionnelles.

Comparaison

Le tableau 4.1 résume les statistiques de performance prédictive [3] sur un échantillon de validation (1000 observations) pour les différentes méthodes (stationnaire, déformation d'espace et convolution). Comme on peut l'observer, les méthodes par déformation d'espace et convolution donnent de meilleurs résultats que celle stationnaire. Le coût de non-utilisation d'une approche non-stationnaire est non-négligeable : en terme de RMSE (*Root Mean Square Error*), la méthode de déformation d'espace (respectivement la méthode de convolution) diminue le RMSE de 21% (respectivement 22%) par rapport à la méthode stationnaire. On note aussi que l'approche par convolution est légèrement meilleure que l'approche par déformation d'espace.

3. Les scores de prédiction sont définis en annexe B.

	Stationnaire	Déformation	Convolution
MAE	2,86	2,17	2,12
RMSE	3,88	3,21	3,17
NMSE	0,19	0,26	0,30
LogS	6400	5517	5268
CRPS	4,97	3,61	3,28

TABLEAU 4.1 – Méthodes stationnaire/non-stationnaire : validation externe sur un échantillon de 1000 observations.

4.3 Application 2 : données de pluie

On s'intéresse à présent aux données pluviométriques suisses [Allard, 2006]. Ces dernières correspondent au cumul pluviométrique sur 24 heures (en mm) mesuré en 467 emplacements fixes à travers la Suisse le 8 mai 1986, soit quelques jours après l'accident de Tchernobyl. Les données sont séparées aléatoirement en deux groupes : un groupe de 400 données d'apprentissage et un groupe de 67 données de validation.

Méthode de déformation d'espace

Les points d'observation dans l'espace d'origine et leurs transformés dans l'espace déformé sont présentés aux figures 4.14a et 4.14b. Une configuration de 169 points supports a été utilisée pour construire l'espace déformé. On peut observer que l'espace d'origine est contracté dans la région nord-est et étiré dans la région sud-ouest. Ainsi, ces régions correspondent aux zones de forte et faible corrélation spatiale. Ce résultat est en cohérence avec la carte de variance produite par la méthode de convolution qui sera présentée ultérieurement. Les figures 4.14c et 4.14d représentent respectivement le variogramme stationnaire dans les espaces géographique et déformé. La modélisation non-stationnaire par déformation d'espace et celle stationnaire conduisent respectivement aux modèles ci-après :

$$\widehat{\gamma}_0(\|\mathbf{h}\|) = 14261 \times \mathrm{Sph}_{81}(\|\mathbf{h}\|), \tag{4.3}$$

$$\widehat{\gamma}_1(\mathbf{h}) = 14981 \times \mathrm{Sph}_{\substack{\mathbf{h}\,\max=213 \\ \mathbf{h}\,\min=60 \\ Az.=27^\circ}}(\mathbf{h}), \tag{4.4}$$

où $\widehat{\gamma}_0(.)$ est un variograme stationnaire isotrope sphérique, de variance 14261 mm^2, de portée 81 km. $\widehat{\gamma}_1(.)$ est un variogramme stationnaire anisotrope sphérique, de variance 14981 mm^2, de portée 213 km dans la direction de forte continuité (azimut 27°) et de portée 60 km dans la direction de faible continuité.

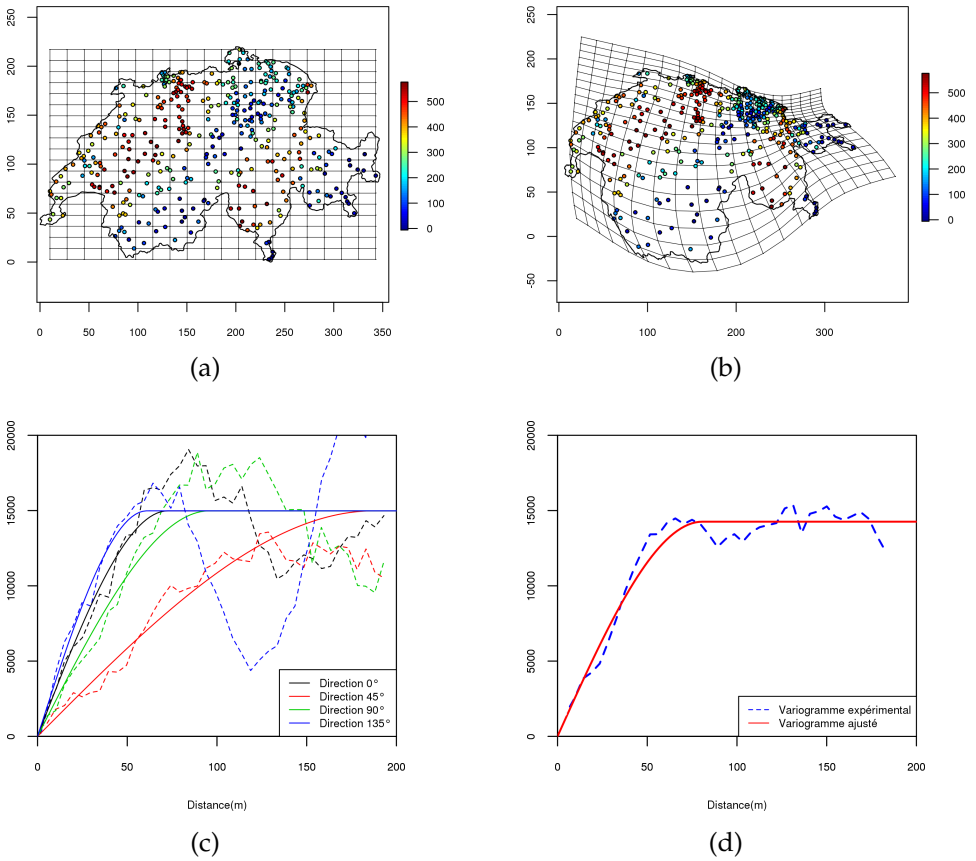


FIGURE 4.14 – Méthode de déformation d'espace : (a, c) données d'apprentissage dans l'espace géographique et modèle de variogramme stationnaire estimé ; (b, d) données d'apprentissage dans l'espace déformé et modèle de variogramme transformé estimé.

Les figures 4.15a et 4.15b montrent les fonctions de score en validation croisée $CV(\lambda)$ et $MSE(\lambda, \omega)$. Suivant la procédure de réglage des hyper-paramètres présentée en section 2.3.4, $\lambda = 163$ km et $\omega = 0,55$ sont les valeurs optimales. La validation externe conduit à ces mêmes valeurs d'hyper-paramètres (figure 4.15d). La figure 4.15c représente le diagramme de Shepard de la solution NMDS (espace déformé). La valeur du *stress* associée à cette solution vaut 7%.

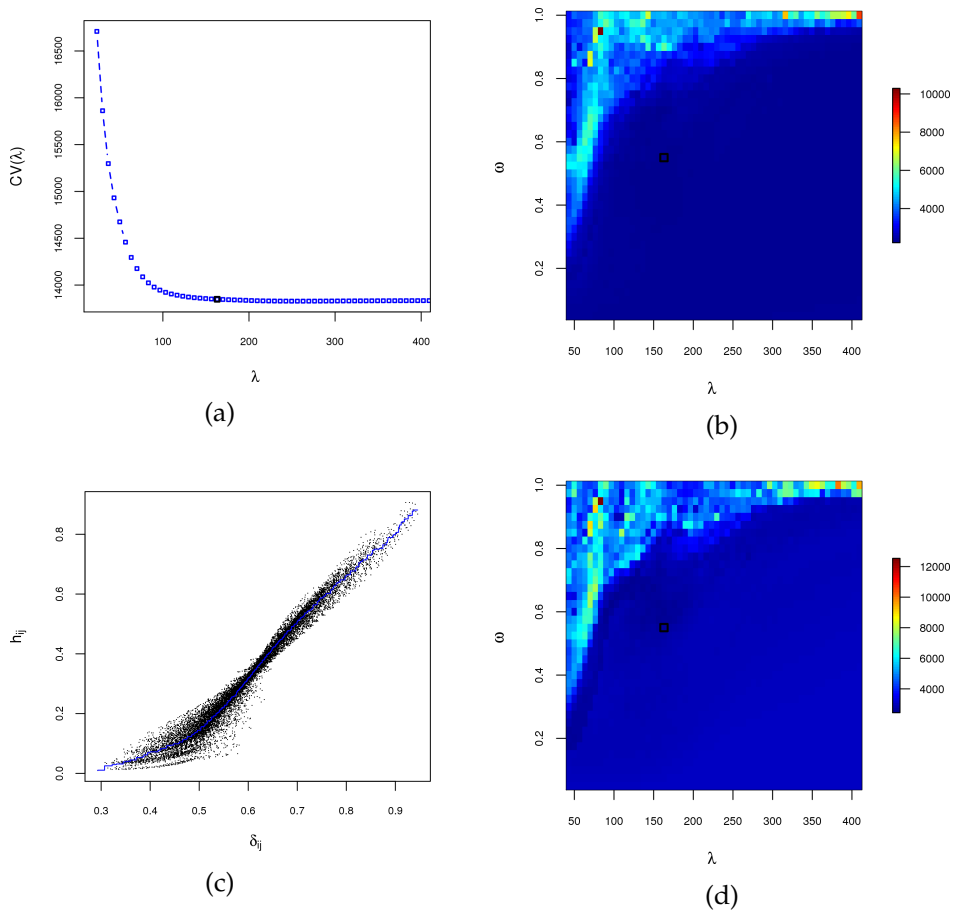


FIGURE 4.15 – Méthode de déformation d'espace : (a) fonction score $CV(\lambda)$ de validation croisée ; (c) diagramme de Shepard de la solution NMDS ; fonction score $MSE(\lambda, \omega)$ en validation croisée (b) et externe (d).

La structure de dépendance spatiale sous les hypothèses de stationnarité et de non-stationnarité est présentée en quelques points aux figures 4.16 et 4.17. Le changement de forme de la structure de dépendance spatiale non-stationnaire s'observe bien d'un site à un autre.

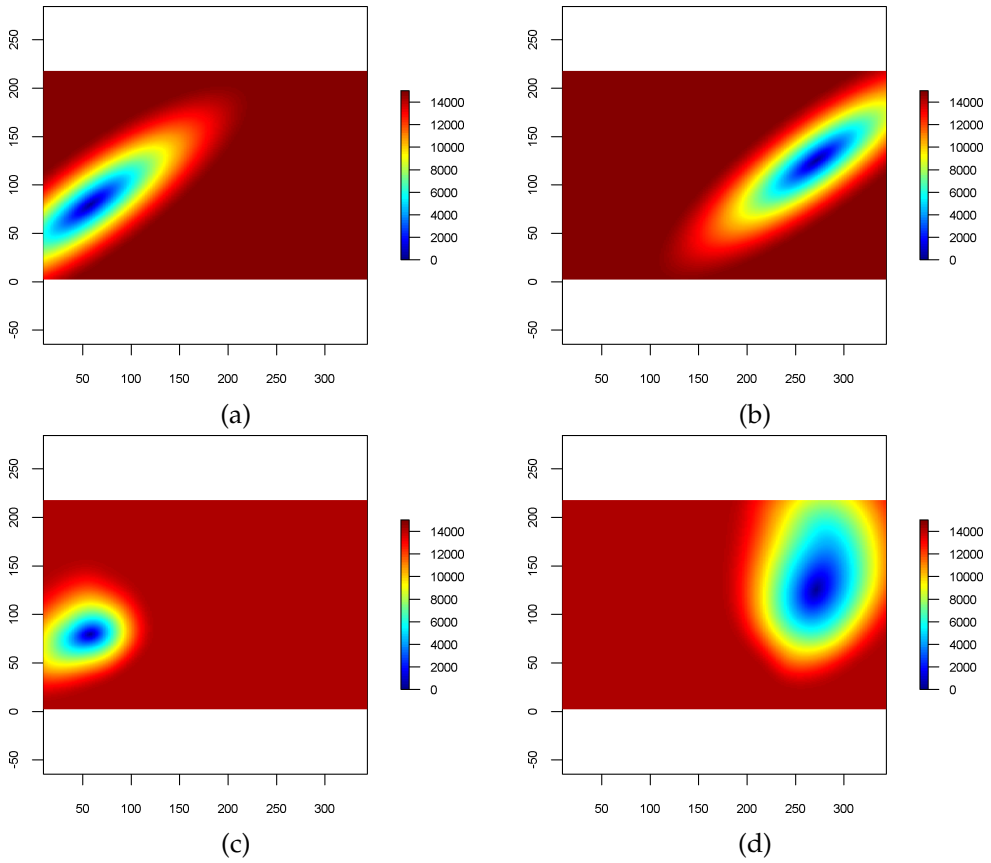


FIGURE 4.16 – Méthode de déformation d'espace : variogramme d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d).

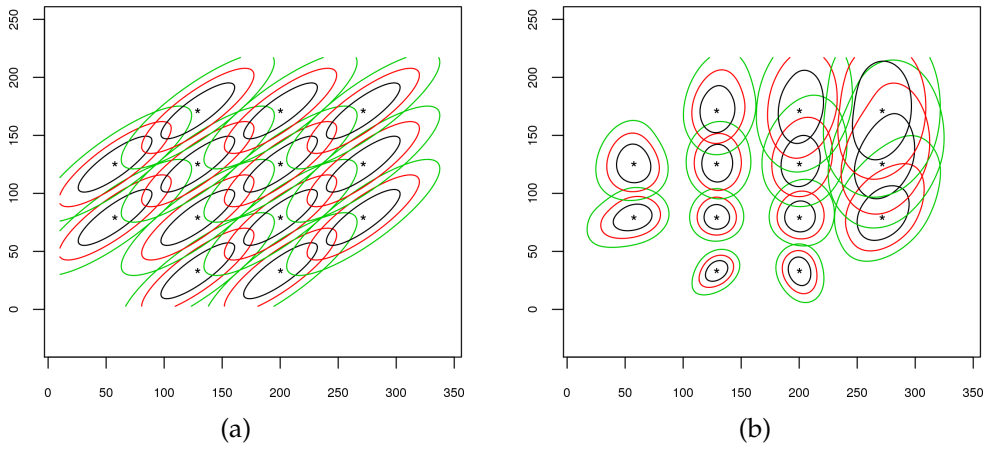


FIGURE 4.17 – Méthode de déformation d'espace : contours du variogramme en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 4000 (noir), 6000 (rouge) et 8000 (vert).

La figure 4.18 montre les résultats de la prédiction spatiale basée sur les modèles stationnaire et non-stationnaire. Comme on peut le constater, les cartes de krigeage et d'écart-type de krigeage associées diffèrent sensiblement selon l'approche. Comparativement au modèle stationnaire basé sur des paramètres structuraux globaux identiques sur tout le domaine d'étude, le modèle non-stationnaire s'adapte à la structure locale variable des données. Quelques simulations conditionnelles (dans le contexte gaussien) du modèle non-stationnaire estimé sont présentées à la figure 4.19.

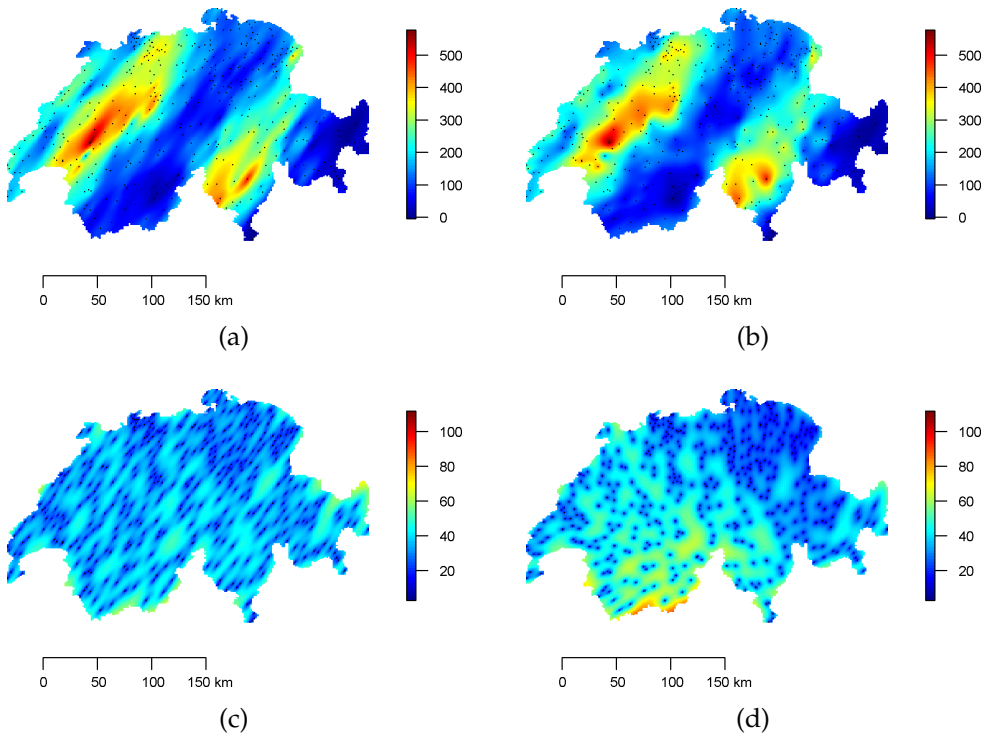
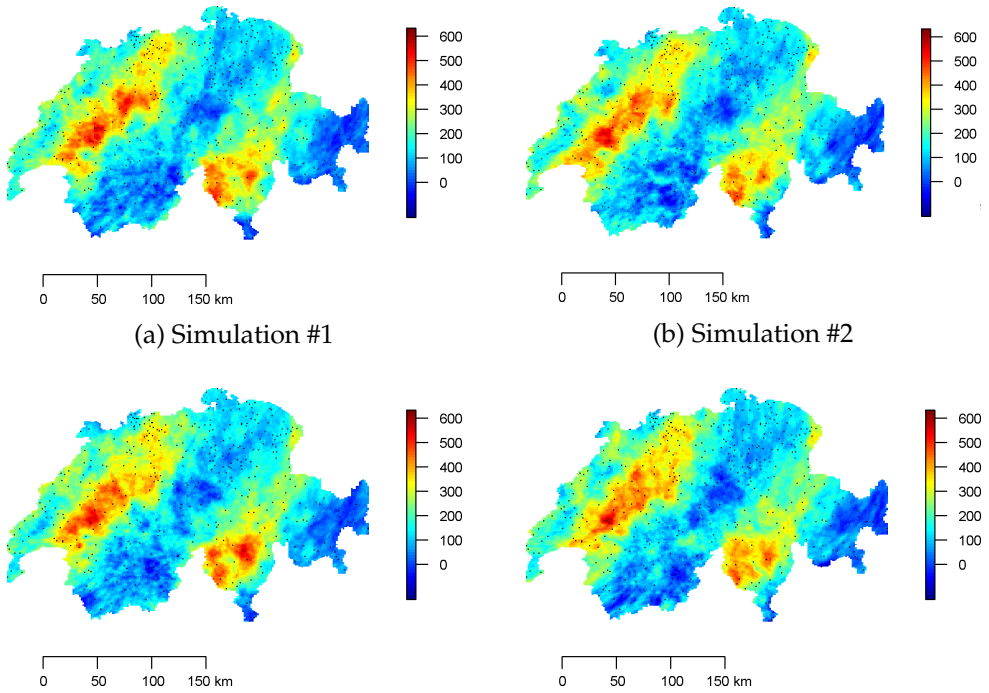


FIGURE 4.18 – Méthode de déformation d'espace : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; (c, d) cartes d'écart-type de krigeage associées.



(a) Simulation #1 (b) Simulation #2

(c) Simulation #3 (d) Simulation #4

FIGURE 4.19 – Méthode de déformation d'espace : simulations conditionnelles.

Méthode de convolution

La figure 4.20a donne une représentation des données d'apprentissage et suggère la présence d'une direction d'anisotropie globale dans les données. Ce constat est confirmé par le calcul de variogrammes expérimentaux directionnels dans le contexte stationnaire (voir figure 4.14c). La direction de forte continuité pour le modèle stationnaire global est donnée par la direction sud/ouest-nord/est. Les figures 4.20b, 4.20c et 4.20d montrent respectivement les paramètres $\widehat{m}(.)$, $\widehat{\sigma}^2(.)$ et $\widehat{\mathbf{\Sigma}}(.)$ estimés aux points supports. Ces estimations brutes sont basées sur une structure de covariance non-stationnaire exponentielle (exemple 3.7). La figure 4.20d révèle une anisotropie géométrique locale variable incluant la direction principale de continuité trouvée par l'approche stationnaire. Notons que l'anisotropie géométrique globale ne saurait décrire des caractéristiques directionnelles locales.

La fonction de score $MSE(\epsilon)$ en validation croisée est donnée par la figure 4.21a. La procédure de sélection présentée à la section 3.4.3 conduit à prendre comme largeur de fenêtre $\epsilon = 46$ km. La fonction de score $MSE(\epsilon)$ en validation externe confirme cette valeur optimale (figure 4.21b). La figure 4.22 fournit les cartes de paramètres lissés sur le domaine d'étude : moyenne, variance, ratio d'anisotropie et azimut. La valeur optimale du paramètre lissage associée au noyau gaussien correspond à $\delta = 11$ km.

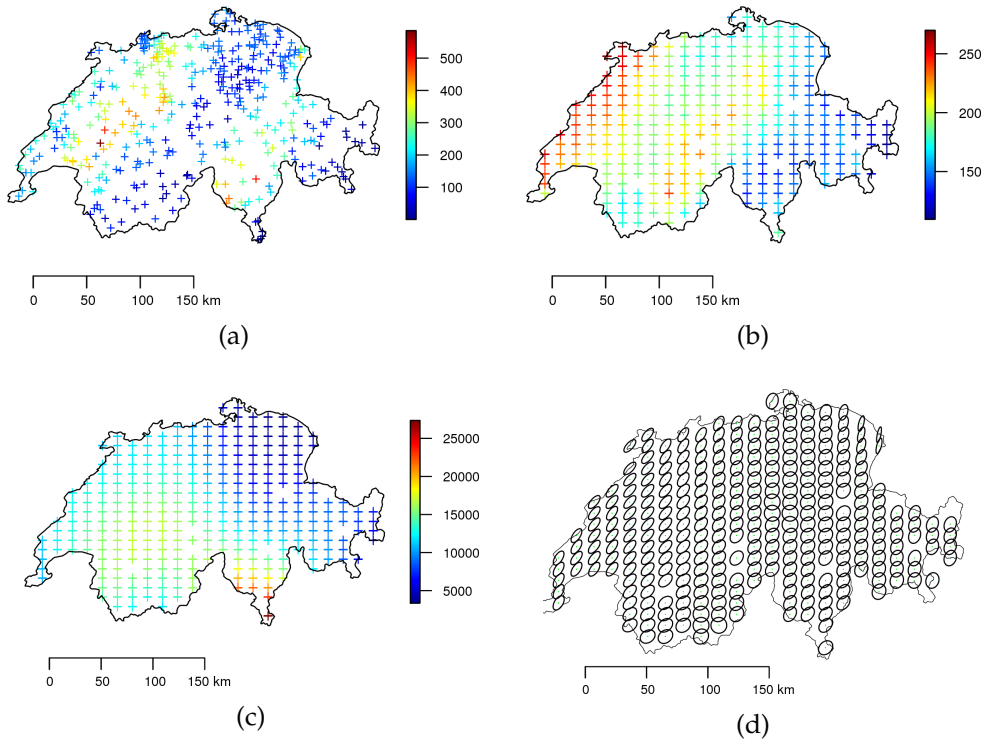


FIGURE 4.20 – Méthode de convolution : (a) données d'apprentissage ; (b) fonction moyenne estimée $\widehat{\mathrm{m}}(.)$ aux points supports ; (c) fonction variance estimée $\widehat{\sigma}^2(.)$ aux points supports ; (d) fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports.

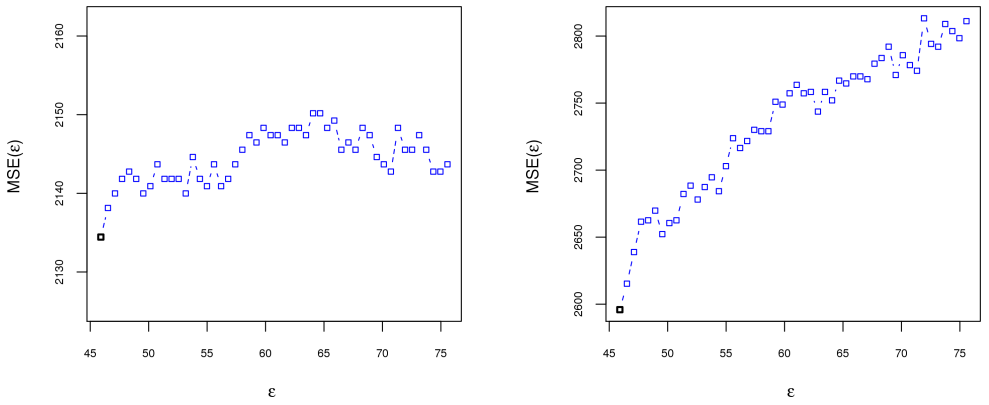


FIGURE 4.21 – Méthode de convolution : fonction de score $MSE(\epsilon)$ en validation croisée (a) et externe (b).

Comme nous l'avons souligné dans la première application en section 4.2, l'approche par convolution est en accord avec celle par déformation d'espace. La carte de variance représentée à la figure 4.20b est en adéquation avec l'action la déformation présentée à la figure 4.14b. En effet, la région nord-est contractée correspond à une région de faible variance, tandis que la région sud-ouest étirée correspond à une région de forte variance.

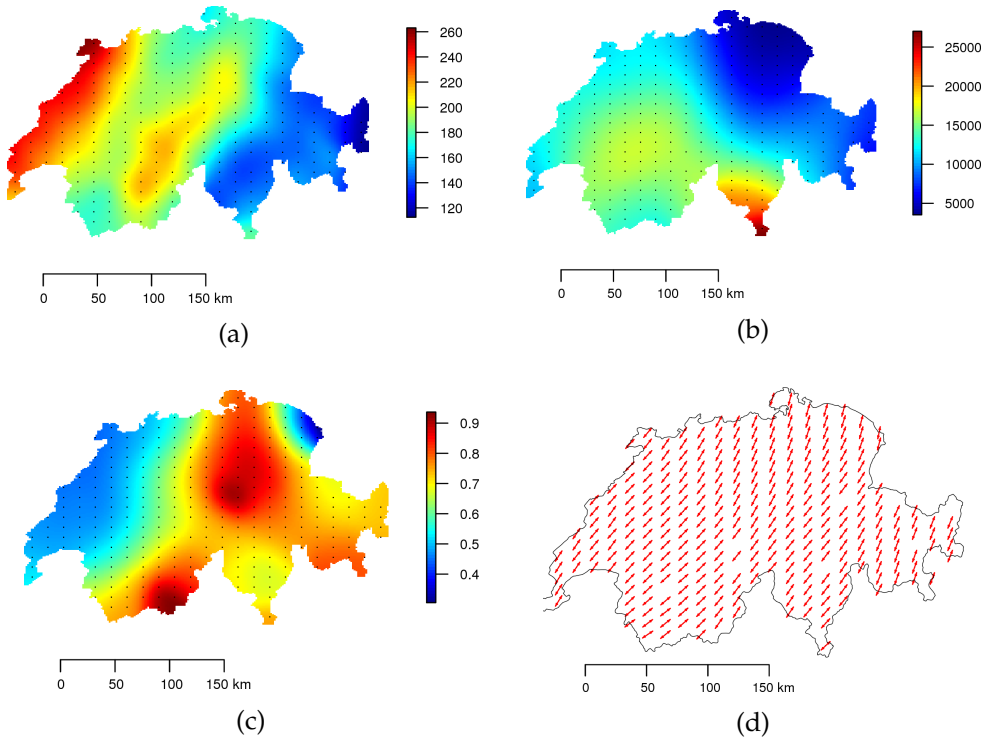


FIGURE 4.22 – Méthode de convolution : paramètres lissés sur le domaine des observations : (a) moyenne, (b) variance, (c) ratio d'anisotropie, (d) azimut.

Selon les hypothèses de stationnarité ou non-stationnarité, une visualisation de la structure de corrélation spatiale en quelques points est donnée aux figures 4.23 et 4.24. La variation de la structure de corrélation spatiale non-stationnaire estimée d'un emplacement à un autre est visible.

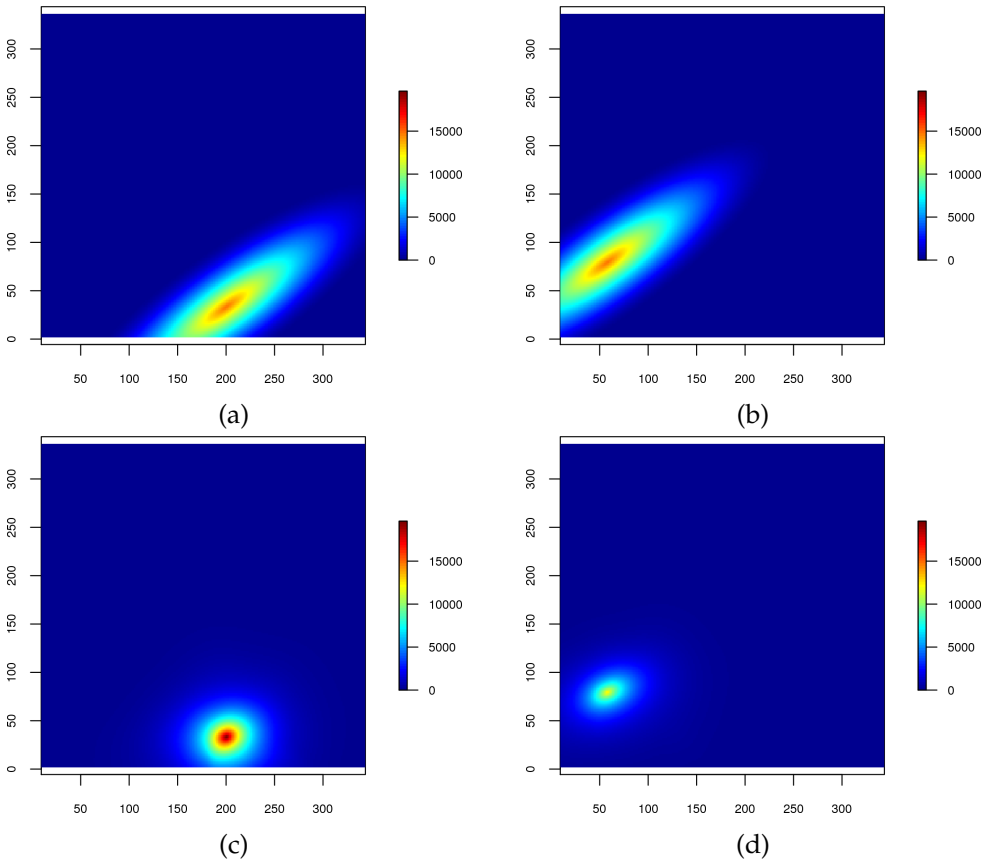


FIGURE 4.23 – Méthode de convolution : covariance d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d).

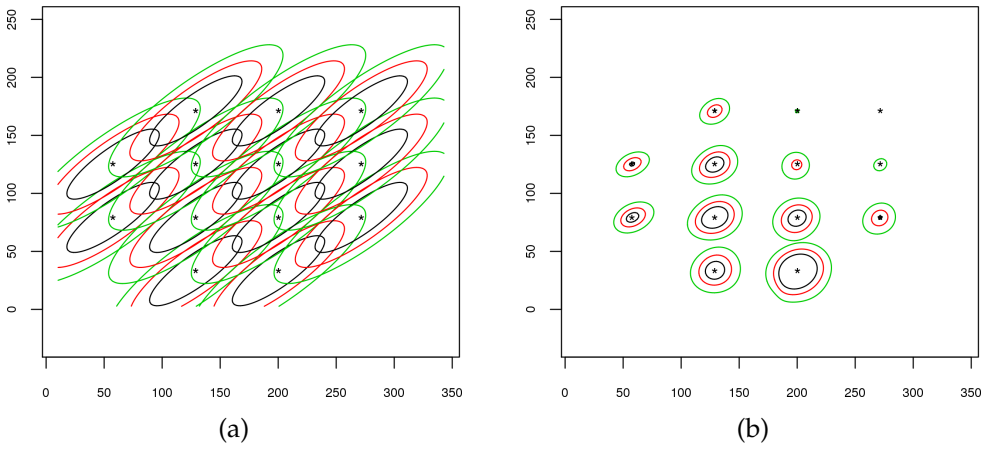


FIGURE 4.24 – Méthode de convolution : contours de la covariance en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 10000 (noir), 8000 (rouge) et 6000 (vert).

Les résultats du krigeage pour les modèles stationnaire et non-stationnaire estimés sont donnés par la figure 4.25. La différence entre la structure de corrélation spatiale en stationnaire et non-stationnaire s'observe clairement sur les cartes de krigeage et d'écart-type de krigeage associées à chacune des approches. On peut voir que l'approche par convolution s'adapte mieux à la structure locale des données que l'approche stationnaire. Certains détails de la régionalisation sont gommés avec l'approche stationnaire. La figure 4.26 montre quelques simulations conditionnelles (dans le cadre gaussien) du modèle non-stationnaire estimé.

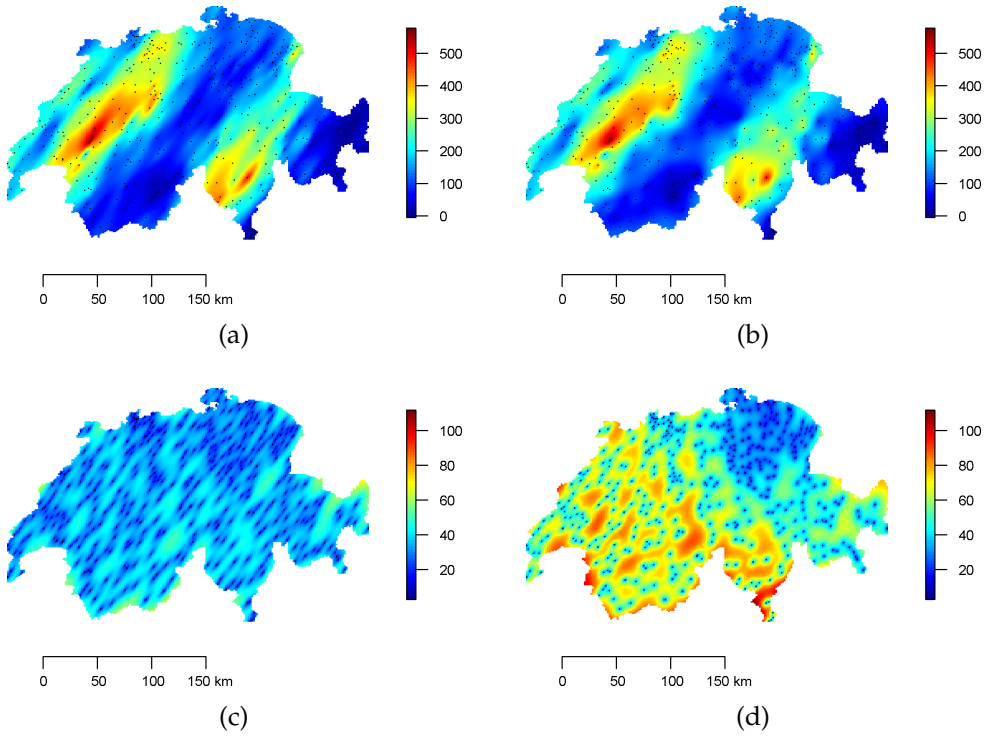
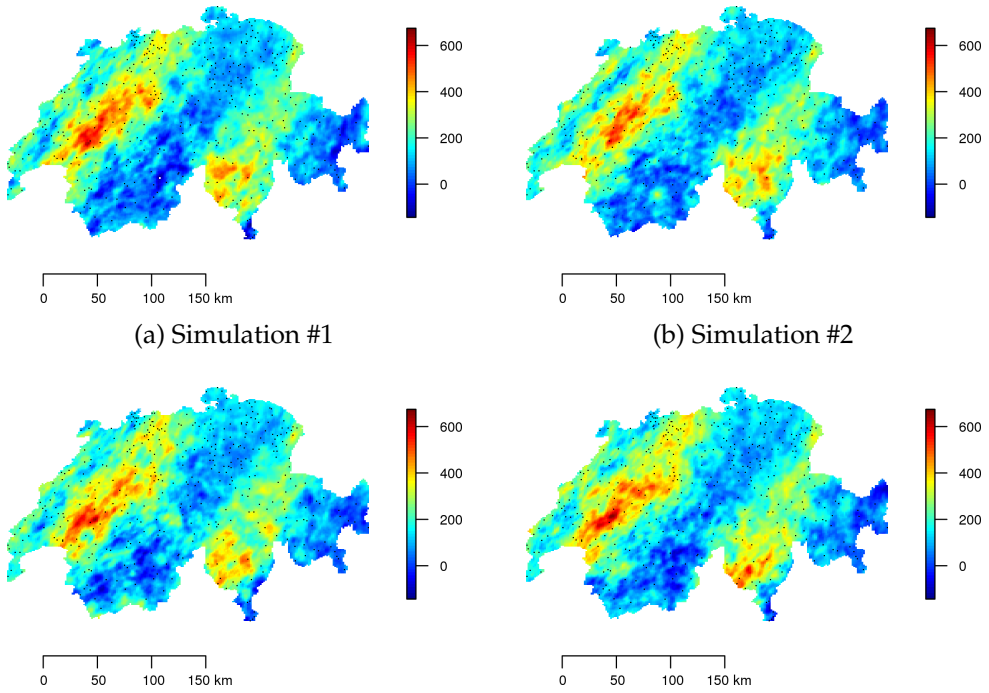


FIGURE 4.25 – Méthode de convolution : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; cartes d'écart-type de krigeage associées (c, d).



(a) Simulation #1 (b) Simulation #2

(c) Simulation #3 (d) Simulation #4

FIGURE 4.26 – Méthode de convolution : simulations conditionnelles.

Comparaison

Les performances en validation externe pour les méthodes stationnaire, déformation d'espace et convolution sont résumées dans le tableau 4.2 [4]. L'analyse du tableau indique la supériorité des méthodes non-stationnaires proposées sur la méthode stationnaire. On note par exemple en terme du critère de RMSE, une amélioration globale de près de 12%. Toutefois, notons que la taille de l'échantillon de validation est assez faible (seulement 67 observations). Au vu des résultats présentés dans les tableaux 4.2 et de la petite taille de l'échantillon de validation, on ne saurait dire si l'approche par convolution ou celle par déformation d'espace est la meilleure.

	Stationnaire	Déformation	Convolution
MAE	42,21	37,36	37,10
RMSE	56,80	50,51	50,90
NMSE	2,24	1,53	0,70
LogS	753	715	710
CRPS	53,31	46,37	50,60

TABLEAU 4.2 – Méthodes stationnaire/non-stationnaire : validation externe sur un échantillon de 67 observations.

4. Les scores de prédiction calculés sont définis en annexe B.

4.4 Application 3 : données de topographie

Dans cette section, la variable régionalisée étudiée est relative aux données topographiques (en m) de la cheminée de brèche (*breccia pipe*) nommé *braden* de la mine d'El Teniente au Chili [Séguret and Celhay, 2013]. Cette dernière est l'un des plus grands gisements de cuivre porphyrique. Sa cheminée de brèche est un cône minéralisé presque vertical, situé au centre de la mine et entouré de minéralisations à un stade précoce. Comme le bord de la cheminée constitue la limite du dépôt et de l'exploitation minière, l'estimer avec précision est important. On dispose d'un ensemble de données de 816 observations séparées aléatoirement en un échantillon d'apprentissage de 616 observations et un échantillon de validation de 200 observations.

Méthode de déformation d'espace

Les figures 4.27a et 4.27b montrent respectivement la configuration des données d'apprentissage dans l'espace géographique et celle de leur déformation. L'espace déformé a été construit avec un ensemble de 179 points supports. On peut constater que la déformation a contracté l'espace géographique dans la région ouest et l'a étiré dans la région est. Ces régions correspondent alors respectivement aux zones de forte et faible corrélation spatiale. Par ailleurs, comme nous le verrons par la suite, ce résultat est en cohérence avec la carte de variance produite par la méthode de convolution. Les variogrammes stationnaires isotropes dans les espaces géographique et déformé [5] sont présentés aux figures 4.27c et 4.27d. La modélisation non-stationnaire par déformation d'espace et celle stationnaire conduisent respectivement aux modèles suivants :

$$\widehat{\gamma}_0(\|\mathbf{h}\|) = 9947 \times \mathrm{Exp}_{96}(\|\mathbf{h}\|) + 32757 \times \mathrm{Sph}_{323}(\|\mathbf{h}\|), \quad (4.5)$$

$$\widehat{\gamma}_1(\|\mathbf{h}\|) = 3050 \times \mathrm{Nug}(\|\mathbf{h}\|) + 40999 \times \mathrm{Sph}_{181}(\|\mathbf{h}\|), \quad (4.6)$$

où $\widehat{\gamma}_0(.)$ est un variograme stationnaire isotrope gigogne (exponentiel et sphérique) de variance totale 42704 m^2. $\widehat{\gamma}_1(.)$ est un variogramme stationnaire isotrope gigogne (effet pépite, sphérique) de variance totale 44049 m^2 et de portée 181 m .

5. L'approche stationnaire n'a pas détecté d'anisotropie géométrique globale.

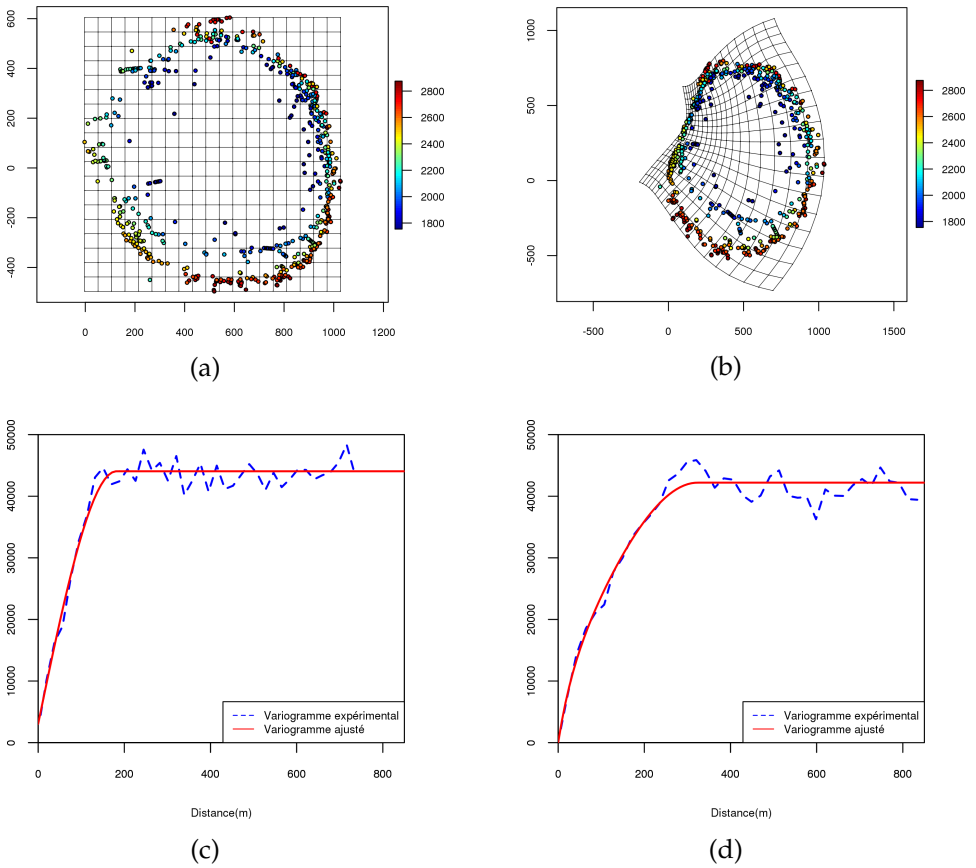


FIGURE 4.27 – Méthode de déformation d'espace : (a, c) données d'apprentissage dans l'espace géographique et modèle de variogramme stationnaire estimé ; (b, d) données d'apprentissage dans l'espace déformé et modèle de variogramme transformé estimé.

La sélection des hyper-paramètres décrite en section 2.3.4 est présentée aux figures 4.28a et 4.28b. La valeur optimale des hyper-paramètres de la fonction de score pour la validation croisée $MSE(\lambda, \omega)$ correspond à $\lambda = 1446$ m et $\omega = 0,90$. Ces valeurs optimales concordent avec celles données par la validation externe (figure 4.28d). Le diagramme de Shepard relatif à l'espace déformé (solution NMDS) est donnée par la figure 4.28c. La solution NMDS correspond à un *stress* valant 8%.

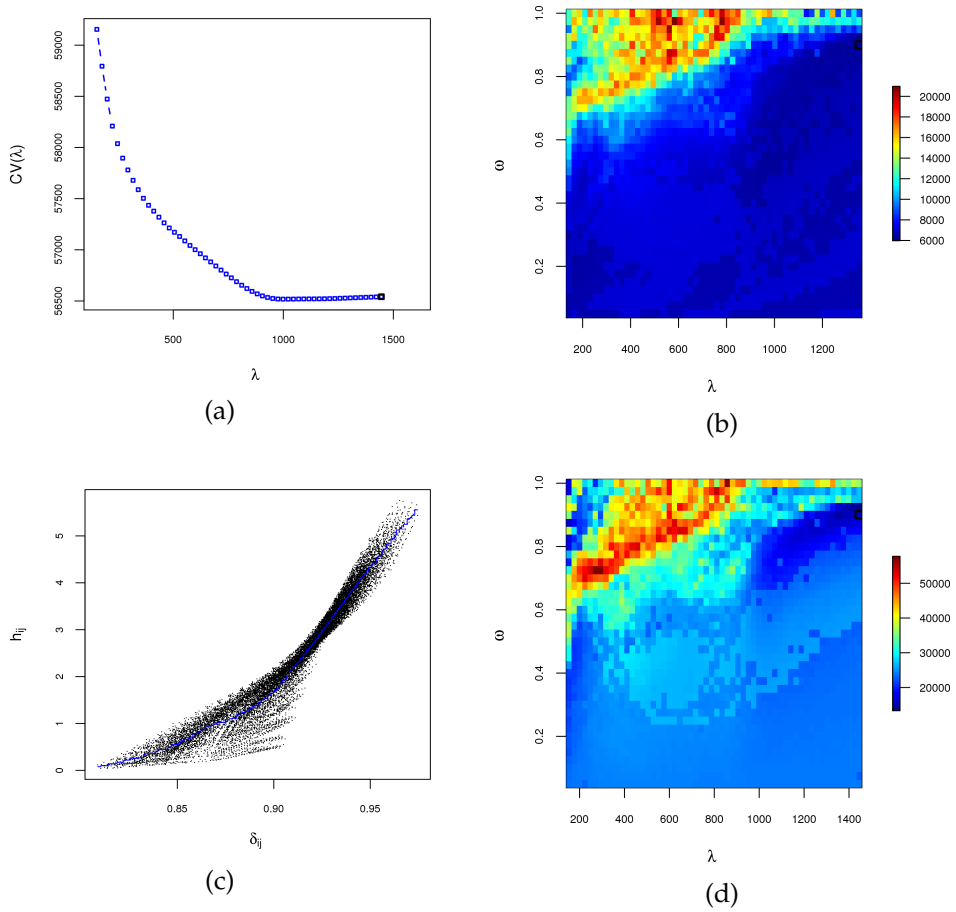


FIGURE 4.28 – Méthode de déformation d'espace : (a) fonction score $CV(\lambda)$ de validation croisée ; (c) diagramme de Shepard de la solution NMDS ; fonction score $MSE(\lambda, \omega)$ en validation croisée (b) et externe (d).

Une représentation graphique de la structure de dépendance spatiale en contexte stationnaire et non-stationnaire est donnée aux figures 4.29 et 4.30. L'effet de la déformation s'observe bien sur la structure de dépendance spatiale non-stationnaire qui varie d'un point à un autre.

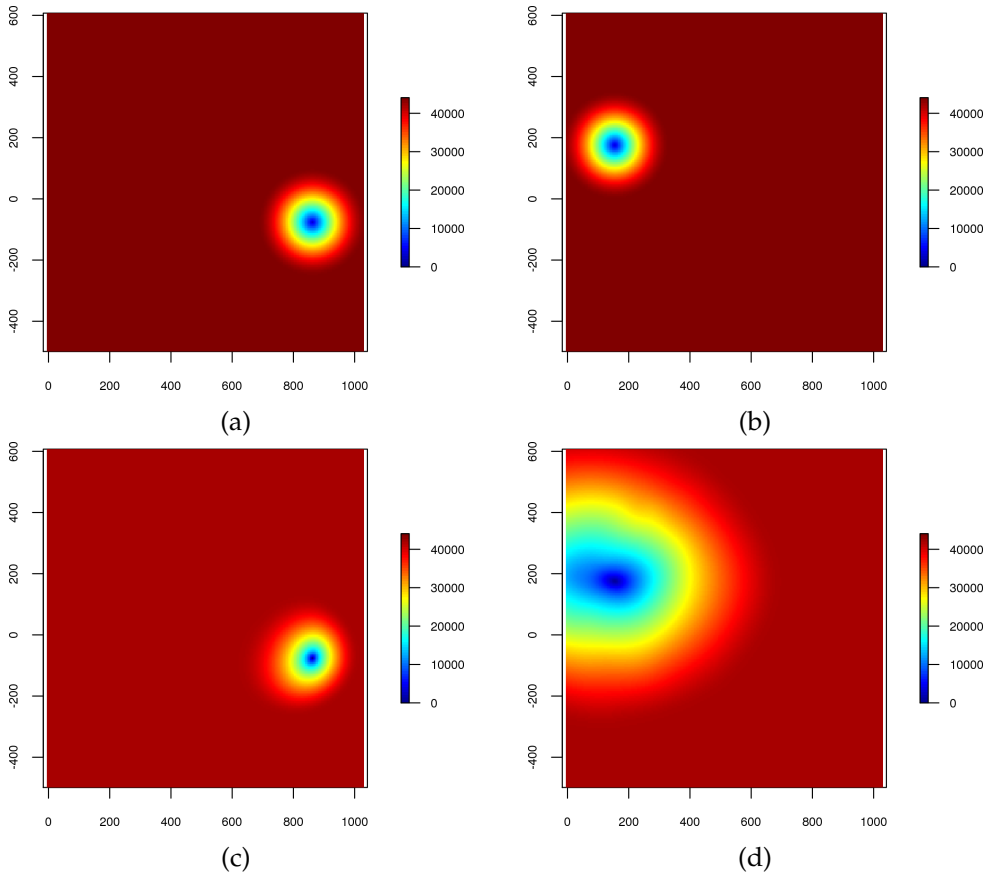


FIGURE 4.29 – Méthode de déformation d'espace : variogramme d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d).

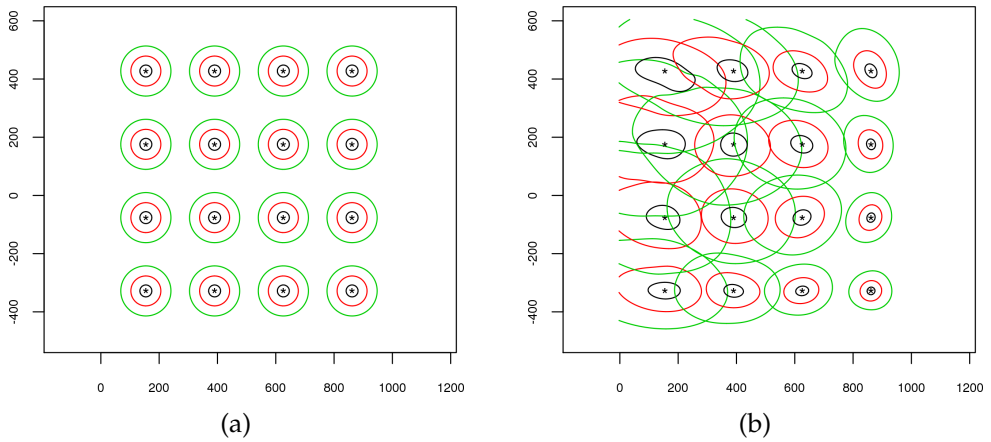


FIGURE 4.30 – Méthode de déformation d'espace : contours du variogramme en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 10000 (noir), 20000 (rouge) et 30000 (vert).

Les résultats du krigeage pour les modèles stationnaire et non-stationnaire sont présentés à la figure 4.31. L'allure générale des cartes de valeurs prédites et celles d'écart-type de krigeage associées à chaque modèle diffère nettement. Cette différence est plus marquée au centre du domaine d'intérêt où il y a très peu d'observations. Quelques simulations conditionnelles (dans le cadre gaussien) du modèle non-stationnaire estimé sont présentées à la figure 4.32.

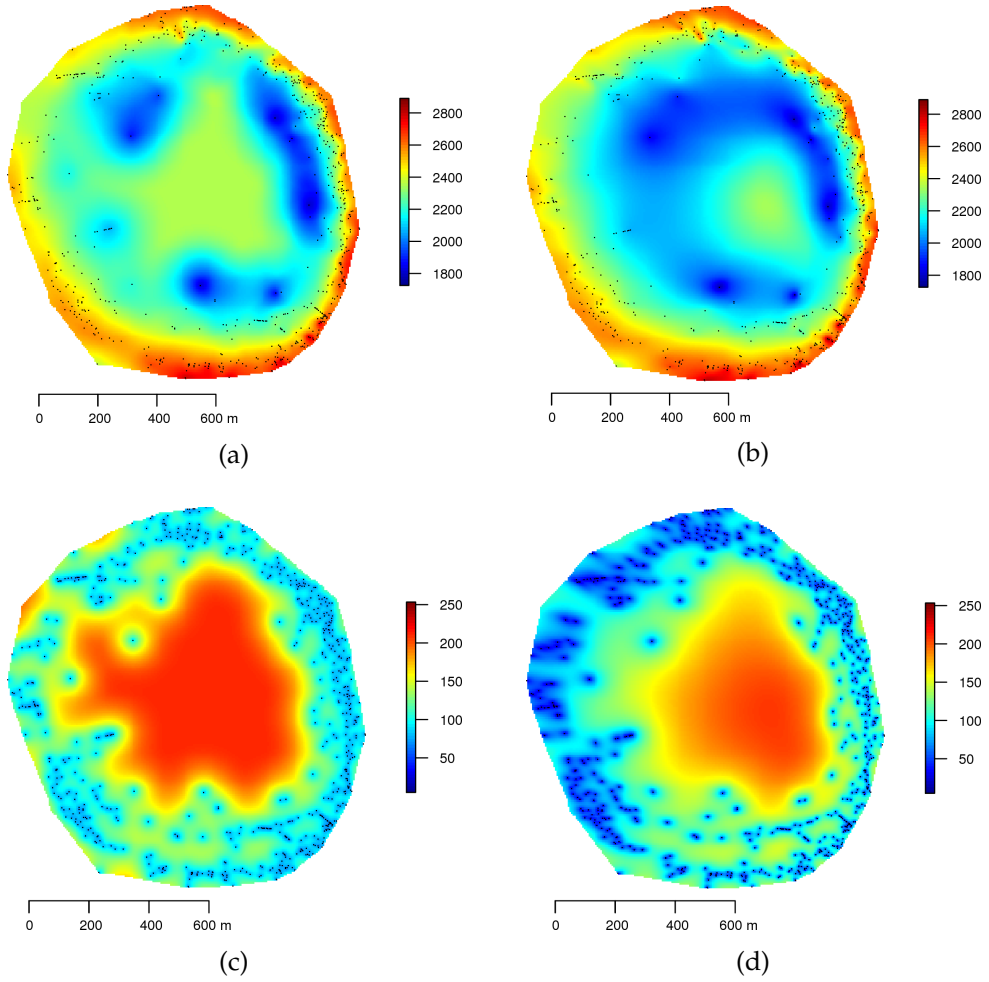
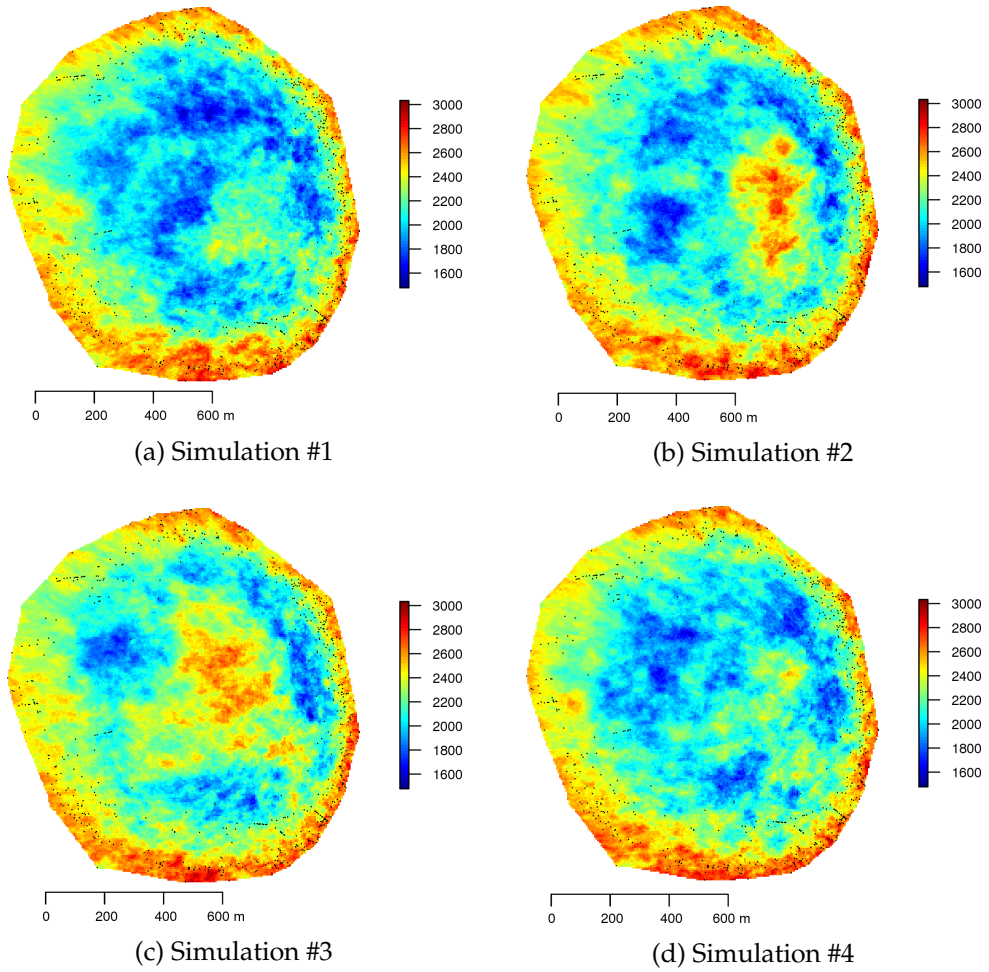


FIGURE 4.31 – Méthode de déformation d'espace : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; (c, d) cartes d'écart-type de krigeage associées.



(a) Simulation #1

(b) Simulation #2

(c) Simulation #3

(d) Simulation #4

FIGURE 4.32 – Méthode de déformation d'espace : simulations conditionnelles.

Méthode de convolution

Les figures 4.33b, 4.33c et 4.33d donnent les estimations brutes des paramètres $\widehat{m}(.)$, $\widehat{\sigma}^2(.)$ et $\widehat{\mathbf{\Sigma}}(.)$ aux points supports. Ces estimations brutes sont basées sur une structure de covariance non-stationnaire exponentielle (exemple 3.7). Elles mettent en évidence la non-stationnarité dans les données. On peut observer par exemple sur la figure 4.33d l'anisotropie géométrique locale variable. Notons que l'approche stationnaire n'a pas détecté d'anisotropie géométrique globale.

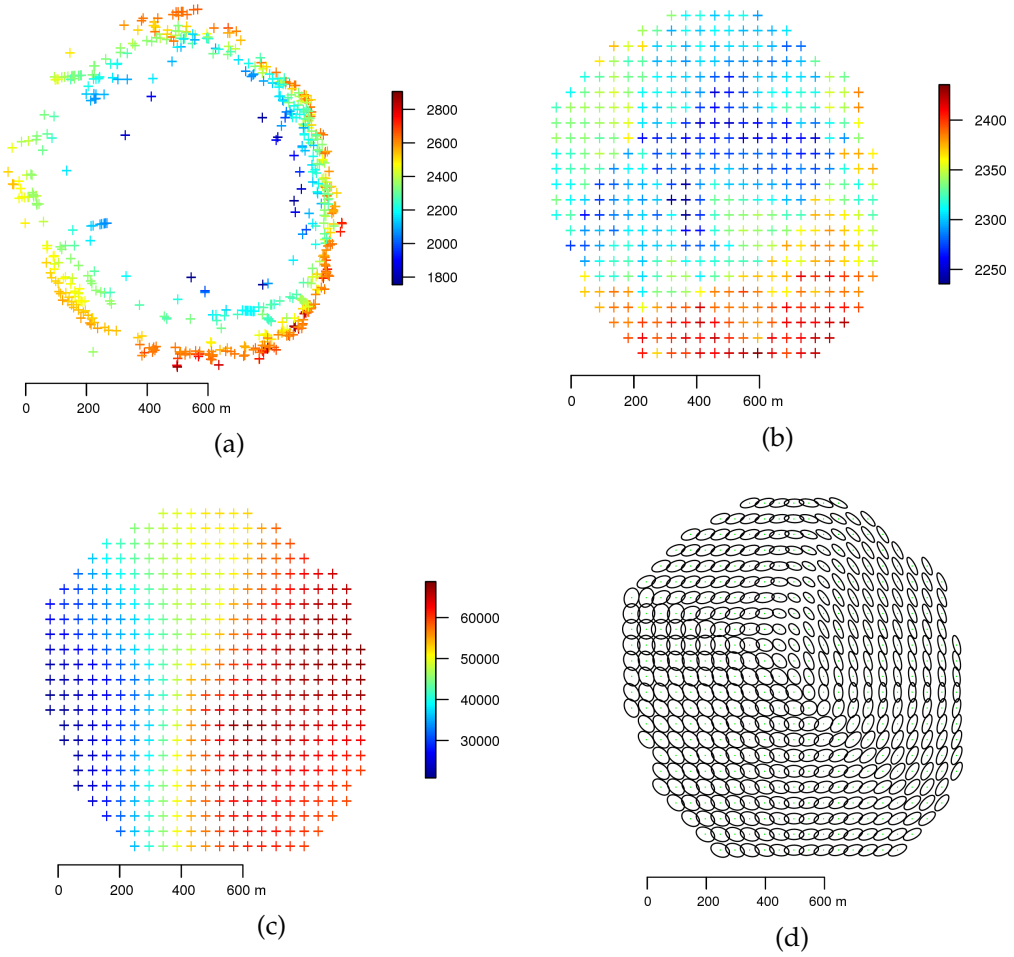
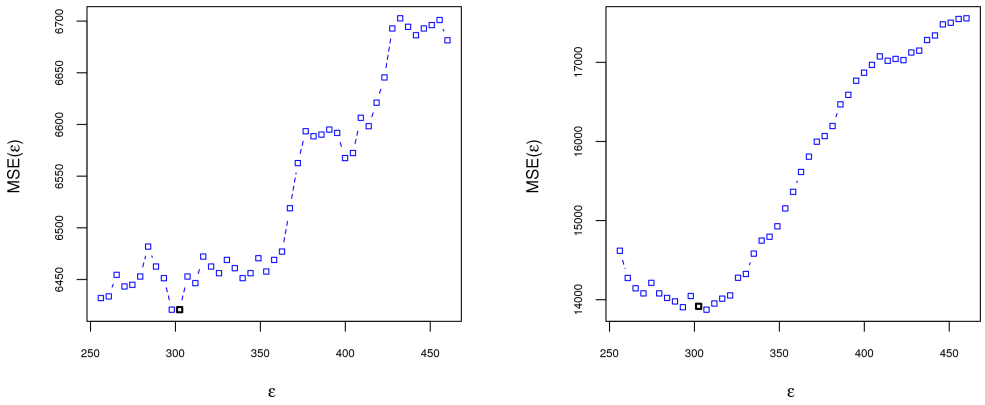


FIGURE 4.33 – Méthode de convolution : (a) données d'apprentissage ; (b) fonction moyenne estimée $\widehat{\mathrm{m}}(.)$ aux points supports ; (c) fonction variance estimée $\widehat{\sigma}^2(.)$ aux points supports ; (d) fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports.

La sélection de la largeur de fenêtre par validation croisée telle que présentée en section 3.4.3 permet d'obtenir la valeur optimale $\epsilon = 302$ m comme illustrée à la figure 4.34a. Cette valeur est confirmée par la validation externe en figure 4.34b. Le résultat du lissage des estimations brutes des paramètres (moyenne, variance, ratio d'anisotropie et azimut) est présenté à la figure 4.35. Suivant la procédure de sélection décrite en section 3.4.3, la valeur optimale du paramètre de lissage associée au noyau gaussien vaut $\delta = 26$ m.

Comme dans les applications précédentes, l'approche par convolution est en accord avec celle par déformation d'espace. La carte de variance représentée à la figure 4.35b est en adéquation avec l'action de la déformation présentée à la figure 4.27b. En effet, la région ouest contractée correspond à une région de faible variance, tandis que la région est étirée correspond à une région de forte variance.



(a) (b)

FIGURE 4.34 – Méthode de convolution : fonction de score $MSE(\epsilon)$ en validation croisée (a) et externe (b).

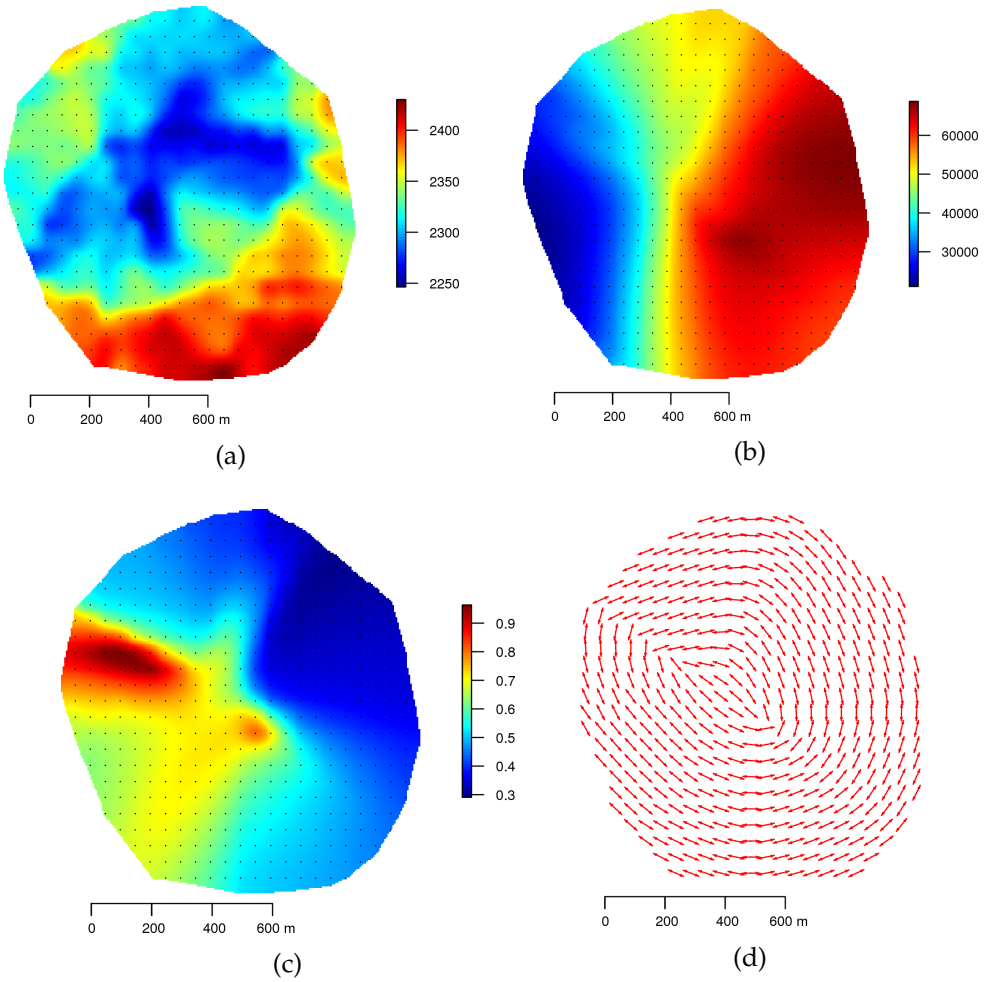


FIGURE 4.35 – Méthode de convolution : paramètres lissés sur le domaine des observations : (a) moyenne, (b) variance, (c) ratio d'anisotropie, (d) azimut.

La structure de corrélation spatiale estimée sous les hypothèses stationnaire et non-stationnaire est visualisée en quelques points sur les figures 4.36 et 4.37. On peut ainsi voir comment la structure de corrélation non-stationnaire change d'un point à un autre, comparativement à celle stationnaire.

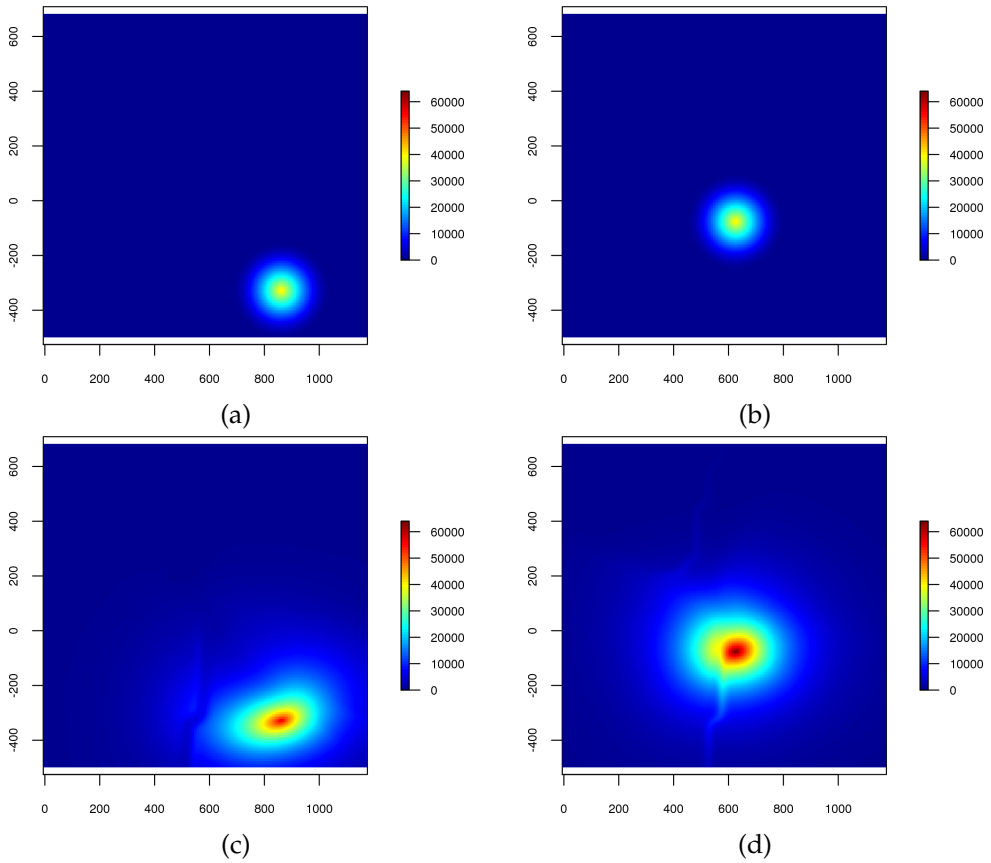


FIGURE 4.36 – Méthode de convolution : covariance d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d).

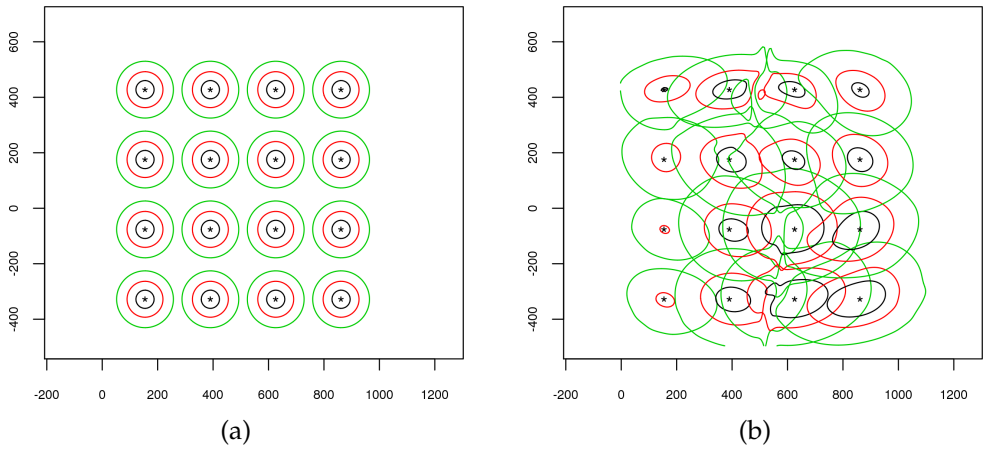


FIGURE 4.37 – Méthode de convolution : contours de la covariance en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 30000 (noir), 20000 (rouge) et 10000 (vert).

Le krigeage basé sur les modèles stationnaire et non-stationnaire conduit aux résultats présentés à la figure 4.31. Les cartes de krigeage et d'écart-type de krigeage pour chaque modélisation sont assez distinctes. Cette dissemblance est plus prononcée au centre du domaine d'intérêt où il y a très peu d'observations. À la figure 4.32 sont présentées quelques simulations conditionnelles (dans le cadre gaussien) du modèle non-stationnaire estimé.

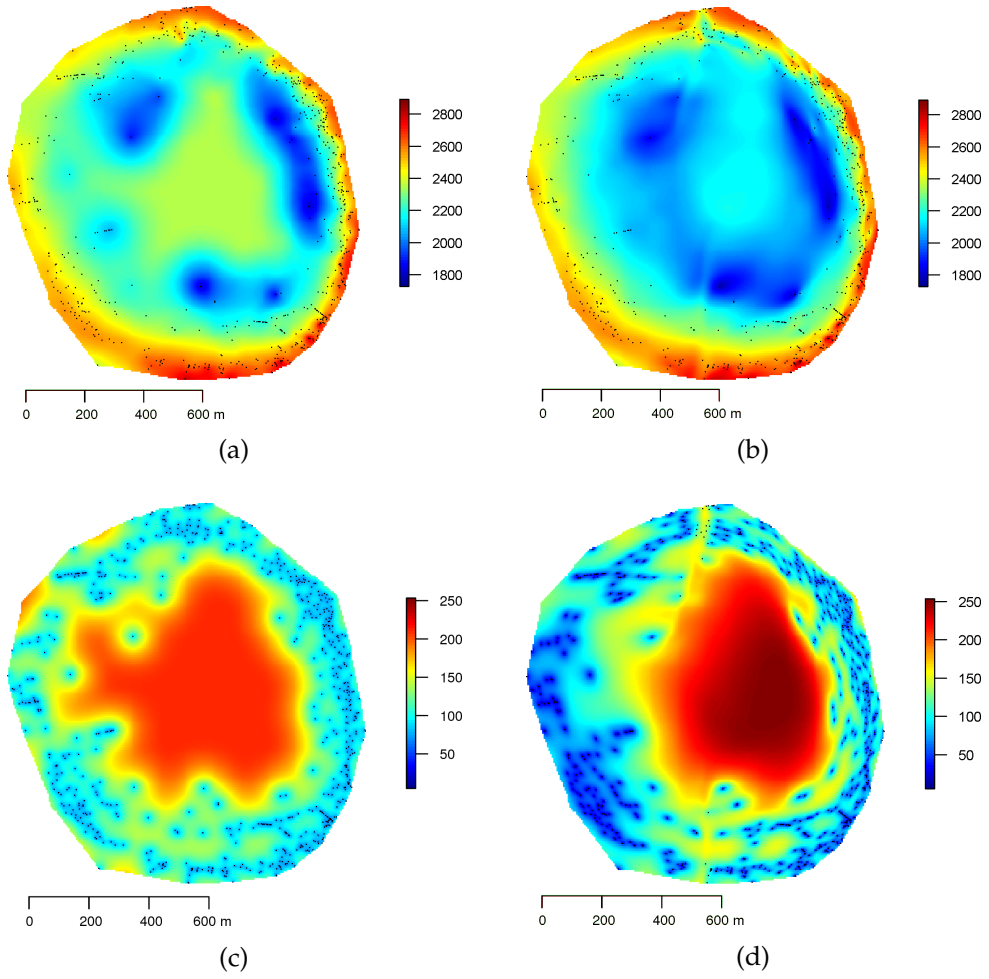
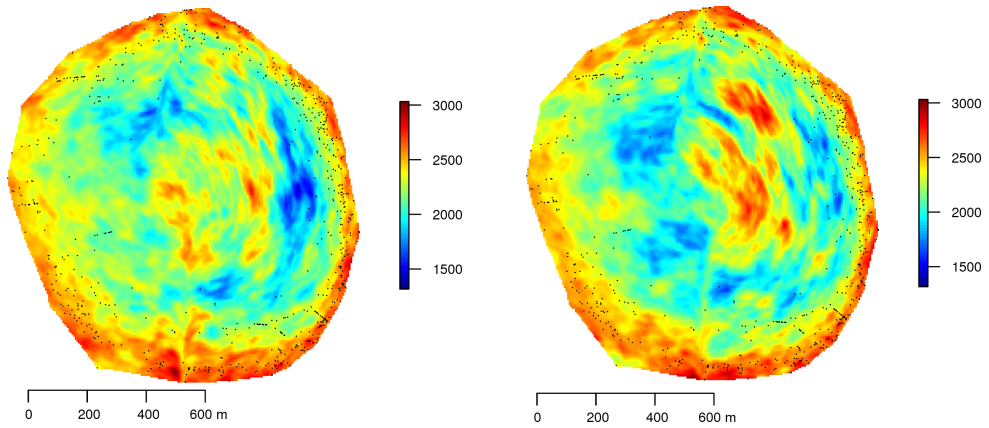
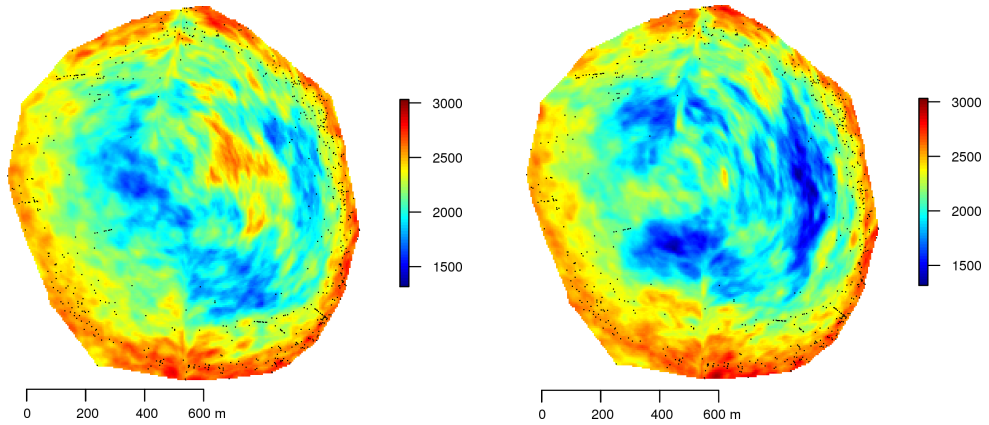


FIGURE 4.38 – Méthode de convolution : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; cartes d'écart-type de krigeage associées (c, d).



(a) Simulation #1 (b) Simulation #2



(c) Simulation #3 (d) Simulation #4

FIGURE 4.39 – Méthode de convolution : simulations conditionnelles.

Comparaison

Les scores de prédiction [6] calculés sur un échantillon de validation (200 observations) pour les méthodes stationnaire, déformation d'espace et convolution sont présentés au tableau 4.3. Les méthodes non-stationnaires sont nettement meilleures que celle stationnaire. Par rapport à l'approche stationnaire, on enregistre par exemple, une amélioration globale de RMSE de l'ordre de 23% pour l'approche par convolution et de 20% pour l'approche par déformation d'espace . La méthode de convolution est légèrement meilleure que celle par déformation d'espace en terme de RMSE.

6. Les scores de prédiction sont définis en annexe B.

	Stationnaire	Déformation	Convolution
MAE	79,74	64,83	61,39
RMSE	154,41	122,80	117,99
NMSE	0,98	1,07	0,74
LogS	2439	2365	2315
CRPS	123,59	98,84	121,11

TABLEAU 4.3 – Méthodes stationnaire/non-stationnaire : validation externe sur un échantillon de 200 observations.

4.5 Conclusion

Le présent chapitre a permis d'éprouver les méthodes non-stationnaires proposées au cours des deux précédents chapitres sur des données réelles. Force est de constater que les méthodes développées s'appliquent bien sur ces dernières. En tant qu'outils exploratoires, elles permettent de mettre en évidence des structures locales variables dans les données. S'agissant de la prédiction spatiale, elles ont montré des performances prédictives supérieures à l'approche stationnaire sur l'ensemble des trois jeux de données utilisés. Par ailleurs, nous avons effectué des simulations conditionnelles basées sur ces approches non-stationnaires. Nous avons ainsi pu faire ressortir l'intérêt d'utiliser de telles méthodes non-stationnaires.

Conclusions et perspectives

Les approches non-stationnaires suscitent un intérêt grandissant dans les applications géostatistiques car elles permettent de mieux caractériser et estimer la structure de dépendance spatiale sous-jacente d'un phénomène naturel. Dans ce travail, deux méthodes destinées à la prédiction et à la simulation conditionnelle de fonctions aléatoires non-stationnaires de second ordre ont été développées. Leur mise en œuvre sur des cas concrets et leur application à des problèmes tels que la prédiction ont été présentées (chapitre 4). Ces méthodes reposent sur les stratégies de modélisation de type déformation d'espace et convolution stochastique. Elles ont été développées dans l'environnement R [R Core Team, 2013] et principalement sur la base du package RGeostats [Renard et al., 2014].

En tant qu'outils exploratoires, les méthodes proposées permettent de mettre en évidence la non-stationnarité dans les données. S'agissant de la prédiction spatiale, elles ont montré des performances prédictives nettement supérieures à l'approche stationnaire (selon plusieurs critères), sur l'ensemble des données synthétiques et réelles utilisées. Elles prennent en compte certaines caractéristiques structurales locales des données qu'une approche stationnaire ne saurait capturer.

Les méthodes proposées ont une variété d'aspects intéressants. Comparativement aux autres méthodes relatives à la déformation d'espace, celle proposée (chapitre 2) ne nécessite pas de réplications. Elle a ainsi une grande portée d'application. En comparaison avec les méthodes existantes, elle détermine la déformation en n'utilisant qu'un ensemble réduit de points (points supports) au lieu de tous les points de données. Il s'agit d'une amélioration majeure pour les grands ensembles de données. En effet, l'utilisation de points supports rend possible la mise en œuvre de la procédure de positionnement multidimensionnel non-métrique pour de grands ensembles de données. En outre, l'approche intègre des contraintes spatiales qui aident à s'assurer que la déformation est bijective. Ce point était un inconvénient majeur lors de l'ajustement de la fonction de déformation. De plus, elle est facile à mettre en œuvre puisqu'elle permet d'utiliser tous les développements déjà réalisés dans le cadre stationnaire via la fonction de déformation. En effet, la prédiction spatiale et la simulation conditionnelle dans le contexte stationnaire sont des sujets déjà

bien maitrisés, pour lesquels des techniques rapides et robustes sont disponibles. La méthode est également un outil d'analyse exploratoire de la non-stationnarité, à travers la fonction de déformation qui résume cette non-stationnarité. La représentation de l'espace déformé permet d'identifier les régions de forte et faible continuité spatiale.

L'approche par convolution proposée (chapitre 3) fournit en terme de modélisation un nouveau modèle pour les fonctions aléatoires non-stationnaires de second ordre. Ce modèle généralise le modèle classique de convolution et permet d'obtenir des classes explicites et flexibles de covariances non-stationnaires. Pour ce qui est de l'inférence, la méthode proposée permet de traiter dans un cadre unique différents types de non-stationnarité (moyenne, variance et covariance). La procédure d'estimation s'affranchit de la stationnarité. Elle repose sur l'hypothèse moins forte de quasi-stationnarité. Cette méthode est aussi un outil d'analyse exploratoire de la non-stationnarité. En effet, les cartes de paramètres (moyenne, variance, azimut) permettent d'explorer et de décrire la non-stationnarité présente dans les données. Le nuage de points variance *versus* moyenne permet par exemple d'identifier la relation entre la moyenne et la variance connue sous le terme d'effet proportionnel [Chilès and Delfiner, 2012]. Les cartes d'azimut permettent de mettre en évidence l'anisotropie géométrique locale variable.

Bien que les deux approches proposées soient assez distinctes par la façon de prendre en compte la non-stationnarité, nous avons mis en évidence un lien entre elles par le biais d'applications. La carte de variance produite par la méthode de convolution est cohérente avec l'action de la fonction de déformation. Les zones de faible variance sont contractées tandis que celles de forte variance sont étirées. Notons par ailleurs que contrairement à la méthode par convolution, la méthode par déformation d'espace ne permet de prendre en compte les autres types de non-stationnarité (moyenne et variance).

Au terme de ce travail et de l'analyse du potentiel des méthodes non-stationnaires développées, il est possible d'énoncer quelques pistes de recherche pour les développements futurs :

Dans la méthode de déformation d'espace proposée, il serait intéressant d'étudier les propriétés statistiques (telle que la cohérence) de l'estimateur non-paramétrique à noyau du variogramme non-stationnaire. Ceci peut être fait dans le contexte asymptotique en suivant le travail de Garcìa-Soidàn et al. [2004] développé dans le cadre stationnaire. L'estimation de la déformation ne doit pas être limitée à la classe de fonctions actuellement utilisées. D'après notre expérience, les fonctions de base radiales de type spline plaque mince fonctionnent bien, mais toute autre une méthode d'interpolation peut se révéler utile. Il serait également intéressant de prendre en compte les informations sur des covariables dans le modèle de déformation d'espace, afin d'améliorer l'estimation de la déformation. Ceci peut être fait en se basant sur les travaux de Schmidt et al. [2011]. Un aspect intéressant

à regarder est la sélection des hyper-paramètres. En effet, les procédures de validation croisée utilisées restent numériquement coûteuses.

Dans la méthode de convolution proposée, l'estimation repose sur l'estimateur non-paramétrique à noyau du variogramme local. Afin de mieux s'adapter à la densité d'échantillonnage variable dans le domaine des observations, il serait intéressant de travailler avec un estimateur non-paramétrique à noyau localement adaptatif. L'idée est d'augmenter (respectivement de réduire) la largeur de fenêtre dans les zones à forte (respectivement à faible) densité d'échantillonnage. L'hypothèse de stationnarité locale qui sous-tend la méthodologie proposée rend la méthode fonctionnelle pour une non-stationnarité variant lentement. Cependant, elle peut être difficile à appliquer sur des données très éparpillés ou présentant des variations spatiales rapides ou brusques. Dans ces cas, il est souhaitable de travailler dans le cadre stationnaire. Il serait intéressant d'étendre cette méthode pour une prise en compte des covariables. Ceci peut se faire en paramétrant la fonction d'anisotropie et de variance en fonction des covariables. On peut pour cela s'inspirer des travaux de Neto et al. [2013] et Risser and Calder [2014]. Jusqu'à présent, l'approche de modélisation non-stationnaire par convolution ne permet pas d'obtenir des classes explicites de covariances non-stationnaires à support compact, ceci reste un problème ouvert. En effet, l'utilisation de telles covariances réduirait considérablement la charge de calcul du krigeage lorsqu'on traite de grands ensembles de données.

Au vu des résultats obtenus dans ce travail, nous pouvons répondre par l'affirmative à la question ci-après : la prise en compte de la non-stationnarité est-elle souhaitable dans les modèles géostatistiques ? En effet, des améliorations substantielles de prédiction et d'estimation des erreurs peuvent être obtenues en utilisant des modèles de structure de dépendance spatiale non-stationnaire.

Liste des figures

1.1 Modèle à point source : réalisations de fonctions aléatoires non-stationnaires. 23
1.2 Modèle de déformation d'espace (1D) : réalisation d'une fonction aléatoire non-stationnaire. 27
1.3 Modèle de déformation d'espace (2D) : réalisation d'une fonction aléatoire non-stationnaire. 28
1.4 Modèle de convolution : représentation de la matrice de noyau. 42
1.5 Modèle de convolution : présence d'effets de bord. 44
1.6 Modèle de convolution : correction d'effets bord. 45
1.7 Modèle de convolution : réalisations de fonctions aléatoires non-stationnaires. 45
1.8 Modèles locaux stationnaires pondérés : réalisation d'une fonction aléatoire non-stationnaire. 52
1.9 Modèle d'EDPS : réalisation d'une fonction aléatoire non-stationnaire.[7] . . 57

2.1 Modèle de déformation d'espace : exemples de noyaux. 67
2.2 Méthode de déformation d'espace : réalisation d'une fonction aléatoire non-stationnaire dans l'espace géographique (a) et l'espace déformé estimé (b) ; (c) déformation vraie et son estimation ; (d) points déformés vrais *versus* points déformés estimés. 80
2.3 Méthode de déformation d'espace : (a, b) variogrammes non-stationnaires estimé $\hat{\gamma}(.,.)$ et vrai $\gamma(.,.)$; (c, d) variogrammes transformés estimé $\hat{\gamma}_0(.)$ et vrai $\gamma_0(.)$. 81
2.4 Méthode de déformation d'espace : (a) réalisation d'une fonction aléatoire non-stationnaire par déformation d'espace ; (b) points de données et points supports (croix rouge) ; points de données dans l'espace déformé vrai (c) et l'espace déformé estimé (d). 82
2.5 Méthode de déformation d'espace : (a) fonction score $CV(\lambda)$ en validation croisée ; fonction score $MSE(\lambda, \omega)$ en validation croisée (b) et externe (d) ; (c) diagramme de Shepard de la solution NMDS. 83
2.6 Méthode de déformation d'espace : variogramme d'un point avec tous les autres points pour les modèles M1 (a, b), M2 (c, d) et M3 (e, f). 85

2.7 Méthode de déformation d'espace : contours du variogramme en quelques points pour les modèles M1 (a), M2 (b) et M3 (c). Les isolignes correspondent aux valeurs : $0,1$ (noir), $0,3$ (rouge) et $0,5$ (vert). 86

2.8 Méthode de déformation d'espace : (a, c, e) résultats d'interpolation pour les modèles M1, M2 et M3 ; (b, d, f) cartes d'écart type de krigeage correspondantes. 87

3.1 Modèle de convolution : (a) données exhaustives ; (b) données d'apprentissage. 108

3.2 Méthode de convolution : (a) fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports ; (b) fonction de variance estimée $\widehat{\sigma}^2(.)$ aux points supports. . . . 108

3.3 Méthode de convolution : fonction de score $MSE(\epsilon)$ en validation croisée (a) et externe (b). 109

3.4 Méthode de convolution : paramètres lissés sur le domaine des observations : (a) variance, (b) ratio d'anisotropie, (c) azimut. 110

3.5 Méthode de convolution : covariance d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d). 111

3.6 Méthode de convolution : contours de la covariance en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 4 (noir), 2 (rouge) et 1 (vert). 111

3.7 Méthode de convolution : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; cartes d'écart-type de krigeage correspondantes (c, d). 112

4.1 Méthode de déformation d'espace : (a, c) données d'apprentissage dans l'espace géographique et modèle de variogramme stationnaire estimé ; (b, d) données d'apprentissage dans l'espace déformé et modèle de variogramme transformé estimé. 117

4.2 Méthode de déformation d'espace : (a) fonction score $CV(\lambda)$ de validation croisée ; (c) diagramme de Shepard de la solution NMDS ; fonction score $MSE(\lambda, \omega)$ en validation croisée (b) et externe (d). 118

4.3 Méthode de déformation d'espace : variogramme d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d). 119

4.4 Méthode de déformation d'espace : contours du variogramme en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 70 (noir), 90 (rouge) et 110 (vert).119

4.5 Méthode de déformation d'espace : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; (c, d) cartes d'écart-type de krigeage associées. 120
4.6 Méthode de déformation d'espace : simulations conditionnelles. 121
4.7 Méthode de convolution : (a) données d'apprentissage ; (b) fonction moyenne estimée $\widehat{\mathrm{m}}(.)$ aux points supports ; (c) fonction variance estimée $\widehat{\sigma}^2(.)$ aux points supports ; (d) fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports. 122
4.8 Méthode de convolution : fonction de score $MSE(\epsilon)$ en validation croisée (a) et externe (b). 123
4.9 Méthode de convolution : paramètres lissés sur le domaine des observations : (a) moyenne, (b) variance, (c) ratio d'anisotropie, (d) azimut. 123
4.10 Méthode de convolution : covariance d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d). 124
4.11 Méthode de convolution : contours de la covariance en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 80 (noir), 60 (rouge) et 40 (vert). . . . 124
4.12 Méthode de convolution : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; cartes d'écart-type de krigeage associées (c, d). 125
4.13 Méthode de convolution : simulations conditionnelles. 126
4.14 Méthode de déformation d'espace : (a, c) données d'apprentissage dans l'espace géographique et modèle de variogramme stationnaire estimé ; (b, d) données d'apprentissage dans l'espace déformé et modèle de variogramme transformé estimé. 128
4.15 Méthode de déformation d'espace : (a) fonction score $CV(\lambda)$ de validation croisée ; (c) diagramme de Shepard de la solution NMDS ; fonction score $MSE(\lambda, \omega)$ en validation croisée (b) et externe (d). 129
4.16 Méthode de déformation d'espace : variogramme d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d). 130
4.17 Méthode de déformation d'espace : contours du variogramme en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 4000 (noir), 6000 (rouge) et 8000 (vert). 130
4.18 Méthode de déformation d'espace : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; (c, d) cartes d'écart-type de krigeage associées. 131
4.19 Méthode de déformation d'espace : simulations conditionnelles. 132

4.20 Méthode de convolution : (a) données d'apprentissage ; (b) fonction moyenne estimée $\widehat{\mathrm{m}}(.)$ aux points supports ; (c) fonction variance estimée $\widehat{\sigma}^2(.)$ aux points supports ; (d) fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports. 133
4.21 Méthode de convolution : fonction de score $MSE(\epsilon)$ en validation croisée (a) et externe (b). 133
4.22 Méthode de convolution : paramètres lissés sur le domaine des observations : (a) moyenne, (b) variance, (c) ratio d'anisotropie, (d) azimut. 134
4.23 Méthode de convolution : covariance d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d). 135
4.24 Méthode de convolution : contours de la covariance en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 10000 (noir), 8000 (rouge) et 6000 (vert). 135
4.25 Méthode de convolution : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; cartes d'écart-type de krigeage associées (c, d). 136
4.26 Méthode de convolution : simulations conditionnelles. 137
4.27 Méthode de déformation d'espace : (a, c) données d'apprentissage dans l'espace géographique et modèle de variogramme stationnaire estimé ; (b, d) données d'apprentissage dans l'espace déformé et modèle de variogramme transformé estimé. 139
4.28 Méthode de déformation d'espace : (a) fonction score $CV(\lambda)$ de validation croisée ; (c) diagramme de Shepard de la solution NMDS ; fonction score $MSE(\lambda, \omega)$ en validation croisée (b) et externe (d). 140
4.29 Méthode de déformation d'espace : variogramme d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d). 141
4.30 Méthode de déformation d'espace : contours du variogramme en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 10000 (noir), 20000 (rouge) et 30000 (vert). 141
4.31 Méthode de déformation d'espace : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; (c, d) cartes d'écart-type de krigeage associées. 142
4.32 Méthode de déformation d'espace : simulations conditionnelles. 143
4.33 Méthode de convolution : (a) données d'apprentissage ; (b) fonction moyenne estimée $\widehat{\mathrm{m}}(.)$ aux points supports ; (c) fonction variance estimée $\widehat{\sigma}^2(.)$ aux points supports ; (d) fonction d'anisotropie estimée $\widehat{\mathbf{\Sigma}}(.)$ aux points supports. 144

4.34 Méthode de convolution : fonction de score $MSE(\epsilon)$ en validation croisée (a) et externe (b). 145
4.35 Méthode de convolution : paramètres lissés sur le domaine des observations : (a) moyenne, (b) variance, (c) ratio d'anisotropie, (d) azimut. 146
4.36 Méthode de convolution : covariance d'un point avec tous les autres points pour les modèles stationnaire estimé (a, b) et non-stationnaire estimé (c, d). 147
4.37 Méthode de convolution : contours de la covariance en quelques points pour les modèles stationnaire estimé (a) et non-stationnaire estimé (b). Les isolignes correspondent aux valeurs : 30000 (noir), 20000 (rouge) et 10000 (vert). 147
4.38 Méthode de convolution : (a, b) résultats d'interpolation pour le modèle stationnaire estimé et le modèle non-stationnaire estimé ; cartes d'écart-type de krigeage associées (c, d). 148
4.39 Méthode de convolution : simulations conditionnelles. 149

Liste des tableaux

1.1 Modèle de convolution : exemples de noyaux et leur covariance associée. . 39

2.1 Modèle de déformation d'espace : évaluation du *stress*. 72
2.2 Méthode de déformation d'espace : validation externe sur un échantillon de 1024 observations. 88

3.1 Méthode de convolution : validation externe sur un échantillon de 1024 observations. 113

4.1 Méthodes stationnaire/non-stationnaire : validation externe sur un échantillon de 1000 observations. 127
4.2 Méthodes stationnaire/non-stationnaire : validation externe sur un échantillon de 67 observations. 137
4.3 Méthodes stationnaire/non-stationnaire : validation externe sur un échantillon de 200 observations. 150

Bibliographie

Abrahamsen, P. (1997). *A Review of Gaussian Random Fields and Correlation Functions*. Norsk Regnesentral/Norwegian Computing Center.

Adler, R. and Taylor, J. (2009). *Random Fields and Geometry*. Springer.

Allard, D. (2006). Validation d'un modèle géostatistique pour l'interpolation : application à un évènement pluvieux. In *Analyse statistique des données spatiales*, pages 403–414. Technip.

Almendral, A., Abrahamsen, P., and Hauge, R. (2008). Multidimensional scaling and anisotropic covariance functions. In *Proceedings of the Eight International Geostatistics Congress*, pages 187–196.

Anderes, E. B. and Chatterjee, S. (2009). Consistent estimates of deformed isotropic gaussian random fields on the plane. *Annals of Statistics*, 37(5) :2324–2350.

Anderes, E. B. and Stein, M. L. (2008). Estimating deformations of isotropic gaussian random fields on the plane. *Annals of Statistics*, 36 :719–741.

Anderes, E. B. and Stein, M. L. (2011). Local likelihood estimation for nonstationary random fields. *Journal of Multivariate Analysis*, 102(3) :506–520.

Atkinson, P. M. and Lloyd, C. D. (2007). Non-stationary variogram models for geostatistical sampling optimisation : An empirical investigation using elevation data. *Computers & Geosciences*, 33(10) :1285–1300.

Banerjee, S., Gelfand, A. E., Knight, J. R., and Sirmans, C. F. (2004). Spatial modeling of house prices using normalized distance-weighted sums of stationary processes. *Journal of Business & Economic Statistics*, 22(2) :206–213.

Barber, J. J. (2002). *Modeling And Prediction Of Nonstationary Spatial Environmental Processes*. PhD thesis, Nc State University.

Bel, L. (2004). *Non Parametric Variogram Estimator : Application to Air Pollution Data*, volume 13 of *Quantitative Geology and Geostatistics*, chapter 3, pages 29–40. Springer Netherlands.

Billingsley, P. (1995). *Probability and measure*. Wiley, third edition.

Boisvert, J. B. and Deutsch, C. V. (2011). Programs for kriging and sequential gaussian simulation with locally varying anisotropy using non-euclidean distances. *Computers & Geosciences*, 37(4) :495–510.

Boisvert, J. B., Manchuk, J. G., and Deutsch, C. V. (2009). Kriging in the presence of locally varying anisotropy using non-euclidean distances. *Mathematical Geosciences*, 41(5) :585–601.

Bolin, D. and Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Ann. Appl. Stat.*, 5(1) :523–550.

Borg, I. and Groenen, P. (2005). *Modern Multidimensional Scaling : Theory and Applications*. Springer Series in Statistics. Springer.

Borg, I., Groenen, P., and Mair, P. (2012). *Applied Multidimensional Scaling*. Springer.

Bornn, L., Shaddick, G., and Zidek, J. V. (2012). Modeling nonstationary processes through dimension expansion. *Journal of the American Statistical Association*, 107(497) :281–289.

Calder, C. A. and Cressie, N. (2007). Some topics in convolution-based spatial modeling. In *Proceedings of the 56th Session of the International Statistical Institute*.

Castro Morales, F., Gamerman, D., and Paez, M. (2013). State space models with spatial deformation. *Environmental and Ecological Statistics*, 20(2) :191–214.

Chang, Y.-M., Hsu, N.-J., and Huang, H.-C. (2010). Semiparametric estimation and selection for nonstationary spatial covariance functions. *Journal of Computational and Graphical Statistics*, 19(1) :117–139.

Chilès, J. P. and Delfiner, P. (2012). *Geostatistics : modeling spatial uncertainty*. Wiley.

Christakos, G. (2012). *Random Field Models in Earth Sciences*. Dover Publications.

Cohen, A. and Jones (1969). Regression on a random field. *Journal of the American Statistical Association*, 64 :1172–1182.

Cox, T. and Cox, A. (2000). *Multidimensional Scaling, Second Edition*. Taylor & Francis.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 70 :209–226.

Damian, D., Sampson, P. D., and Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, 12(2) :161–178.

Desassis, N. and Renard, D. (2012). Automatic variogram modeling by iterative least squares : Univariate and multivariate cases. *Mathematical Geosciences*, pages 1–18.

D'Hondt, O., López-Martínez, C., Ferro-Famil, L., and Pottier, E. (2007). Spatially nonstationary anisotropic texture analysis in sar images. *IEEE T. Geoscience and Remote Sensing*, 45(12-1) :3905–3918.

Diggle, P. and Ribeiro, P. (2007). *Model-based geostatistics*. Springer.

Dryden, I. L. and Mardia, K. (1998). *Statistical Shape Analysis*. John Wiley & Sons.

Ecker, M., De Oliveira, V., and Isakson, H. (2013). A note on a non-stationary point source spatial model. *Environmental and Ecological Statistics*, 20(1) :59–67.

Ecker, M. D. and Oliveira, V. D. (2008). Bayesian spatial modeling of housing prices subject to a localized externality. *Communications in Statistics - Theory and Methods*, 37(13) :2066–2078.

Felder, J. (2011). *Développement de méthodes de traitement d'images pour la détermination de paramètres variographiques locaux*. PhD thesis, École Nationale Supérieure des Mines de Paris.

Fuentes, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics*, 12(5) :469–483.

Fuentes, M. (2002a). Interpolation of nonstationary air pollution processes : a spatial spectral approach. *Statistical Modelling : An International Journal*, 2(4) :281–298.

Fuentes, M. (2002b). Spectral methods for nonstationary spatial processes. *Biometrika*, 89(1) :197–210.

Fuglstad, G.-A. S. L. (2013). Non-stationary spatial modelling with applications to spatial prediction of precipitation. *arXiv :1306.0408*.

Fuglstad, G.-A. S. L. R. (2014). Do we need non-stationarity in spatial models ? *arXiv :1409.0743*.

Gaetan, C. and Guyon, X. (2009). *Spatial Statistics and Modeling*. Springer.

Garcìa-Soidàn, P. H., Febrero-Bande, M., and Gonzalez-Manteiga, W. (2004). Nonparametric kernel estimation of an isotropic variogram. *Journal of Statistical Planning and Inference*, 121(1) :65–92.

Gelfand, A. E., Diggle, P., and Guttorp, P. (2010). *Handbook of spatial statistics*. Taylor & Francis Group.

Genton, M. G. and Perrin, O. (2004). On a time deformation reducing nonstationary stochastic processes to local stationarity. *Journal of Applied Probability*, 41(1) :236–249.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477) :359–378.

Gosoniu, L. and Vounatsou, P. (2011). Non-stationary partition modeling of geostatistical data for malaria risk mapping. *Journal of Applied Statistics*, 38(1) :3–13.

Gosoniu, L., Vounatsou, P., Sogoba, N., Maire, N., and Smith, T. (2009). Mapping malaria risk in west africa using a bayesian nonparametric non-stationary model. *Computational Statistics & Data Analysis*, 53(9) :3358–3371.

Gradshteyn and Ryzhik (2007). *Table of Integrals, Series, and Products*. Elsevier Science.

Guillot, G., Senoussi, R., and Monestiez, P. (2001). *A Positive Definite Estimator of the Non Stationary Covariance of Random Fields*, volume 11 of *Quantitative Geology and Geostatistics*, pages 333–344. Springer Netherlands.

Guttorp, P. and Schmidt, A. M. (2013). Covariance structure of spatial and spatiotemporal processes. *Wiley Interdisciplinary Reviews : Computational Statistics*, 5(4) :279–287.

Haas, T. C. (1990a). Kriging and automated variogram modeling within a moving window. *Atmospheric Environment. Part A. General Topics*, 24(7) :1759–1769.

Haas, T. C. (1990b). Lognormal and moving window methods of estimating acid deposition. *Journal of the American Statistical Association*, 85(412) :950–963.

Harris, P., Charlton, M., and Fotheringham, A. S. (2010). Moving window kriging with geographically weighted variograms. *Stochastic Environmental Research and Risk Assessment*, 24(8) :1193–1209.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, 5(2) :173–190.

Higdon, D. (2002). *Space and space-time modeling using process convolutions*, pages 37–56. Springer, London, springer edition.

Higdon, D., Swall, J., and Kern, J. (1999). *Non-stationary spatial modeling*. Bayesian Statistics 6. Oxford Univ Press, New York.

Holland, D., Saltzman, N., Cox, L. H., and Nychka, D. (1999). Spatial prediction of sulfur dioxide in eastern united states. In *geoENV-II - geostatistics for environmental applications*, pages 65–76. Kluwer.

Hughes-Oliver, J., Heo, T.-Y., and Ghosh, S. (2009). An autoregressive point source model for spatial processes. *Environmetrics*, 20(5) :575–594.

Hughes-Oliver, J. M. and Gonzàlez-Farìas, G. (1999). Parametric covariance models for shock-induced stochastic processes. *Journal of Statistical Planning and Inference*, 77(1) :51–72.

Hughes-Oliver, J. M., Gonzalez-Farias, G., Lu, J. C., and Chen, D. (1998a). Parametric nonstationary correlation models. *Statistics & Probability Letters*, 40(3) :267–278.

Hughes-Oliver, J. M., Lu, J. C., Davis, J. C., and Gyurcsik, R. S. (1998b). Achieving uniformity in a semiconductor fabrication process using spatial modeling. *Journal of the American Statistical Association*, 93(443) :1252–1252.

Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8 :20–38.

Iovleff, S. and Perrin, O. (2004). Estimating a nonstationary spatial structure using simulated annealing. *Journal of Computational and Graphical Statistics*, 13(1) :90–105.

Kern, J. C. (2000). *Bayesian process-convolution approaches to specifying spatial dependence structure*. PhD thesis, Duke University.

Kim, H. M., Mallick, B. K., and Holmes, C. C. (2005). Analyzing nonstationary spatial data using piecewise gaussian processes. *Journal of the American Statistical Association*, 100(470) :653–668.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1) :1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling : A numerical method. *Psychometrika*, 29(2) :115–129.

Lantuéjoul, C. (2002). *Geostatistical Simulation : Models and Algorithms*. Springer.

Lantuéjoul, C. and Desassis, N. (2012). Simulation of a gaussian random vector : A propagative version of the gibbs sampler. In *The 9th International Geostatistics Congress*.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields : the stochastic partial differential equation approach. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73(4) :423–498.

Lloyd, C. D. (2011). *Local models for spatial analysis*. CRC Press, Boca Raton, Fla.

Machuca-Mory, D. and Deutsch, C. (2012). Non-stationary geostatistical modeling based on distance weighted statistics and distributions. *Mathematical Geosciences*, pages 1–18.

Magneron, C., Jeannee, N., Le Moine, O., and Bourillet, J. F. (2010). *Integrating Prior Knowledge and Locally Varying Parameters with Moving-GeoStatistics : Methodology and Application to Bathymetric Mapping*, volume 16 of *geoENV VII - Geostatistics for Environmental Applications*. Springer, Dordrecht.

Mardia, K. and Goodall, C. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In *Multivariate Environmental Statistics*, pages 347–386. Elsevier Science Publishers.

Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate Analysis*. Academic Press, London.

Matérn, B. (1986). *Spatial Variation*. Lecture Notes in Statistics. Springer-Verlag, New york.

Mateu, J., Fernandez-Avilas, G., and Montero, J. (2010). On a class of non-stationary, compactly supported spatial covariance functions. *Stochastic Environmental Research and Risk Assessment*, pages 1–13.

Matheron, G. F. (1971). *The theory of regionalized variables and its applications*, volume 5 of *Les cahiers du Centre de Morhologie Mathématique de Fontainebleau*. École Nationale Supérieure des Mines de Paris.

Matsuo, T., Nychka, D., and Paul, D. (2011). Nonstationary covariance modeling for incomplete data : Monte carlo em approach. *Computational Statistics & Data Analysis*, 55(6) :2059–2073.

McBratney, A. B. and Minasny, B. (2013). Spacebender. *Spatial Statistics*, 4(0) :57–67.

McLeish, D. L. (1982). A robust alternative to the normal distribution. *Canadian Journal of Statistics*, 10(2) :89–102.

Meiring, W., Monestiez, P., Sampson, P. D., and Guttorp, P. (1997). *Developments In The Modelling Of Nonstationary Spatial Covariance structure From Space-Time Monitoring Data*, volume 1 of *Quantitative Geology and Geostatistics*, pages 162–173. Kluwer Academic Publisher, Dordrecht.

Neto, J. H. V., Schmidt, A. M., and Guttorp, P. (2013). Accounting for spatially varying directional effects in spatial covariance structures. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, pages 103–122.

Nott, D. J. and Dunsmuir, W. T. M. (2002). Estimation of nonstationary spatial covariance structure. *Biometrika*, 89(4) :819–829.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2014). A multiresolution gaussian process model for the analysis of large spatial data sets. *Journal of Computational and Graphical Statistics*, pages xxx–xxx.

Nychka, D. and Saltzman, N. (1998). Design of air quality networks. In *Case Studies in Environmental Statistics*, volume 132, pages 51–76. Lectures Notes in Statistics Springer Verlag.

Nychka, D., Wikle, C., and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling : An International Journal*, 2(4) :315–331.

Oehlert, G. W. (1993). Regional trends in sulfate wet deposition. *Journal of the American Statistical Association*, 88(422) :390–399.

Oliver, D. (1995). Moving averages for gaussian simulation in two and three dimensions. *Mathematical Geology*, 27(8) :939–960.

Paciorek, C. and Schervish, M. (2004). Nonstationary covariance functions for gaussian process regression. In *Proc. of the Conf. on Neural Information Processing Systems (NIPS)*, pages 273–280. MIT Press.

Paciorek, C. J. (2003). *Nonstationary Gaussian Processes For Regression And Spatial Modelling*. PhD thesis, Carnegie Mellon University.

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

Perrin, O. and Meiring, W. (1999). Identifiability for non-stationary spatial structure. *Journal of Applied Probability*, 36(4) :1244–1250.

Perrin, O. and Meiring, W. (2003). Nonstationarity in r-n is second-order stationarity in r-2n. *Journal of Applied Probability*, 40(3) :815–820.

Perrin, O. and Monestiez, P. (1998). *Modeling of non-stationary spatial covariance structure by parametric radial basis deformations*, volume 11 of *Quantitative Geology and Geostatistics*, pages 175–186. Springer Netherlands.

Perrin, O. and Senoussi, R. (2000). Reducing non-stationary random fields to stationarity and isotropy using a space deformation. *Statistics & Probability Letters*, 48(1) :23–32.

Pintore, A. and Holmes, C. (2004). Spatially adaptive non-stationary covariance functions via spatially adaptive spectra. Technical report, University of Oxford.

Porcu, E., Mateu, J., and Christakos, G. (2009). Quasi-arithmetic means of covariance functions with potential applications to space-time data. *Journal of Multivariate Analysis*, 100(8) :1830–1844.

Porcu, E., Matkowski, J., and Mateu, J. (2010). On the non-reducibility of non-stationary correlation functions to stationary ones under a class of mean-operator transformations. *Stochastic Environmental Research and Risk Assessment*, 24(5) :599–610.

R Core Team (2013). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Renard, D., Bez, N., Desassis, N., Beucher, H., and Ors, F. (2014). *RGeostats : Geostatistical Package*. R package version 10.0.5.

Risser, M. D. and Calder, C. A. (2014). Regression-based covariance functions for nonstationary spatial modeling. *arXiv :1410.1494*.

Sampson, P. D. (2006). *Spatial Covariance*, pages 2059–2067. John Wiley & Sons, Ltd.

Sampson, P. D. (2010). *Constructions for Nonstationary Spatial Processes*, pages 119–130. Taylor & Francis Group.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric-estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417) :108–119.

Schabenberger, O. and Gotway, C. A. (2005a). Non-stationary covariance. In *Statistical methods for spatial data analysis*, pages 421–430. Chapman & Hall/CRC.

Schabenberger, O. and Gotway, C. A. (2005b). *Statistical methods for spatial data analysis*. Chapman & Hall/CRC.

Schmidt, A. M., Guttorp, P., and O'Hagan, A. (2011). Considering covariates in the covariance structure of spatial processes. *Environmetrics*, 22(4) :487–500.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 65 :743–758.

Schoenberg, I. (1938). Metric spaces and positive definite functions. *Transactions of the American Statistical Association*, 44 :522–536.

Séguret, S. A. and Celhay, F. (2013). Geometric modeling of a breccia pipe - comparing five approaches. *Apcom - Application of Computers and Operations research in the Mineral Industry*, 1 :257–266.

Smith, R. L. (1996). Estimating nonstationary spatial correlations. Technical report, University of North Carolina.

Stein, M. (2005). Nonstationary spatial covariance functions. *http ://www-personal.umich.edu/ jizhu/jizhu/covar/Stein-Summary.pdf*.

Stein, M. L. (1999). *Interpolation of spatial data : some theory for kriging*. Springer.

Stephenson, J., Holmes, C., Gallagher, K., and Pintore, A. (2005). *A statistical technique for modelling non-stationary spatial processes*, volume 14 of *Geostatistics Banff 2004, Vols 1 and 2*. Springer, Dordrecht.

Swall, J. (1999). *Non-Stationary Spatial Modeling Using A Process Convolution Approach*. PhD thesis, Duke University.

Vanmarcke, E. (2010). *Random Fields : Analysis and Synthesis*. World Scientific.

Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69(2) :275–294.

Vera, J., Macias, R., and Angulo, J. (2008). Non-stationary spatial covariance structure estimation in oversampled domains by cluster differences scaling with spatial constraints. *Stochastic Environmental Research and Risk Assessment*, 22(1) :95–106.

Vera, J. F., Macias, R., and Angulo, J. M. (2009). A latent class mds model with spatial constraints for non-stationary spatial covariance estimation. *Stochastic Environmental Research and Risk Assessment*, 23(6) :769–779.

Wand, M. and Jones, C. (1995). *Kernel smoothing*. Monographs on Statistics and Applied Probability. Chapman and Hall.

Whittle, P. (1963). Stochastic processes in several dimensions. *Bull. Internat.Statist. Inst.*, 40 :974–994.

Wolpert, R. L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika*, 85(2) :251–267.

Yaglom, A. and Silverman, R. (2004). *An Introduction to the Theory of Stationary Random Functions*. Dover Publications.

Zhang, H. and Wang, Y. (2010). Kriging and cross-validation for massive spatial data. *Environmetrics*, 21(3-4) :290–304.

Zhu, Z. and Wu, Y. (2010). Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data. *Journal of Computational and Graphical Statistics*, 19(1) :74–95.

Annexe A

Rappels sur les fonctions aléatoires

Cette annexe est un simple rappel des définitions et propriétés relatives aux fonctions aléatoires ; elle est synthétique et non exhaustive. Pour une présentation sur la théorie des fonctions aléatoires voir par exemple : Matheron [1971], Yaglom and Silverman [2004], Adler and Taylor [2009], Vanmarcke [2010] et Christakos [2012].

Définitions

Fonction aléatoire. Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et G un domaine de $\mathbb{R}^p$. Une fonction aléatoire sur G à valeurs dans $\mathbb{R}$ est une fonction de deux variables, notée $Z(\mathbf{x}, \omega)$, telle que pour chaque $\mathbf{x}_0 \in G$, $Z(\mathbf{x}_0, .)$ est une variable aléatoire sur $(\Omega, \mathcal{A}, P)$ et pour chaque $\omega_0 \in \Omega$, $Z(., \omega_0)$ est une fonction de $G \to \mathbb{R}$.

En règle générale, on se contente de noter la fonction aléatoire $Z(\mathbf{x}, \omega)$ par $\{Z(\mathbf{x}), \mathbf{x} \in G\}$ ou $Z(.)$ ou même Z.

Lois fini-dimensionnelles. Une fonction aléatoire est caractérisée par les distributions fini-dimensionnelles pour tout ensemble fini de points $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}, n \in \mathbb{N}^\star$:

$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(v_1, \ldots, v_n) = P\left(Z(\mathbf{x}_1) \leq v_1, \ldots, Z(\mathbf{x}_n) \leq v_n\right), \tag{A.1}$$

satisfaisant les conditions de Kolmogorov de symétrie et de cohérence. Ces conditions imposent que :

1. la loi de distribution fini-dimensionnelle doit être invariante par permutation : $F_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(v_1, \ldots, v_n) = F_{\mathbf{x}_{\tau(1)},\ldots,\mathbf{x}_{\tau(n)}}(v_{\tau(1)}, \ldots, v_{\tau(n)})$, $\tau(.)$ étant une permutation de $\{1, \ldots, n\}$ (condition de symétrie) ;

2. $F_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(v_1, \ldots, v_m, \infty, \ldots, \infty) = F_{\mathbf{x}_1,\ldots,\mathbf{x}_m}(v_1, \ldots, v_m), n > m$ (condition de cohérence).

Inversement une famille de distributions fini-dimensionnelles (A.1) vérifiant les conditions de Kolmogorov peut être considérée comme définissant une fonction aléatoire. L'ensemble des distributions fini-dimensionnelles, pour tout entier n et tous les choix possibles de $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ dans G constitue la loi spatiale de la fonction aléatoire. La donnée de la loi spatiale est une information extrêmement riche qui fournit de très nombreuses caractéristiques sur $Z(.)$. Billingsley [1995] montre que certaines propriétés comme la continuité ou la dérivabilité des réalisations de $Z(.)$, la loi des valeurs supérieure et inférieure de $Z(.)$ sur un ensemble, la connexité des ensembles où $Z(.)$ est supérieure à une valeur donnée, échappent à la loi spatiale. Si la distribution jointe F est une distribution multivariée gaussienne, alors $Z(.)$ est appelée fonction aléatoire gaussienne.

Réalisation. On appelle réalisation d'une fonction aléatoire $\{Z(\mathbf{x}), \mathbf{x} \in G\}$, désignée par $\{z(\mathbf{x}), \mathbf{x} \in G\}$ l'ensemble des valeurs numériques $z(\mathbf{x}_i)$ obtenues, pour tous les points $\mathbf{x}_i$ de l'espace, par un tirage au sort particulier effectué selon la loi F. Il y a le même rapport entre une fonction aléatoire et une de ses réalisations qu'entre une variable aléatoire Z et une valeur numérique z obtenue par un tirage au sort selon la loi de répartition de Z. Matheron [1971] appelle $\{z(\mathbf{x}), \mathbf{x} \in G\}$ variable régionalisée pour mettre l'accent sur la nature spatiale de l'espace des index G. Inversement, une variable régionalisée observée peut être interprétée comme une réalisation particulière d'une certaine fonction aléatoire.

Stationnarité stricte. Une fonction aléatoire $Z(.)$ est dite strictement stationnaire si toutes les lois fini-dimensionnelles sont invariantes par translation, c'est-à-dire :

$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(v_1, \ldots, v_n) = F_{\mathbf{x}_1+\mathbf{h},\ldots,\mathbf{x}_n+\mathbf{h}}(v_1, \ldots, v_n), \tag{A.2}$$

pour tout $n \in \mathbb{N}^\star$, tout ensemble de points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ et tout vecteur $\mathbf{h} \in \mathbb{R}^p$.

Dans ce cas, les lois fini-dimensionnelles F ne dépendent que de la configuration relative des points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ et non de leur position. Cette hypothèse présente deux défauts majeurs. En effet, elle ne nous renseigne pas sur l'existence ou non des moments de $Z(.)$ et n'est que partiellement vérifiable puisqu'elle est faite pour tout n.

Moments. On appelle moments d'une fonction aléatoire $Z(.)$ la famille $\{M_{\alpha_1\ldots\alpha_n}(\mathbf{x}_1, \ldots, \mathbf{x}_n)\}$ définie pour tous les entiers n, les points d'appui $\mathbf{x}_1, \ldots, \mathbf{x}_n$ et les entiers $\alpha_1, \ldots, \alpha_n$, par :

$$M_{\alpha_1\ldots\alpha_n}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathbb{E}\left([Z(\mathbf{x}_1)]^{\alpha_1} \ldots [Z(\mathbf{x}_n)]^{\alpha_n}\right). \tag{A.3}$$

En particulier le moment d'ordre 1 $m(\mathbf{x})$ (moyenne), ne dépend que de $\mathbf{x}$ et le moment d'ordre 2 est une fonction $M_{1,1}(\mathbf{x},\mathbf{y})$ de deux variables, symétriques en $\mathbf{x}$ et $\mathbf{y}$; on note $C(\mathbf{x},\mathbf{y})$ la covariance des variables aléatoires $Z(\mathbf{x})$ et $Z(\mathbf{y})$, et $\sigma(\mathbf{x})$ la variance de $Z(\mathbf{x})$:

$$C(\mathbf{x},\mathbf{y}) = M_{1,1}(\mathbf{x},\mathbf{y}) - m(\mathbf{x})m(\mathbf{y}) = \mathbb{C}\mathrm{ov}\left(Z(\mathbf{x}),Z(\mathbf{y})\right) \tag{A.4}$$

$$\sigma^2(\mathbf{x}) = M_2(\mathbf{x}) - m(\mathbf{x})^2 = \mathbb{V}\left(Z(\mathbf{x})\right). \tag{A.5}$$

Naturellement, ces divers moments n'existent pas toujours. Lorsque la variance $\sigma^2(\mathbf{x})$ existe en tout point $\mathbf{x}$ ($\sigma^2(\mathbf{x}) < +\infty$), la covariance $C(\mathbf{x},\mathbf{y})$ existe aussi pour les points $\mathbf{x}$ et $\mathbf{y}$. La fonction aléatoire $Z(.)$ est alors dite de variance finie ou d'ordre 2.

Définie positivité. La fonction de covariance $C(\mathbf{x},\mathbf{y})$ d'une fonction aléatoire d'ordre 2 $Z(.)$ est définie positive :

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j C(\mathbf{x}_i,\mathbf{x}_j) \geq 0, \tag{A.6}$$

pour tout $n \in \mathbb{N}^\star$, tout ensemble de points d'appui $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\} \subset \mathbb{R}^p$ et tout ensemble de réels $\{\lambda_1,\ldots,\lambda_n\}$.

Stationnarité d'ordre 2. Une fonction aléatoire $Z(.)$ est dite stationnaire d'ordre 2 ou de second ordre si ses deux premiers moments existent et vérifient :

$$\mathbb{E}(Z(\mathbf{x})) = m \in \mathbb{R},\ \forall \mathbf{x} \in \mathbb{R}^p \tag{A.7}$$

$$\mathbb{C}\mathrm{ov}(Z(\mathbf{x}),Z(\mathbf{y})) = C(\mathbf{x}-\mathbf{y}),\ \forall(\mathbf{x},\mathbf{y}) \in \mathbb{R}^p \times \mathbb{R}^p. \tag{A.8}$$

Propriétés géométriques

Continuité stochastique. Une fonction aléatoire $Z(.)$ est dite continue en moyenne quadratique en un point $\mathbf{x}_0 \in \mathbb{R}^p$ si l'on a :

$$\lim_{\|\mathbf{x}-\mathbf{x}_0\|\to 0} \mathbb{E}\left(\left(Z(\mathbf{x}) - Z(\mathbf{x}_0)\right)^2\right) = 0. \tag{A.9}$$

Proposition A.1. *Supposons que $m(.)$ soit continue. Alors, la fonction aléatoire $Z(.)$ est continue en moyenne quadratique en $\mathbf{x}_0 \in \mathbb{R}^p$ si et seulement si sa fonction de covariance $C(\mathbf{x},\mathbf{y})$ est continue en $\mathbf{x} = \mathbf{y} = \mathbf{x}_0$.*

Si $C(.,.)$ est continue sur la diagonale : $\lim_{\mathbf{y}\to\mathbf{x}} C(\mathbf{x},\mathbf{y}) = C(\mathbf{x},\mathbf{x})$, *alors elle est continue partout.*

Dérivation stochastique. Une fonction aléatoire $Z(.)$ est dérivable en moyenne quadratique, par rapport à x_1 au point $\mathbf{x} = (x_1, \ldots, x_p)$ et a en ce point pour dérivée partielle en x_1 la fonction aléatoire $\partial Z/\partial x_1$ si :

$$\lim_{|h_1|\to 0} \mathbb{E}\left\{\left(\frac{Z(x_1+h_1, x_2, \ldots, x_p) - Z(x_1, x_2, \ldots, x_p)}{h_1}\right)^2\right\} \text{ existe et est finie.} \tag{A.10}$$

On définit de la même manière les dérivées en moyenne quadratique par rapport à $x_2, \ldots, x_p$. La dérivée dans une direction quelconque $\boldsymbol{\alpha}$ de cosinus directeur $(\alpha_1, \ldots, \alpha_p)$ est donnée par :

$$\frac{\partial Z}{\partial \boldsymbol{\alpha}} = \alpha_1 \frac{\partial Z}{\partial x_1} + \ldots + \alpha_p \frac{\partial Z}{\partial x_p}. \tag{A.11}$$

Proposition A.2. *Soit une fonction aléatoire $Z(.)$ de fonction de covariance $C(.,.)$ et d'espérance $m(.)$ dérivable. Si la dérivée $\partial^2 C(\mathbf{x}, \mathbf{y})/\partial \mathbf{x}_i \partial \mathbf{y}_i$ existe et est finie pour tout $i = 1, \ldots, p$ au point $(\mathbf{x}_0, \mathbf{x}_0)$, alors $Z(.)$ est dérivable en moyenne quadratique en $\mathbf{x}_0$. La fonction de covariance de $\partial Z/\partial x_i$ est alors donnée par $\partial^2 C(\mathbf{x}, \mathbf{y})/\partial \mathbf{x}_i \partial \mathbf{y}_i$.*

Ainsi, les fonctions aléatoires de second ordre dérivables une fois en moyenne quadratique sont caractérisées par des covariances deux fois dérivables.

Intégration stochastique. Soient deux fonctions aléatoires $Z_1(.)$ et $Z_2(.)$ de moyennes nulles, de variances finies et indépendantes d'ordre 2 ; on note $C_1(.,.)$ et $C_2(.,.)$ leur covariance respective. Soit un volume V, partitionné en n volumes Δv_i. On considère les variables aléatoires I_n de la forme :

$$I_n = \sum_{i=1}^{n} Z_1(\mathbf{x}_i) Z_2(\Delta v_i), \text{ avec } Z_2(\Delta v_i) = \int_{\Delta v_i} Z_2(d\mathbf{x}), \tag{A.12}$$

où les points $\mathbf{x}_i$ sont pris quelconques, à l'intérieur du volume Δv_i.

On appelle intégrale stochastique de Stieltjes $\int_V Z_1(\mathbf{x}) Z_2(d\mathbf{x})$ la limite en moyenne quadratique, si elle existe, des sommes I_n lorsque $n \to +\infty$ et $\max \Delta v_i \to 0$.

Proposition A.3. *Étant données deux fonctions aléatoires $Z_1(.)$ et $Z_2(.)$ de moyennes nulles, de variances finies, indépendantes d'ordre 2, pour que l'intégrale stochastique $I = \int_V Z_1(\mathbf{x}) Z_2(d\mathbf{x})$ existe il faut et il suffit que l'intégrale ordinaire de Stieltjes :*

$$S^2 = \int_V \int_V C_1(\mathbf{x}, \mathbf{y}) C_2(d\mathbf{x}, d\mathbf{y}) \tag{A.13}$$

existe ; dans ce cas on a $\mathbb{E}(I^2) = S^2$.

La proposition A.3 subsiste lorsqu'on fait tendre vers $\mathbb{R}^p$ le volume V. En remplaçant dans la définition précédente les fonctions aléatoires $Z_1(.)$ ou $Z_2(.)$ par une fonction de pondération $p(\mathbf{x})$, on définit les intégrales stochastiques $\int_{\mathbb{R}^p} p(\mathbf{x})Z_2(d\mathbf{x})$ et $\int_{\mathbb{R}^p} Z_1(\mathbf{x})p(d\mathbf{x})$.

Application à la convolution stochastique. On peut également convoluer deux fonctions aléatoires sans corrélation $Z_1(.)$ et $Z_2(.)$; on obtient alors une autre fonction aléatoire $Z_3(.)$ dont la covariance est $C_3(.,.)$:

$$Z_3(\mathbf{x}) = \int_{\mathbb{R}^p} Z_1(\mathbf{x}-\mathbf{u})Z_2(d\mathbf{u}), \tag{A.14}$$

$$C_3(\mathbf{x},\mathbf{y}) = \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} C_1(\mathbf{x}-\mathbf{u},\mathbf{y}-\mathbf{v})C_2(d\mathbf{u},d\mathbf{v}). \tag{A.15}$$

Annexe B

Scores de prédiction

Dans la pratique géostatistique, on est amené à comparer différents modèles et à choisir celui qui fonctionne le mieux. La performance d'un modèle géostatistique est généralement reliée à sa capacité de prédiction. Évaluer cette performance est extrêmement important en pratique car cela guide le choix du modèle. Cela permet aussi d'avoir une mesure de la qualité du modèle finalement choisi. Les modèles sont comparés selon certains critères (scores). Dans cette annexe, nous en présentons brièvement quelques uns. Ces scores sont calculés sur la base de méthodes de type validation croisée (*cross-validation*) ou externe (*external validation*).

Validation croisée. La validation croisée est un outil classique en géostatistique pour évaluer la capacité d'un modèle à retrouver des valeurs connues et pour comparer différents modèles. Le principe de la validation croisée consiste à supprimer à tour de rôle chacune des données disponibles, et à ré-estimer chaque point supprimé à partir des données restantes (par krigeage) sans ré-estimer le modèle, puis à comparer la prédiction obtenue avec la vraie valeur mise de côté [Chilès and Delfiner, 2012]. La pertinence de la validation croisée s'appuie sur trois propriétés de l'estimateur de krigeage : celui-ci est sans biais, de variance minimale et surtout est un interpolateur exact. Ce dernier point justifie le besoin de supprimer une donnée et de la ré-estimer pour évaluer le modèle. En validation croisée, les données sont utilisées à la fois pour estimer le modèle et pour évaluer sa capacité prédictive. De ce fait, les modèles ayant les meilleurs résultats en validation croisée ne sont pas nécessairement ceux fournissant les meilleures prédictions aux points non-observés. La validation croisée telle que présentée est la *Leave One Out Cross Validation*. Cependant, il existe d'autres variantes de la validation croisée telle que la *k-Fold Cross Validation*. Cette dernière consiste à partitionner l'ensemble des observations en k sous-ensembles, puis à enlever une partition et prédire les points de celle-ci à partir de ceux des autres partitions.

Validation externe. Généralement, lorsque la taille des données le permet, l'approche par validation croisée est remplacée par une validation externe. Cette dernière consiste tout simplement à séparer l'échantillon global en deux sous-ensembles : un échantillon d'apprentissage (*training data*) et un échantillon de validation ou test (*validation or test data*). L'échantillon d'apprentissage sert uniquement à estimer le modèle et l'échantillon de validation ou test est utilisé uniquement pour évaluer les capacités prédictives du modèle. Cette approche permet ainsi de corriger le défaut de l'approche par validation croisée.

Scores de prédiction. Plusieurs critères peuvent être utilisés pour comparer l'ensemble des valeurs vraies (données observées) avec l'ensemble des prédictions obtenues en ces points par validation croisée [Chilès and Delfiner, 2012, Zhang and Wang, 2010, Gneiting and Raftery, 2007].

Soient les observations $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ aux n emplacements fixés $\mathbf{s}_1, \ldots, \mathbf{s}_n$. On désigne par $\widehat{Y}_{-i}(\mathbf{s}_i)$ le krigeage à l'emplacement $\mathbf{s}_i$ en utilisant toutes les observations exceptée $Y(\mathbf{s}_i)$ et $\hat{\sigma}^2_{-i}(\mathbf{s}_i)$ la variance de krigeage associée. Nous pouvons calculer :

1. le biais ou la moyenne des erreurs (*Mean Error*) :

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_{-i}(\mathbf{s}_i) - Y(\mathbf{s}_i)\right)$$

2. la moyenne des erreurs absolues (*Mean Absolute Error*) :

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{Y}_{-i}(\mathbf{s}_i) - Y(\mathbf{s}_i)\right|$$

3. la racine carrée de la moyenne des erreurs au carré (*Root Mean Square Error*) :

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_{-i}(\mathbf{s}_i) - Y(\mathbf{s}_i)\right)^2\right]^{\frac{1}{2}}$$

4. la moyenne des carrés des erreurs standardisées (*Normalized Mean Square Error*) :

$$NMSE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{Y}_{-i}(\mathbf{s}_i) - Y(\mathbf{s}_i)}{\hat{\sigma}_{-i}(\mathbf{s}_i)}\right)^2$$

5. le score logarithmique (*Logarithmic Score*) :

$$LogS = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\log\left(2\pi\hat{\sigma}^2_{-i}(\mathbf{s}_i)\right) + \frac{1}{2}\left(\frac{\hat{Y}_{-i}(\mathbf{s}_i) - Y(\mathbf{s}_i)}{\hat{\sigma}_{-i}(\mathbf{s}_i)}\right)^2\right)$$

6. l'indice continu de probabilité ordonnée (*Continued Ranked Probability Score*) :

$$CRPS = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{+\infty}(F_i(y) - \mathbb{1}\{Y(\mathbf{s}_i) \leq y\})^2 dy$$

où $F_i(y)$ est la fonction de répartition prédictive de $Y(\mathbf{s}_i)$: $F_i(y) = P(Y(\mathbf{s}_i) \leq y | Y(\mathbf{s}_j), i \neq j)$.

Plus les critères RMSE, LogS et CRPS sont petits meilleurs ils sont ; plus les critères MAE et ME sont proches de 0, meilleurs ils sont ; pour le NMSE, plus il proche de 1 meilleur il est. Les scores ME, MAE et RMSE mesurent les performances de ré-estimation, mais ne disent rien au sujet de la variance de l'erreur associée. Les scores NMSE et LogS prennent en compte la prédiction en même temps que la variance de l'erreur associée. Les critères ME, MAE, RMSE et NMSE ne dépendent pas de la distribution des observations. Le score LogS est équivalent à la pseudo-vraisemblance dans le cas où Y est une fonction aléatoire gaussienne. Le critère de CRPS correspond à la distance entre la fonction de répartition de la variable prédite et l'observation (elle-même exprimée sous la forme d'une fonction de répartition). Le score CRPS est généralement calculé dans le cadre gaussien où il admet une expression explicite (sans intégrale) [Zhang and Wang, 2010]. Bien que ces les scores LogS et CRPS soient généralement calculés dans le contexte gaussien, ils sont assez robustes.

Annexe C

Démonstrations

Cette annexe contient les preuves des propositions et corollaires énoncés dans les chapitres 2 et 3.

Preuve proposition 2.1

Considérons un ensemble de points d'appui $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset G$ et un ensemble de réels $\{\lambda_1, \ldots, \lambda_n\}$ tel que $\sum_{i=1}^{n} \lambda_i = 0, n \in \mathbb{N}^*$. Posons $\mathbf{x}_i^* = f(\mathbf{x}_i), \forall i = 1, \ldots, n$. Nous avons :

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j\gamma_0 \circ (f \times f)(\mathbf{x}_i, \mathbf{x}_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j\gamma_0(f(\mathbf{x}_i), f(\mathbf{x}_j)) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j\gamma_0(\mathbf{x}_i^\star, \mathbf{x}_j^\star), \text{ avec } \mathbf{x}_1^\star, \ldots, \mathbf{x}_n^\star \in D \text{ et} \sum_{i=1}^{n} \lambda_i = 0 \\
&\leq 0, \text{ car } \gamma_0(.,.) \text{ est un variogammme valide sur } D.
\end{aligned}$$

D'où $\gamma_0 \circ (f \times f)$ est un variogramme valide sur G.

Preuve proposition 2.3

Soit (γ_0, f) une solution de (2.3). Posons $(\tilde{\gamma_0}, \tilde{f})$ tel que $\tilde{f}(\mathbf{x}) = \mathbf{A}f(\mathbf{x}) + \mathbf{b}$ et $\tilde{\gamma}_0(\|\mathbf{u}\|) = \gamma_0(\|\mathbf{A}^{-1}\mathbf{u}\|)$. Nous avons :

$$\begin{aligned}
\tilde{\gamma}_0(\|\tilde{f}(\mathbf{x}) - \tilde{f}(\mathbf{y})\|) &= \tilde{\gamma}_0(\|\mathbf{A}f(\mathbf{x}) + \mathbf{b} - \mathbf{A}f(\mathbf{y}) - \mathbf{b}\|) \\
&= \tilde{\gamma}_0(\|\mathbf{A}(f(\mathbf{x}) - f(\mathbf{y}))\|) \\
&= \gamma_0(\|\mathbf{A}^{-1}\mathbf{A}(f(\mathbf{x}) - f(\mathbf{y}))\|) \\
&= \gamma_0(\|f(\mathbf{x}) - f(\mathbf{y})\|).
\end{aligned}$$

D'où $(\tilde{\gamma_0}, \tilde{f})$ est aussi une solution de (2.3).

Preuve proposition 3.1

Pour tout $\mathbf{x} \in G$, l'espérance de Z s'écrit :

$$\begin{aligned}
\mathbb{E}(Z(\mathbf{x})) &= \mathbb{E}\left(\int_{\mathbb{R}^p} f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))W(d\mathbf{u})\right) \\
&= \int_{\mathbb{R}^p} \mathbb{E}\left(f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))W(d\mathbf{u})\right) \text{ (Fubini)} \\
&= \int_{\mathbb{R}^p} \mathbb{E}\left(f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))\right)\mathbb{E}\left(W(d\mathbf{u})\right), \text{ car } T(\mathbf{u}) \perp W(d\mathbf{u}) \\
&= 0, \text{ car } \quad \mathbb{E}\left(W(d\mathbf{u})\right) = 0 \text{ et } \mathbb{E}\left(f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))\right) \text{ est supposée finie et intégrable sur } \mathbb{R}^p.
\end{aligned}$$

Z étant centrée, pour tout $\mathbf{x}, \mathbf{y} \in G$, sa covariance s'écrit :

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(Z(\mathbf{x}), Z(\mathbf{y})) &= \mathbb{E}(Z(\mathbf{x})Z(\mathbf{y})) \\
&= \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \mathbb{E}(f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))f_{\mathbf{y}}(\mathbf{v}; T(\mathbf{v}))W(d\mathbf{u})W(d\mathbf{v})) \text{ (Fubini)} \\
&= \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \mathbb{E}(f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))f_{\mathbf{y}}(\mathbf{v}; T(\mathbf{v})))\mathbb{E}(W(d\mathbf{u})W(d\mathbf{v})), \text{ car } T(\mathbf{u}) \perp W(d\mathbf{u}) \\
&= \lambda \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \mathbb{E}(f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))f_{\mathbf{y}}(\mathbf{v}; T(\mathbf{v})))\delta_{\mathbf{u}}(d\mathbf{v})d\mathbf{u} \\
&= \lambda \int_{\mathbb{R}^p} \mathbb{E}(f_{\mathbf{x}}(\mathbf{u}; T(\mathbf{u}))f_{\mathbf{y}}(\mathbf{u}; T(\mathbf{u})))d\mathbf{u} \\
&= \lambda \int_{\mathbb{R}^p}\int_{\mathcal{T}} f_{\mathbf{x}}(\mathbf{u}; t)f_{\mathbf{y}}(\mathbf{u}; t)\mu(dt)d\mathbf{u} \\
&= \int_{\mathcal{T}}\int_{\mathbb{R}^p} f_{\mathbf{x}}(\mathbf{u}; t)f_{\mathbf{y}}(\mathbf{u}; t)d\mathbf{u}M(dt) \\
&= \int_{\mathcal{T}} C(\mathbf{x}, \mathbf{y}; t)M(dt).
\end{aligned}$$

L'intégrale $C(\mathbf{x}, \mathbf{y}; t)$ pour tout $t \in \mathcal{T}$ existe et est finie, car le produit de deux fonctions de carré intégrable est intégrable. En outre, si pour $n \in \mathbb{N}^*$, nous choisissons une collection de points arbitraires $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \mathbb{R}^p$ et $\{\alpha_1, \ldots, \alpha_n\} \subset \mathbb{R}$, nous avons :

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j C(\mathbf{x}_i, \mathbf{x}_j; t) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \int_{\mathbb{R}^p} f_{\mathbf{x}_i}(\mathbf{u}; t)f_{\mathbf{x}_j}(\mathbf{u}; t)d\mathbf{u} \\
&= \int_{\mathbb{R}^p} \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j f_{\mathbf{x}_i}(\mathbf{u}; t)f_{\mathbf{x}_j}(\mathbf{u}; t)d\mathbf{u} \\
&= \int_{\mathbb{R}^p} \left(\sum_{i=1}^{n} \alpha_i f_{\mathbf{x}_i}(\mathbf{u}; t)\right)\left(\sum_{j=1}^{n} \alpha_j f_{\mathbf{x}_j}(\mathbf{u}; t)\right) d\mathbf{u} \\
&= \int_{\mathbb{R}^p} \left(\sum_{i=1}^{n} \alpha_i f_{\mathbf{x}_i}(\mathbf{u}; t)\right)^2 d\mathbf{u} \geq 0.
\end{aligned}$$

D'où $C(.,.;t)$ est définie positive sur $\mathbb{R}^p$, pour tout $t \in \mathcal{T}$ et intégrable sur $\mathcal{T}$ par rapport à la mesure positive $M(.)$ pour toute paire $(\mathbf{x},\mathbf{y}) \in \mathbb{R}^p \times \mathbb{R}^p$ (car, $\mathbb{V}(Z(\mathbf{x})) < +\infty$, $\forall \mathbf{x} \in G$). Alors, $C : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$, $(\mathbf{x},\mathbf{y}) \mapsto C(\mathbf{x},\mathbf{y}) = \int_{\mathcal{T}} C(\mathbf{x},\mathbf{y};t)M(dt)$ est une covariance valide sur $\mathbb{R}^p \times \mathbb{R}^p$, d'après Matérn [1986].

La démonstration de la proposition 3.3 repose sur le lemme suivant :

Lemme C.1.

$$\forall (\mathbf{x},\mathbf{y}) \in \mathbb{R}^p \times \mathbb{R}^p,\ \int_{\mathbb{R}^p} k_{\mathbf{x}}(\mathbf{u};t)k_{\mathbf{y}}(\mathbf{u};t)d\mathbf{u} = \pi^{-\frac{p}{2}}t^{-p}\left|\frac{\mathbf{\Sigma_x}+\mathbf{\Sigma_y}}{2}\right|^{-\frac{1}{2}} \exp\left(-\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2}\right).$$

Preuve

Posons : $\tilde{C}(\mathbf{x},\mathbf{y};t) = \int_{\mathbb{R}^p} k_{\mathbf{x}}(\mathbf{u};t)k_{\mathbf{y}}(\mathbf{u};t)d\mathbf{u}$.

Comme $k_{\mathbf{x}}(.;t)$ est la densité de la distribution $\mathcal{N}\left(\mathbf{x}, \frac{t^2\mathbf{\Sigma_x}}{4}\right)$, nous avons :

$$k_{\mathbf{x}}(\mathbf{u};t) = (2\pi)^{-\frac{p}{2}}\left|\frac{t^2\mathbf{\Sigma_x}}{4}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{u}-\mathbf{x})^T\left(\frac{t^2\mathbf{\Sigma_x}}{4}\right)^{-1}(\mathbf{u}-\mathbf{x})\right).$$

Soient $\mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \frac{t^2\mathbf{\Sigma_x}}{4}\right)$ et $\mathbf{Y} \sim \mathcal{N}\left(\mathbf{y}, \frac{t^2\mathbf{\Sigma_y}}{4}\right)$ deux vecteurs aléatoires gaussiens indépendants, de densités respectives $f_{\mathbf{X}}(.)$ et $f_{\mathbf{Y}}(.)$. Ainsi, nous avons : $f_{\mathbf{X}}(\mathbf{u}-\mathbf{x}) = k_{\mathbf{x}}(\mathbf{u};t)$ et $f_{\mathbf{Y}}(\mathbf{u}) = k_{\mathbf{y}}(\mathbf{u};t)$. D'où :

$$\tilde{C}(\mathbf{x},\mathbf{y};t) = \int_{\mathbb{R}^p} f_{\mathbf{X}}(\mathbf{x}-\mathbf{u})f_{\mathbf{Y}}(\mathbf{u})d\mathbf{u} = \int_{\mathbb{R}^p} f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}-\mathbf{u},\mathbf{u})d\mathbf{u}.$$

Considérons le C^1-difféomorphisme $(\mathbf{X},\mathbf{Y}) \to (\mathbf{U},\mathbf{V})$ avec $\mathbf{U} = \mathbf{X}+\mathbf{Y}$ et $\mathbf{V} = \mathbf{Y}$, dont le Jacobien vaut 1. Nous avons :

$$\int_{\mathbb{R}^p} f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}-\mathbf{u},\mathbf{u})d\mathbf{u} = \int_{\mathbb{R}^p} f_{\mathbf{U},\mathbf{V}}((\mathbf{x}-\mathbf{u})+\mathbf{u},\mathbf{u})d\mathbf{u} = \int_{\mathbb{R}^p} f_{\mathbf{U},\mathbf{V}}(\mathbf{x},\mathbf{u})d\mathbf{u} = f_{\mathbf{U}}(\mathbf{x}) = f_{\mathbf{X}+\mathbf{Y}}(\mathbf{x}).$$

Comme $\mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \frac{t^2\mathbf{\Sigma_x}}{4}\right)$ et $\mathbf{Y} \sim \mathcal{N}\left(\mathbf{y}, \frac{t^2\mathbf{\Sigma_y}}{4}\right)$ sont indépendants, alors $\mathbf{X}+\mathbf{Y} \sim \mathcal{N}\left(\mathbf{y}, \frac{t^2\mathbf{\Sigma_x}}{4}+\frac{t^2\mathbf{\Sigma_y}}{4}\right)$. Par conséquent,

$$\begin{aligned}
\tilde{C}(\mathbf{x},\mathbf{y};t) &= (2\pi)^{-\frac{p}{2}}\left|\frac{t^2\mathbf{\Sigma_x}+t^2\mathbf{\Sigma_y}}{4}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{y})^T\left(\frac{t^2\mathbf{\Sigma_x}+t^2\mathbf{\Sigma_y}}{4}\right)^{-1}(\mathbf{x}-\mathbf{y})\right) \\
&= (2\pi)^{-\frac{p}{2}}t^{-p}\left|\frac{\mathbf{\Sigma_x}+\mathbf{\Sigma_y}}{2}\right|^{-\frac{1}{2}}2^{\frac{p}{2}} \exp\left(-(\mathbf{x}-\mathbf{y})^T\left(\frac{\mathbf{\Sigma_x}+\mathbf{\Sigma_y}}{2}\right)^{-1}(\mathbf{x}-\mathbf{y})\frac{1}{t^2}\right) \\
&= \pi^{-\frac{p}{2}}t^{-p}\left|\frac{\mathbf{\Sigma_x}+\mathbf{\Sigma_y}}{2}\right|^{-\frac{1}{2}} \exp\left(-\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2}\right).
\end{aligned}$$

Preuve proposition 3.3

La covariance $C(\mathbf{x},\mathbf{y};t)$ s'écrit :

$$C(\mathbf{x},\mathbf{y};t) = \int_{\mathbb{R}^p} f_{\mathbf{x}}(\mathbf{u};t) f_{\mathbf{y}}(\mathbf{u};t) d\mathbf{u} = g(\mathbf{x};t) g(\mathbf{y};t) \int_{\mathbb{R}^p} k_{\mathbf{x}}(\mathbf{u};t) k_{\mathbf{y}}(\mathbf{u};t) d\mathbf{u}.$$

En appliquant le lemme C.1, nous obtenons :

$$C(\mathbf{x},\mathbf{y};t) = g(\mathbf{x};t) g(\mathbf{y};t) \pi^{-\frac{p}{2}} t^{-p} \left| \frac{\mathbf{\Sigma}_{\mathbf{x}} + \mathbf{\Sigma}_{\mathbf{y}}}{2} \right|^{-\frac{1}{2}} \exp\left(-\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2} \right).$$

D'où le résultat :

$$\begin{aligned} C^{NS}(\mathbf{x},\mathbf{y}) &= \int_0^{+\infty} C(\mathbf{x},\mathbf{y};t) M(dt) \\ &= \pi^{-\frac{p}{2}} \left| \frac{\mathbf{\Sigma}_{\mathbf{x}} + \mathbf{\Sigma}_{\mathbf{y}}}{2} \right|^{-\frac{1}{2}} \int_0^{+\infty} g(\mathbf{x};t) g(\mathbf{y};t) t^{-p} \exp\left(-\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2} \right) M(dt). \end{aligned}$$

Preuve corollaire 3.4

Si $g(\mathbf{x};t) \propto 1$ et $M(dt) = t^p \xi(dt)$, pour $\xi(.)$ une mesure positive finie sur $\mathbb{R}_+$ telle que : $R^S(\tau) = \int_0^{+\infty} \exp\left(-\frac{\tau^2}{t^2} \right) \xi(dt), \tau \geq 0$, alors :

$$\begin{aligned} C^{NS}(\mathbf{x},\mathbf{y}) &\propto \pi^{-\frac{p}{2}} \left| \frac{\Sigma_{\mathbf{x}} + \Sigma_{\mathbf{y}}}{2} \right|^{-\frac{1}{2}} \int_0^{+\infty} \exp\left(-\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2} \right) \xi(dt) \\ &\propto \pi^{-\frac{p}{2}} \left| \frac{\Sigma_{\mathbf{x}} + \Sigma_{\mathbf{y}}}{2} \right|^{-\frac{1}{2}} R^S\left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})} \right). \end{aligned}$$

D'où le résultat :

$$R^{NS}(\mathbf{x},\mathbf{y}) = \frac{C^{NS}(\mathbf{x},\mathbf{y})}{\sqrt{C^{NS}(\mathbf{x},\mathbf{x})}\sqrt{C^{NS}(\mathbf{y},\mathbf{y})}} = \phi_{\mathbf{xy}} R^S\left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})} \right).$$

Posons : $\tilde{C}^{NS}(\mathbf{x},\mathbf{y}) = \int_0^{+\infty} g(\mathbf{x};t) g(\mathbf{y};t) t^{-p} \exp\left(-\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2} \right) M(dt)$.

Preuve corollaire 3.8

Si $g(\mathbf{x};t) \propto t^{\nu(\mathbf{x})}, \nu(\mathbf{x}) > 0, \forall \mathbf{x} \in \mathbb{R}^p$ et $M(dt) = 2t^{p-1}h(t^2)\mathbb{1}_{[0,+\infty[}(t)dt$, avec $h(.)$ la densité d'une loi Gamma $\mathcal{G}a(1, 1/4a^2), a > 0$, alors :

$$\tilde{C}^{NS}(\mathbf{x},\mathbf{y}) = \int_0^{+\infty} \frac{1}{2a^2} t^{\nu(\mathbf{x})+\nu(\mathbf{y})-1} \exp\left(-\left(\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2} + \frac{t^2}{4a^2}\right)\right) dt.$$

En faisant le changement de variable $w = \frac{t^2}{4a^2}$ et en utilisant l'expression sous forme d'intégrale de la fonction de Bessel [Gradshteyn and Ryzhik, 2007, McLeish, 1982], nous obtenons :

$$\begin{aligned}
\tilde{C}^{NS}(\mathbf{x},\mathbf{y}) &= \int_0^{+\infty} (4a^2w)^{(\nu(\mathbf{x})+\nu(\mathbf{y})-2)/2} \exp\left(-w - \frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{4a^2w}\right) dw \\
&= (2a)^{2(\nu(\mathbf{x},\mathbf{y})-1)} \int_0^{+\infty} w^{\nu(\mathbf{x},\mathbf{y})-1} \exp\left(-w - \frac{\left(\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}/a\right)^2}{4w}\right) dw \\
&= 2^{\nu(\mathbf{x},\mathbf{y})-1} a^{2(\nu(\mathbf{x},\mathbf{y})-1)} \left(\frac{\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}}{a}\right)^{\nu(\mathbf{x},\mathbf{y})} K_{\nu(\mathbf{x},\mathbf{y})}\left(\frac{\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}}{a}\right).
\end{aligned}$$

Nous avons :

$$C^{NS}(\mathbf{x},\mathbf{y}) = \pi^{-\frac{p}{2}} \left|\frac{\boldsymbol{\Sigma}_\mathbf{x} + \boldsymbol{\Sigma}_\mathbf{y}}{2}\right|^{-\frac{1}{2}} \tilde{C}^{NS}(\mathbf{x},\mathbf{y}).$$

D'où,

$$R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} \frac{2^{1-\nu(\mathbf{x},\mathbf{y})}}{\sqrt{\Gamma(\nu(\mathbf{x}))\Gamma(\nu(\mathbf{y}))}} \left(\frac{\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}}{a}\right)^{\nu(\mathbf{x},\mathbf{y})} K_{\nu(\mathbf{x},\mathbf{y})}\left(\frac{\sqrt{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}}{a}\right).$$

Preuve corollaire 3.9

Si $g(\mathbf{x};t) \propto t^{-\alpha(\mathbf{x})}, \alpha(\mathbf{x}) > 0, \forall \mathbf{x} \in \mathbb{R}^p$ et $M(dt) = 2t^{p+3}h(t^2)\mathbb{1}_{[0,+\infty[}(t)dt$, avec $h(.)$ la densité d'une loi inverse Gamma $\mathcal{IG}(1,a^2), a > 0$, alors :

$$\tilde{C}^{NS}(\mathbf{x},\mathbf{y}) = \int_0^{+\infty} 2a^2 t^{-\alpha(\mathbf{x})-\alpha(\mathbf{y})-1} \exp\left(-\left(\frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{t^2} + \frac{a^2}{t^2}\right)\right) dt.$$

En faisant le changement de variable $w = \frac{1}{t^2}$, nous obtenons :

$$\begin{aligned}
\tilde{C}^{NS}(\mathbf{x},\mathbf{y}) &= a^2 \int_0^{+\infty} w^{(\alpha(\mathbf{x})+\alpha(\mathbf{y})-2)/2} \exp\left(-\left(a^2 w + Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})w\right)\right) dw \\
&= a^{-2(\alpha(\mathbf{x},\mathbf{y})-1)}\Gamma(\alpha(\mathbf{x},\mathbf{y})) \int_0^{+\infty} \exp\left(-Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})w\right) \frac{a^{2\alpha(\mathbf{x},\mathbf{y})}}{\Gamma(\alpha(\mathbf{x},\mathbf{y}))} w^{\alpha(\mathbf{x},\mathbf{y})-1} \exp\left(a^2 w\right) dw \\
&= a^{-2(\alpha(\mathbf{x},\mathbf{y})-1)}\Gamma(\alpha(\mathbf{x},\mathbf{y}))\mathbb{E}\left(\exp\left(-Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})W\right)\right),\ W \sim \mathcal{G}a(\alpha(\mathbf{x},\mathbf{y}),a^2) \\
&= a^{-2(\alpha(\mathbf{x},\mathbf{y})-1)}\Gamma(\alpha(\mathbf{x},\mathbf{y}))\left(1 + \frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{a^2}\right)^{-\alpha(\mathbf{x},\mathbf{y})}.
\end{aligned}$$

Nous avons :

$$C^{NS}(\mathbf{x},\mathbf{y}) = \pi^{-\frac{p}{2}} \left|\frac{\mathbf{\Sigma_x} + \mathbf{\Sigma_y}}{2}\right|^{-\frac{1}{2}} \tilde{C}^{NS}(\mathbf{x},\mathbf{y}).$$

D'où,

$$R^{NS}(\mathbf{x},\mathbf{y}) = \phi_{\mathbf{xy}} \frac{\Gamma(\alpha(\mathbf{x},\mathbf{y}))}{\sqrt{\Gamma(\alpha(\mathbf{x}))\Gamma(\alpha(\mathbf{y}))}} \left(1 + \frac{Q_{\mathbf{xy}}(\mathbf{x}-\mathbf{y})}{a^2}\right)^{-\alpha(\mathbf{x},\mathbf{y})}.$$

Annexe D

Liste des publications et communications

Les résultats obtenus au cours de cette recherche ont fait l'objet de plusieurs publications et communications dont la liste est donnée ci-dessous :

Articles :

1. F. Fouedjio, N. Desassis and T. Romary (2014). Estimation of Space Deformation Model for Non-stationary Random Functions, *Spatial Statistics*, Soumis.
2. F. Fouedjio, N. Desassis and J. Rivoirard (2014). A Generalized Convolution Model and Estimation for Non-stationary Random Functions, *Journal of the American Statistical Association*, Soumis.

Communications orales :

1. C. Lantuéjoul, N. Desassis and F. Fouedjio (2014). Iterative Simulation of A Gaussian Random Vector : A Propagative Version of the Gibbs Sampler, *Gaussian Random Field Simulation Workshop*, November 2014, Bern, Switzerland.
2. F. Fouedjio (2014). *Modèles de Variogramme et Covariance Non-stationnaire : Inférence et Applications*, Cours : Les Méthodes de la Géostatistique, Octobre 2014, École Nationale Supérieure des Mines de Paris.
3. F. Fouedjio, N. Desassis and T. Romary (2014). Estimation of Space Deformation Model for Non-stationary Random Functions, *21th International Conference on Compuatational Statistics*, August 2014, Geneva, Switzerland.
4. F. Fouedjio, N. Desassis and J. Rivoirard (2014). A Generalized Convolution Model and Estimation for Non-stationary Random Functions, *10th Conference on Geostatistics for Environmental Applications*, July 2014, Paris, France.

5. F. Fouedjio, N. Desassis and T. Romary (2014). Estimation de Fonctions Aléatoires Non-stationnaires de Second Ordre par Déformation Spatiale, *46èmes journées de statistique de la SFDS*, Juin 2014, Rennes, France.
6. C. Lantuéjoul, N. Desassis and F. Fouedjio (2014). Simulation of a Gaussian random vector : A propagative Approach to the Gibbs Sampler, *Kriging and Gaussian Processes for Computer Experiments CHORUS Workshop*, April 2014, Paris, France.
7. F. Fouedjio, N. Desassis and J. Rivoirard (2014). A Generalized Convolution Model and Estimation for Non-stationary Random Functions, *Workshop Rencontres Statistiques au Sommet de Rochebrune*, April 2014, Rochebrune, France.
8. F. Fouedjio, N. Desassis and T. Romary (2014). Estimation of Non-stationary Random Fields through Space Deformation, *11th German Probability and Statistics Days Conference*, March 2014, Ulm, Germany.
9. F. Fouedjio, N. Desassis and T. Romary (2013). Modélisation Géostatistique Globale Non-stationnaire par Déformation d'Espace, *Journées de Géostatistique*, Septembre 2013, Fontainebleau, France.

Posters :

1. F. Fouedjio, N. Desassis and T. Romary (2014). Estimation of Non-stationary Random Fields through Space Deformation, *10th Conference on Geostatistics for Environmental Applications*, July 2014, Paris, France.
2. F. Fouedjio, N. Desassis and T. Romary (2013). Non-stationary Global Geostatistical Modelling through Space Deformation, *Summer School on Topics in Space-Time Modelling and Inference*, May 2013, Aalborg, Denmark.

Contributions à la modélisation et à l'inférence des fonctions aléatoires non-stationnaires de second ordre

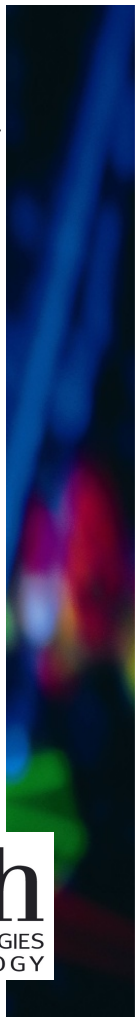
Résumé : Les fonctions aléatoires stationnaires ont été utilisées avec succès dans les applications géostatistiques depuis plusieurs décennies. La structure de dépendance spatiale sous-jacente de la fonction aléatoire est alors représentée par un variogramme ou une covariance stationnaire. Cependant, dans certaines situations, il y a très peu de raisons de s'attendre à une structure de dépendance spatiale stationnaire sur l'ensemble du domaine d'intérêt. Dans cette thèse, deux approches de modélisation non-stationnaire de fonctions aléatoires sont considérées : déformation d'espace et convolution stochastique. Pour chacune d'elle, nous développons une méthodologie statistique d'estimation de la structure de dépendance spatiale non-stationnaire, dans le contexte d'une réalisation unique. Par ailleurs, nous montrons également comment dans ce cadre non-stationnaire, les prédictions spatiales et les simulations conditionnelles peuvent être menées. Les méthodes d'inférence développées permettent de capturer des structures de dépendance variables tout en garantissant la cohérence globale du modèle final. L'évaluation de leur performance selon plusieurs critères, sur des données synthétiques et réelles montre qu'elles donnent de meilleurs résultats de prédiction qu'une méthode stationnaire. Au delà de la prédiction, elles peuvent également servir comme outil pour une analyse exploratoire de la non-stationnarité.

Mots-clefs : non-stationnarité, variogramme, covariance, déformation, convolution, noyau, krigeage, simulation.

Contributions to Modelling and Inference of Second Order Non-stationary Random Functions

Abstract: Stationary Random Functions have been sucessfully applied in geostatistical applications for decades. The underlying spatial dependence structure of the Random Function is represented by a stationary variogram or covariance. However, in some instances, there is little reason to expect the spatial dependence structure to be stationary over the whole region of interest. In this manuscript, two non-stationary modelling approaches for Random Functions are considered: space deformation and stochastic convolution. For each of them, we develop a statistical methodology for estimating the non-stationary spatial dependence structure, in the context of a single realization. Moreover, we also show how spatial predictions and conditional simulations can be carried out in this non-stationary framework. The developed inference methods allow to capture varying spatial structures while guaranteeing the global consistency of the final model. The assessment of their performance on both synthetic and real datasets show that they outperform stationary method, according to several criteria. Beyond the prediction, they can also serve as a tool for exploratory analysis of the non-stationarity.

Keywords: non-stationarity, variogram, covariance, deformation, convolution, kernel, kriging, simulation.



MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY