



HAL
open science

Nouveaux outils inférentiels pour processus max-stables

Thomas Opitz, Jean-Noel Bacro, Pierre Ribereau

► **To cite this version:**

Thomas Opitz, Jean-Noel Bacro, Pierre Ribereau. Nouveaux outils inférentiels pour processus max-stables. 43èmes Journées de Statistique, 2011, Tunis, Tunisie. hal-00818137

HAL Id: hal-00818137

<https://hal.science/hal-00818137>

Submitted on 26 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NOUVEAUX OUTILS INFÉRENTIELS POUR PROCESSUS MAX-STABLES

Thomas Opitz & Jean-Noël Bacro & Pierre Ribereau

Institut de Mathématique et de Modélisation (UMR 5149)

Université Montpellier 2, Case Courrier 051, Place Eugène Bataillon, F-34095 Montpellier

Résumé. La modélisation du comportement extrême de phénomènes aléatoires spatiaux est indispensable dans l'évaluation et la gestion des risques environnementaux. L'environnement, la climatologie sont des champs d'application privilégiés (températures, précipitations, neige, marée, vent ...). Les champs max-stables spatiaux, en tant que limites de suites normalisées de processus spatiaux, sont des candidats naturels pour modéliser la dépendance spatiale entre événements extrêmes. Nous présentons d'abord la théorie des valeurs extrêmes univariée et multivariée, théorie qui est à la fois fondement et outil pour les champs max-stables. Après avoir défini les processus max-stables et rappelé quelques-unes de leurs propriétés d'intérêt, nous faisons le point sur les principales méthodes inférentielles actuelles. Des outils graphiques originaux et une approche d'inférence innovante pour caractériser la structure de dépendance extrême entre paires de sites sont alors proposés.

Mots-clés : théorie des valeurs extrêmes ; processus max-stables ; modélisation spatiale

Abstract. Modelling the extreme behaviour of random spatial phenomena is indispensable for the management of environmental risks. Primary fields of application are environment and climatology (temperatures, precipitation, snowfall, storm tides/surges, wind ...). Max-stable fields, which arise as limits of normalised sequences of spatial processes, are obvious candidates for the modelisation of spatial dependence between extreme events. First we present univariate and multivariate extreme value theory which is at the same time foundation and tool for max-stable processes. Then we define max-stable processes and recapitulate some of their properties of interest, together with the current principal inferential methods. Finally we propose novel graphical tools and a new inference approach to characterize the extremal dependence structure between site pairs.

Keywords : Extreme Value Theory ; max-stable processes ; spatial modelisation

1 Introduction

Les applications statistiques de la théorie des valeurs extrêmes supposent souvent une prédiction au-delà des plus grandes observations, c'est-à-dire une extrapolation. Les résultats fondamentaux de la théorie concernent donc, de façon naturelle, le cadre asymptotique. Une certaine régularité dans la queue des distributions est nécessaire pour obtenir asymptotiquement une loi limite non-dégénérée. En pratique, on fait l'hypothèse que cette régularité est présente dans les observations, ou en d'autres termes que l'on est assez loin dans la queue pour avoir une bonne approximation du comportement asymptotique.

Pour plus de détails sur les rappels qui suivent, nous renvoyons à la monographie de S. Resnick [4] qui couvre surtout les aspects théoriques.

2 Extrêmes univariés

2.1 Maxima

Soient X, X_1, \dots, X_n variables aléatoires iid, $X \sim F$. On suppose l'existence de suites normalisantes $a_n > 0, b_n$ telles que

$$\frac{\max_{i=1, \dots, n} X_i - b_n}{a_n} \xrightarrow{d} Z \quad (n \rightarrow \infty), \quad (1)$$

où Z est une variable aléatoire non-dégénérée de loi G . On dit alors que F appartient au domaine d'attraction de G , noté $F \in \mathcal{D}(G)$. La loi G de Z est nécessairement max-stable, c'est-à-dire qu'il existe des fonctions $\alpha(t), \beta(t)$ telles que $G^t(\alpha(t)x + \beta(t)) = G(x)$ pour $t > 0$. Les lois max-stables univariées peuvent être paramétrisées de la façon suivante :

$$G(x) = G_{\xi, \mu, \sigma}(x) = \exp \left(- \left(1 + \xi \frac{x - \mu}{\sigma} \right)_+^{-\frac{1}{\xi}} \right).$$

Cette écriture, dites *loi généralisée des extrêmes*, comporte trois paramètres : un paramètre de forme ξ , caractéristique de la queue de distribution, un paramètres de localisation μ et un d'échelle $\sigma > 0$. La loi Fréchet unité $G(x) = \exp(-\frac{1}{x})$ est obtenue en posant $\xi = \mu = \sigma = 1$.

Dans ce qui suit, nous utilisons la notation $\tilde{X} = \frac{X - b_n}{a_n}$.

2.2 Dépassement de seuils

Soit $F \in \mathcal{D}(G)$. Lorsque $n \rightarrow \infty$, les dépassement de seuils $Y = \tilde{X} - u \mid \tilde{X} > u$ convergent en loi vers une loi de Paréto généralisée (GPD), de fonction de répartition

$$F_{\text{GPD}}(y) = 1 - \left(1 + \frac{\xi y}{\sigma_u} \right)_+^{-\frac{1}{\xi}},$$

où $\sigma_u = \sigma + \xi u$.

2.3 Processus de Poisson

On peut (facilement) montrer que les deux représentations précédentes du comportement asymptotique sont équivalentes à

$$nP(\tilde{X} \in \cdot) \Rightarrow \mu(\cdot) \quad (n \rightarrow \infty), \quad (2)$$

avec $\mu(\cdot \mid z, \infty) = \left(1 + \xi \frac{z - \mu}{\sigma_u} \right)_+^{-\frac{1}{\xi}}$.

En définissant $N_n = \left\{ \tilde{X}_i \mid i = 1, 2, \dots, n \right\}$, (2) nous amène à la convergence faible $N_n \Rightarrow N$ ($n \rightarrow \infty$) avec un processus de Poisson N de mesure d'intensité μ .

3 Extrêmes multivariés

Soient $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \dots$ des vecteurs aléatoires iid t.q. $\mathbf{X} = (X^{(1)}, \dots, X^{(d)}) \sim F(F_1, \dots, F_d)$, avec les fonctions de répartitions marginales F_i de \mathbf{X} .

La théorie des extrêmes univariés s'applique aux marginales univariées et les notions présentées précédemment se généralisent au cadre multivarié.

\triangle Les notations suivantes sont vectorielles, et les opérations entre vecteurs se font par composante des vecteurs.

Soient $\mathbf{a}_n, \mathbf{b}_n$ vecteurs normalisants tels que

$$\frac{\max_{i=1, \dots, n} \mathbf{X}_i - \mathbf{b}_n}{\mathbf{a}_n} \xrightarrow{d} \mathbf{Z} \sim G(G_1, \dots, G_d) .$$

De nouveau, G est max-stable : $G^t(\boldsymbol{\alpha}(t)\mathbf{x} + \boldsymbol{\beta}(t)) = G(\mathbf{x})$ pour tout $t > 0$ et des vecteurs $\boldsymbol{\alpha}(t), \boldsymbol{\beta}(t)$. De plus, les marginales G_i sont des lois extrêmes univariées.

G permet la représentation

$$G(x) = \exp(-\mu([-\infty, \mathbf{x}]^c)) .$$

avec la *mesure exponent* μ , une mesure de Radon soumise à certaines contraintes.

Soit le processus ponctuel

$$N_n = \left\{ \widetilde{\mathbf{X}}_i \mid i = 1, \dots, n \right\} ,$$

alors

$$N_n \Rightarrow N ,$$

où N est un processus de Poisson de mesure d'intensité μ .

Sans perte de généralité on peut toujours se ramener à des marginales Fréchet unité (notées G_i^*). La mesure exponent $\mu = \mu^*$, définie sur $[0, \infty]^d$, vérifie alors une propriété d'homogénéité d'ordre -1 : $\mu^*(tB) = t^{-1}\mu^*(B)$ pour tout B ensemble de Borel. Cette propriété permet une factorisation de μ^* à l'aide de la *mesure spectrale*.

Pour ce faire, choisissons deux normes $\|\cdot\|_a, \|\cdot\|_r$ afin de définir la transformation vers des *coordonnées angulaires*

$$T : \mathbb{R}_+^d \rightarrow \mathbb{R}_+ \times S_+, \quad T(\mathbf{x}) = \left(\|\mathbf{x}\|_r, \frac{\mathbf{x}}{\|\mathbf{x}\|_a} \right) ,$$

La mesure spectrale ρ est une mesure de Radon finie sur S_+ , et on obtient

$$\mu^*(d(r, a)) = r^{-2} dr \times \rho(da) .$$

Il n'y a pas de paramétrisation finie pour la mesure spectrale. Elle caractérise la structure de dépendance extrême d'un vecteur aléatoire dans le domaine d'attraction d'une loi extrême.

4 Champs max-stables

La max-stabilité d'un champ $\{Z(x) \mid x \in \mathbb{R}^d\}$ se définit par la max-stabilité de ses lois fini-dimensionnelles. Comme avant, les champs max-stables se définissent comme processus limites pour une suite de champs aléatoires iid, sous normalisation linéaire par une suite de champs déterministe $a_n(x), b_n(x)$.

4.1 Construction

Nous mentionnons une méthode pour construire des champs max-stables qui est à la fois une caractérisation (voir [1] et [5] pour plus de détails).

Soient (V_i, U_i) les points d'un processus de Poisson de mesure d'intensité $v^{-2} dv \times du$ sur $]0, \infty[\times \mathbb{R}^d$, et soit $f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ une fonction non-négative telle que $\int f(x, y) dy = 1$ pour tout x . Alors

$$Z(x) = \max_{i=1, 2, \dots} (V_i f(x, U_i))$$

définit un champs max-stable sur \mathbb{R}^d avec des marginales Fréchet unité.

En prenant f comme la densité d'une loi normale multivariée, on obtient un processus max-stable dit *processus gaussien de valeurs extrêmes* ([5]).

4.2 Méthodes et outils inférentiels

L'inférence est basée sur les lois fini-dimensionnelles, en particulier sur les lois bivariées pour deux sites spatiaux à distance Δx . Une statistique scalaire résumant la structure de dépendance est le coefficient extrême $\theta(\Delta x) = -\log G(1, 1; \Delta x)$, qui est situé entre 1 (dépendance complète) et 2 (indépendance).

Les outils de géostatistique comme par exemple le variogramme incitent à utiliser des équivalents adaptés au cadre max-stable. Le madogramme, le F -madogramme et le λ -madogramme tiennent compte du fait que les seconds moments des marginales n'existent pas en général (voir par exemple [2]).

Les méthodes d'inférence usuelles sont essentiellement fondées sur l'ajustement de versions paramétriques du coefficient extrême ou sur des approches par maximum de vraisemblance composite. Le plus souvent, seules les expressions analytiques des lois bivariées sont disponibles, ce qui rend toute approche par maximum de vraisemblance classique impossible. Récemment [3] ont proposé une approche par vraisemblance composite sur les paires de maxima par période. L'approche que nous proposons cherche quant à elle à exploiter des propriétés de la mesure spectrale et les processus limites de Poisson des extrêmes multivariés. Différents résultats théoriques sont présentés, et des approches orientées "dépassements de seuil" ou "plus grandes statistiques d'ordre" sont proposées. De nouveaux outils graphiques et des méthodes inférentielles paramétriques innovantes, respectant la structure spatiale ou spatio-temporelle des observations extrêmes sont explicités.

Références

- [1] L. De Haan. A spectral representation for max-stable processes. *The Annals of Probability*, 12(4) :1194–1204, 1984.
- [2] P. Naveau, A. Guillou, D. Cooley, and J. Diebolt. Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1) :1, 2009.
- [3] S.A. Padoan, M. Ribatet, and S.A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489) :263–277, 2010.
- [4] S.I. Resnick. *Extreme values, regular variation and point processes*. Springer Verlag, 2007.
- [5] R.L. Smith. Max-stable processes and spatial extremes. *Unpublished manuscript*, 205, 1990.