ESTIMATION DE FONCTIONS ALÉATOIRES NON-STATIONNAIRES DE SECOND ORDRE PAR DÉFORMATION SPATIALE

Francky Fouedjio [1], Nicolas Desassis [2], Thomas Romary [3] & Jacques Rivoirard [4]

[1, 2, 3, 4,] *Centre de Géosciences/Géostatistique, MINES ParisTech, 35 rue Saint-Honoré, 77305 Fontainebleau, France*

[1] *francky.fouedjio@mines-paristech.fr,* [2] *nicolas.desassis@mines-paristech.fr,* [3] *thomas.romary@mines-paristech.fr,* [4] *jacques.rivoirard@mines-paristech.fr,*

Résumé. Les fonctions aléatoires stationnaires ont été utilisées avec succès dans des applications géostatistiques depuis plusieurs décennies. Cependant, dans certaines situations, l'hypothèse d'une structure de dépendance spatiale homogène sur l'ensemble du domaine d'intérêt se révèle inappropriée. Une approche ingénieuse pour la modélisation de la structure de dépendance spatiale non-stationnaire est la déformation de l'espace. Jusqu'à présent, cette approche n'a pu être fonctionnelle que dans un contexte de données issues de plusieurs réalisations indépendantes d'une fonction aléatoire et présente un certain nombre de manquements. Dans ce travail, nous proposons une approche pour la modélisation géostatistique non-stationnaire utilisant la déformation de l'espace dans le contexte d'une réalisation unique avec des données éventuellement espacés irrégulièrement. La procédure d'estimation combine les outils de lissage par noyaux, positionnement multidimensionnel, fonctions de base radiales, pour transformer la fonction aléatoire initialement non-stationnaire vers un nouvel espace déformé où elle est stationnaire et isotrope. Les techniques classiques pour la prédiction et la simulation peuvent être appliqués dans l'espace déformé. Les résultats prédits et simulés sont ensuite transposés vers l'espace d'origine. Sur un jeu de données synthétiques, la méthode est capable de retrouver la vraie déformation. Un schéma de comparaison du krigeage ordinaire sous les hypothèses stationnaire et non-stationnaire démontre que l'approche proposée a de meilleures performances de prédiction à la fois sur un jeu de données synthétiques et sur un jeu de données de sol. Enfin, nous montrons également que la méthode peut être considérée comme un outil de visualisation de la non-stationnarité.

Mots-clés. non-stationnarité, variogramme global, krigeage, dissimilarité, déformation spatiale, positionnement multidimensionnel, lissage par noyaux, fonctions de base radiales.

Abstract. Second order stationary random functions have been successfully applied in geostatistical applications for decades. However, in some situations, the assumption of an homogeneous spatial dependence structure across the entire domain of interest is unrealistic. An ingenious approach for modeling non-stationary spatial dependence structure is space deformation. So far, this approach has been really functional in the context of data from several independent realizations of a random function and has a number of shortcomings. In this work, we propose an approach for non-stationary geostatistical modeling using space deformation in the context of a single realization with possibly irregularly spaced data. Our estimation procedure combines aspects of kernel smoothing, multidimensional scaling and radial basis functions to transform the originally non-stationary random function towards a new deformed space where it is stationary and isotropic. Standard techniques for prediction and simulation can be applied in the deformed space. The predicted and simulated results are then mapped back into the original space. In a simulation study, the method is able to find the true deformation. A comparison scheme of ordinary kriging under stationary and non-stationary assumptions demonstrates that the proposed approach has better prediction performance on both a simulation case and soil dataset. Finally, we also show that the method can be seen as a vizualisation tool of the non-stationarity.

Keywords. non-stationarity, global variogram, kriging, dissimilarity, space deformation, multi-dimensional scaling, kernel smoothing, radial basis functions.

1 Introduction

Dans l'étude statistique des phénomènes spatiaux, la modélisation de la structure de dépendance spatiale est fondamentale. Elle est exploitée par les techniques classiques de prédiction à l'instar du krigeage et les méthodes d'estimation comme les simulations conditionnelles. La description de celle-ci se fait couramment à l'aide d'outils statistiques, tels que le covariogramme ou le variogramme, calculés sur l'ensemble du champ considéré et sous une hypothèse de stationnarité, pour des raisons soit de parcimonie ou de commodité mathématique. De la sorte, on limite la complexité de la composante spatiale du phénomène analysé. Cette hypothèse que la structure de dépendance spatiale est invariante par translation sur l'ensemble du domaine d'intérêt peut être appropriée, lorsque ce dernier est de taille faible ou 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'autre alternative raisonnable. Bien que justifiée et conduisant à une analyse raisonnable, l'hypothèse de stationarité n'est souvent pas plausible au vu de certaines données spatiales collectées dans la pratique. La non-stationnarité peut survenir en raison de nombreux facteurs dont les caractéristiques topographiques de la région d'intérêt ou d'autres effets localisés. Ces influences locales, se traduisent par le fait que les données peuvent obéir à un variogramme ou covariogramme dont certaines caractéristiques varient à travers le champ, rendant l'inférence difficile. Pour ce type de structures spatialement non-stationnaires, si l'on veut faire des prédictions spatiales, il n'est pas approprié d'utiliser les méthodes stationnaires classiques. En effet, l'application des modèles stationnaires à des données spatiales vraisemblablement non-stationnaires serait susceptible de produire des estimations ou des prédictions moins précises, y compris une évaluation incorrecte de l'erreur d'estimation.

L'idée principale sous-jacente à la méthode proposée est celle de la déformation spatiale du support d'une fonction aléatoire non-stationnaire. Cette idée originale due à Sampson and Guttorp [7] consiste à modéliser la non-stationnarité en réduisant la structure de dépendance non-stationnaire en une structure de dépendance stationnaire et isotrope, grâce à une anamorphose, c'est-à-dire une transformation de l'espace géographique. Une limite majeure de l'approche actuelle est qu'elle nécessite une réplication des observations. Or, dans de nombreuses applications géostatistiques, les observations sont issues d'une réalisation unique de fonctions aléatoires. La méthode proposée ici, s'affranchit de cette hypothèse forte de réplication. D'autres lacunes associées à cette approche sont : la non garantie de la condition de bijection de la fonction déformation ; l'ajustement du modèle devient un challenge computationnel pour de grands ensembles de données. Pour prendre en compte ces manquements, nous proposons un modèle de déformation pour une structure de dépendance spatiale hétérogène basée d'une part, sur l'inclusion de contraintes spatiales, et d'autre part, sur la transformation non pas de l'ensemble des points de données, mais plutôt d'un sous ensemble de points de référence. L'approche proposée comprend l'estimation de la déformation spatiale au travers d'une combinaison d'outils suivants : lissage par noyaux, positionnement multidimensionnel et fonctions de bases radiales ; ainsi qu'une estimation rationnelle et automatique du modèle de variogramme de la fonction aléatoire dans l'espace déformé.

Le présent document est organisé comme suit : la section 2 décrit le modèle de déformation spatiale. La méthodologie d'estimation du modèle est présentée en section 3.

2 Modèle de déformation spatiale

Les ingrédients du modèle déformation pour une structure de dépendance spatiale hétérogène sont les suivants :

- soit une fonction aléatoire non-stationnaire $Z = \{Z(\mathbf{s}), \mathbf{s} \in \mathcal{G} \subset \mathbb{R}^d, d \geq 2\}$ définie sur un domaine d'intérêt $\mathcal{G}$ fixe et continu de l'espace euclidien $\mathbb{R}^d$, avec des moments au moins jusqu'à l'ordre 2 et rendant compte du phénomène sous-jacent étudié ;
- soit $Y = \{Y(\mathbf{u}), \mathbf{u} \in \mathcal{D} \subset \mathbb{R}^{d'}\}$ une fonction aléatoire indicée par $\mathbb{R}^{d'}$, stationnaire d'ordre 2 et isotrope, de variogramme γ_0 ;
- soit f une transformation ou déformation, déterministe, bijective et bicontinue de $\mathcal{G}$(espace géographique) sur $\mathcal{D}$(espace déformé).

On considère que la fonction aléatoire Z est gouverné par le modèle suivant :

$$\{Z(\mathbf{s}), \mathbf{s} \in \mathcal{G}\} = \{Y(f(\mathbf{s})), \mathbf{s} \in \mathcal{G}\} \tag{1}$$

qui peut s'écrire de façon équivalente :

$$\{Z(f^{-1}(\mathbf{u})), \mathbf{u} \in \mathcal{D}\} = \{Y(\mathbf{u}), \mathbf{u} \in \mathcal{D}\} \tag{2}$$

La spécification du modèle en (1) conduit alors, à modéliser le variogramme de la fonction aléatoire non-stationnaire Z de la façon suivante :

$$\gamma(\mathbf{s}, \mathbf{s}') = \frac{1}{2} Var[Z(\mathbf{s}) - Z(\mathbf{s}')] = \gamma_0(\parallel f(\mathbf{s}) - f(\mathbf{s}') \parallel), \quad \forall \quad \mathbf{s}, \mathbf{s}' \in \mathcal{G}, \tag{3}$$

où $\parallel . \parallel$ représente la norme euclidienne classique dans $\mathbb{R}^{d'}$.

Le modèle de variogramme obtenue en (3) conduit à variogramme valide, c'est-à-dire conditionnellement défini négatif. Sa validité repose sur la proposition suivante :

Proposition 1. *Si $\gamma : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^+$ est un variogramme, alors $\gamma \circ (f \times f)$ est un variogramme sur $\mathcal{D} \times \mathcal{D}$, pour toute fonction $f : \mathcal{G} \to \mathcal{D}$.*

Ainsi, il est possible de s'appuyer sur un variogramme dans un autre espace, par l'intermédiaire d'une fonction qui lie les deux espaces. Concrètement, la déformation spatiale f opère comme suit : dans les régions de corrélation spatiale relativement faible, la transformation étire l'espace $\mathcal{G}$, tandis que dans les régions de corrélation spatiale relativement élevée, elle le contracte, de sorte qu'un variogramme isotrope peut être modélisé dans l'espace $\mathcal{D}$. Ainsi, au lieu de travailler dans l'espace d'origine non-stationnaire $\mathcal{G}$, le variogramme de Z est défini par rapport à l'espace latent $\mathcal{D}$, où la stationnarité et l'isotropie sont supposées. Le variogramme est modélisé comme une fonction dépendant de la différence de positions entre les points dans l'espace déformé par f. Tout le problème faisant intervenir le champ observé Z est donc transposé, par la déformation f, à Y champ aléatoire stationnaire et isotrope pour lequel les méthodes usuelles, comme le krigeage et les simulations conditionnelles s'appliquent directement. Les résultats obtenus sur Y se transposent ensuite à Z, par la déformation inverse f^{-1}. Par la suite, les termes $\mathcal{G}$-espace ou $\mathcal{G}$-plan (en dimension 2) feront référence au système de coordonnées géographiques, de même que $\mathcal{D}$-espace ou $\mathcal{D}$-plan à leur représentation déformée.

Il est important de noter que le modèle de déformation spatiale défini en (3) n'est identifiable qu' à une homothétie près pour γ_0 et à une transformation isométrique près (translation, rotation, symmétrie) pour f (Perrin and Meiring [5], Perrin and Senoussi [6]). Toutes les transformations des coordonnées dans le $\mathcal{D}$-espace qui conserve les distances entre les points sont équivalentes. La classe des modèles de déformation spatiale englobe celle des modèles stationnaires isotropes et celle des modèles d'anisotropie géométrique. En prenant, $f = Id_{\mathcal{G}}$ (fonction identité), dans le modèle en (3), on obtient le modèle stationnaire isotrope. De même, pour $f(\mathbf{s}) = \mathbf{A}\mathbf{s}, \forall\ \mathbf{s} \in \mathcal{G}$ (transformation linéaire), on obtient le modèle d'anisotropie géométrique.

Dans le modèle défini en (3), les fonctions γ_0 et f sont inconnues et donc à estimer. Nous nous plaçons dans le cadre suivant :

1. on dispose d'une seule réalisation $\mathbf{Z} = [Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)]'$ le $n \times 1$ vecteur des observations (denses) associées aux emplacements connus (éventuellement irréguliers) $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathcal{G} \subset \mathbb{R}^d$ du champ aléatoire parent Z ;
2. on suppose que le champ aléatoire non-stationnaire Z est de moyenne et de variance constantes, c'est-à-dire que la seule source de non-stationnarité est l'anisotropie spatialement variable ;
3. on autorise en principe que $d \neq d'$ bien que dans la plupart des articles traitant ce modèle on ait $d = d'$. Les espaces $\mathcal{G}$ et $\mathcal{D}$ sont considérés de dimension 2 qui correspond à l'espace physique facilement perçu, bien que le $\mathcal{D}$-espace puisse être de dimension $d' \geq 2$.

3 Méthodologie d'estimation

On considère le problème d'estimation de la structure de dépendance spatiale non-stationnaire définie à l'équation (3) et celui de la prédiction spatiale, dans un contexte de réalisation unique, où les hypothèses habituelles de stationnarité et d'isotropie ne sont plus directement applicables. L'approche méthodologique proposée comprend cinq grandes étapes qui se déclinent de la manière suivante :

1. définir un estimateur non-paramétrique du variogramme (non-stationnaire) qui servira de mesure de dissimilarité entre les points de l'espace d'origine $\mathcal{G}$;
2. déterminer une représentation des points de l'espace géographique dans le pseudo espace $\mathcal{D}$, au moyen de la procédure de positionnement multidimensionnel appliquée à une matrice de dissimilarités définies entre ces points ;
3. estimer la déformation spatiale f dans la classe de transformations non-paramétriques de type splines plaque mince ;
4. estimer le modèle de variogramme γ_0 dans le nouvel espace (virtuel) où la stationnarité et l'isotropie sont supposées, au travers d'un mélange de modèles de base, offrant une plus grande flexibilité pour adapter la structure observée ;
5. prédire ou faire des simulations conditionnelles sur une grille cible.

Bibliographie

[1] I. Borg and P.J.F. Groenen. *Modern Multidimensional Scaling : Theory and Applications.* Springer Series in Statistics. Springer, 2005.

[2] J. P. Chilès and P. Delfiner. *Geostatistics : modeling spatial uncertainty.* Wiley, 1999.

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

[4] Peter Guttorp and Paul D. Sampson. *Methods for estimating heterogeneous spatial covariance functions with environmental applications*, volume Volume 12, pages 661–689. Elsevier, 1994.

[5] O. Perrin and W. Meiring. Identifiability for non-stationary spatial structure. *Journal of Applied Probability*, 36(4) :1244–1250, 1999.

[6] O. Perrin and R. Senoussi. Reducing non-stationary random fields to stationarity and isotropy using a space deformation. *Statistics & Probability Letters*, 48(1) :23–32, 2000.

[7] P. D. Sampson and P. Guttorp. Nonparametric-estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417) :108–119, 1992.

[8] M.P. Wand and C. Jones. *Kernel smoothing*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1995.