1

Statistique Spatiale pour l'environnement

Session organisée par Liliane Bel

Application de la statistique spatiale à divers problèmes environnementaux
Adresse de l'organisateur :

Liliane Bel
Laboratoire Mathématique
Université Paris Sud
91405 Orsay Cedex France
E-mail : Liliane.Bel@math.u-psud.fr
<http://www.math.u-psud.fr/~bel/>

Dépendance des extrêmes bivariés : quelles applications au cadre spatial ?

par **Jean-Noel Bacro**

Les données spatialisées sont aujourd'hui très répandues et la modélisation à l'aide de processus spatiaux est utilisée dans de nombreux domaines de recherche (environnement, écologie, climatologie, ...). La caractérisation du comportement des extrêmes de tels processus est le plus souvent d'un intérêt fondamental et les approches usuelles qu'offre la statistique spatiale sont alors inadaptées. Récemment, différents travaux ont permis d'étendre au cas spatial des résultats de la théorie des extrêmes multivariée (Schlather et Tawn (2003), Ancona-Navarrete et Tawn (2002)) : fonction coefficient extrêmal à distance h, $\theta(h)$, $1 \leq \theta(h) \leq 2$, inspirée de la notion de coefficient extrêmal θ, et fonction coefficient de dépendance de queue à distance h, $\eta(h)$, inspirée de la modélisation proposée par Ledford et Tawn (1996) pour décrire la dépendance des extrêmes multivariés. Les notions de dépendance ou d'indépendance asymptotique des extrêmes sont essentielles et jouent un rôle de premier plan dans les problématiques de modélisation. Deux estimateurs de la fonction $\theta(\cdot)$ ont été proposés dans la littérature : Schlather et Tawn (2003), dans un cadre maximum de vraisemblance sur répétitions indépendantes d'observations à distance h ; Poncet, Cooley et Naveau (2005), dans une approche fondée sur une caractérisation spatiale du processus d'intérêt. Nous proposons une généralisation de ce dernier estimateur et comparons sur la base de simulations les résultats obtenus par ces trois estimateurs. Les cas de modèles spatiaux dépendants et asymptotiquement indépendants sont considérés. Le même cadre de simulations permet par ailleurs de comparer les performances des quatre estimateurs usuels de la fonction de dépendance de queue $\eta(h)$ (Ledford-Tawn (1996), Hill (1975), Peng (1999) et Draisma *et al.* (2001)). Les performances de l'ensemble des estimateurs, en particulier à faible distance h, sont alors discutées.

Adresse :

Jean-Noel Bacro
I3M, UMR CNRS 5149
Universite Montpellier II, CC 51
4 place Eugene Bataillon 34095 Montpellier cedex 5
E-mail : bacro@math.univ-montp2.fr
<>

Spécifier la distribution d'un processus spatial latent. Application en épidémiologie végétale.

par **Samuel Soubeyrand**

Dans certaines études environnementales, des variables d'intérêt observées dans l'espace dépendent d'une structure spatiale sous-jacente qui n'est pas expliquée par les covariables observées. Dans ce type de situations, des modèles hiérarchiques peuvent être construits pour analyser les données. Au premier niveau de ces modèles, la variable d'intérêt est modélisée conditionnellement à un processus aléatoire spatial latent (ou caché). Au second niveau, c'est le processus latent qui est modélisé.

Si le processus latent est mal spécifié, c'est-à-dire mal modélisé, alors les conclusions de l'analyse risquent d'être faussées. Nous proposons donc une méthode, basée sur des résidus, aidant à la spécification du processus latent lorsque ce dernier est formé d'effets aléatoires partagés ou bien d'effets aléatoires spatialement corrélés. Cette méthode est utilisée pour construire des modèles décrivant, à différentes échelles spatiales, la propagation depuis une source ponctuelle de maladies des végétaux.

Adresse :

Samuel SOUBEYRAND
Unité de biométrie, INRA
Domaine St Paul, Site Agroparc
84914 Avignon Cedex 9,France
E-mail : `soubeyra@avignon.inra.fr`
<>

Estimation de l'intensité d'un processus ponctuel spatial bruité.

par **Lionel Cucala**

Lors de l'étude d'un semis de points, une étape importante est l'estimation de la fonction d'intensité du processus ponctuel sous-jacent, qui traduit la variation à grande échelle du nombre d'évènements dans le domaine étudié. Cette estimation s'effectue généralement par des techniques non-paramétriques de type noyau. Malheureusement, la collecte des données sur le terrain engendre des erreurs de localisation des évènements. Dans cette étude, nous prenons en compte la distribution de ces erreurs de localisation, supposée connue, en modifiant les estimateurs à noyau classiques. Nous utilisons des techniques de déconvolution élaborées dans le cadre de l'estimation de densité de données bruitées. Une analyse asymptotique de ces estimateurs est menée, ainsi qu'une étude de simulation et une application à des données réelles.

Adresse :

Lionel CUCALA
Laboratoire de Statistique et Probabilités
bâtiment 1R1 Université Paul Sabatier 31062 Toulouse Cedex 9
E-mail : `cucala@cict.fr`
<>

Echelles géographiques non emboitées dans les modèles de régression de Poisson

par **Lea Fortunato**

Les études écologiques utilisent des données (indicateurs de santé et expositions) mesurées sur des unités géographiques. Ces données peuvent être agrégées sur différentes échelles non emboitées. La partition du domaine d'intérêt sur laquelle les données d'exposition sont disponibles est la partition "source", la partition sur laquelle les données d'exposition doivent être reconstruites est la partition "cible". Nous supposons que les indicateurs de santé sont disponibles sur la partition "cible". Nous proposons de comparer deux modèles hiérarchiques Bayésiens : le premier est basé sur la relation entre les mesures latentes de l'exposition sur la partition "cible" et celles observées sur la partition "source", tandis que le deuxième modèle prend en compte la structure de risque sur la partition "cible", conditionnellement aux mesures observées de l'exposition et la relation individuelle exposition-effet. Ces modèles sont développés dans le cadre de la régression de Poisson. Leurs performances sont étudiées à partir de simulations.

Adresse :

Lea FORTUNATO
INSERM-U754,
Epidémiologie environnementale des cancers, 16 avenue Paul Vaillant-Couturier
94807 Villejuif Cedex France
E-mail : `fortunato@vjf.inserm.fr`
<`http://ifr69.vjf.inserm.fr/~webu170/index.htm`>

Comportement d'un estimation du coefficient de corrélation pour des données spatiales non alignées

par **Nicolas Desassis**

L'utilisation accrue des Systèmes d'Information Géographiques (SIG) rend disponible de grandes bases de données spatiales que l'on peut vouloir mettre en relation. Provenant souvent de campagnes de mesures différentes, les variables à mettre en relation sont généralement observées en des sites différents mais on aimerait utiliser leurs corrélations spatiales pour pouvoir quantifier leur lien. Si le coefficient de corrélation linéaire est utilisé, une possibilité est de modéliser les deux variables comme la réalisation d'un champ aléatoire bivarié gaussien incluant un modèle de covariance spatiale. Lorsque les données sont observées aux mêmes sites, le coefficient de corrélation et les paramètres de la fonction de covariance peuvent être estimés simultanément par maximisation numérique de la vraisemblance. Pour des sites d'observation des deux variables non-alignés, nous considérons ce non-alignement comme un problème aux données manquantes et nous utilisons un algorithme EM amélioré par un accélérateur quasi-Newton pour maximiser la vraisemblance. Cette alternative à l'algorithme MCMC utilisé en contexte bayésien permet une estimation rapide des paramètres. Une étude sur simulations a été conduite nous permettant de décrire le comportement de l'estimateur du coefficient de corrélation en fonction des caractéristiques géométriques de l'échantillonnage. Ainsi, nous avons observé que le non-alignement des sites d'observation pouvait dans une certaine mesure être compensé par de fortes corrélations spatiales des variables. Ce résultat est néanmoins conditionnel à une bonne estimation de la portée, estimation qui dépend elle-même de la vraie valeur du coefficient de corrélation.

Adresse :

Nicolas Desassis
Unité de biométrie, INRA
Domaine St Paul, Site Agroparc
84914 Avignon Cedex 9,France
E-mail : nicolas.desassis@avignon.inra.fr
<>