Estimation of the extremal coefficient function of a stationary random field. Application to rainfalls maxima.

L. Bel[1], J.-N. Bacro[2] and Ch. Lantuéjoul[3]

[1] Université Paris-Sud
[2] Université de Montpellier II
[3] Ecole des Mines de Paris

Abstract: Spatial environmental processes often exhibit dependence in their maxima. In order to simulate these processes we need to characterize and quantify this dependence. We investigate estimators of the extremal coefficient on simulations of different spatial models and on real data of African rainfalls.

Keywords: Extremal coefficient; Random Field; Estimation; Rainfalls.

1 Introduction

Spatial data sets are now easy to obtain and the spatial process modelling is widely used in most of environmental research. The characterization of the extremal behaviour of such processes are often of fundamental interest and the classical methods of spatial statistics are not adapted. Recently, some notions from the multivariate extreme values theory have been extended to the spatial context (Schlather and Tawn (2003), Ancona-Navarrete and Tawn (2002)) : extremal coefficient function at distance h, $\theta(h)$, $1 \leq \theta(h) \leq 2$, coming from the extremal coefficient notion, and tail dependence coefficient function at distance h, $\eta(h)$, coming from the joint tail models proposed by Ledford and Tawn (1996) in order to describe the dependance for multivariate extremes.

2 Estimation of the extremal coefficient function $\theta(\cdot)$

Let $\{Z(s), s \in \mathbf{R}^2\}$ be a stationary max-stable spatial process with standard Frechet marginal distribution function $F(x) = e^{-\frac{1}{x}}$, $x > 0$. We have (Schlather and Tawn, 2003):

$$P\left(Z(s) < z(s), \forall s \in \mathbf{R}^2\right) = \exp\left(-\int \max_{s \in \mathbf{R}^2} \frac{g(x,s)}{z(s)} \nu(dx)\right)$$

where $g(\cdot,\cdot)$ is a non-negative function such that $\int g(s,x)\nu(dx) = 1$ for $s \in \mathbf{R}^2$. Then,

$$P\left(Z(s) < z, Z(s+h) < z\right) = e^{-\frac{\theta(h)}{z}}$$

with $\theta(h) = \int \max\{g(x,s), g(x,s+h)\}\nu(dx)$.

2.1 Estimation of $\theta(\cdot)$

Two estimators have been proposed in the literature:

1. *Maximum likelihood approach* (Schlather and Tawn, 2003)
Let ${}^t\big(Z^{(1)}(s), Z^{(1)}(s+h)\big), {}^t\big(Z^{(2)}(s), Z^{(2)}(s+h)\big), \ldots, {}^t\big(Z^{(n)}(s), Z^{(n)}(s+h)\big)$ denote n i.i.d. repetitions of ${}^t(Z(s), Z(s+h)) \equiv {}^t(Z_1, Z_2)$. The function $\theta(h) \equiv \theta$ is estimated through the following censored likelihood:
$l(\theta) = \text{card}\left\{j : \max_{1,2} Z_i^{(j)}\bar{Z}_i > z\right\}\log(\theta) - \theta\sum_{j=1}^{n}\left(\max\left\{z, \max_{i=1,2}\left(Z_i^{(j)}\bar{Z}_i\right)\right\}\right)^{-1}$
where $\bar{Z}_i = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{Z_i^{(j)}}$ and z is a sufficiently large value.

Let $\widehat{\theta}_{ST}(h)$ denote the corresponding estimator.

2. *Madogram function approach* $\nu(\cdot)$ (Poncet, Cooley and Naveau (2005)).
The madogram function $\nu(h) = \frac{1}{2}\mathrm{E} \mid Z(s+h) - Z(s) \mid$ is one of the functions usually used in geostatistic to charaterize the spatial structure of a process $Z(\cdot)$ at distance h.
Define $\nu_F(h) \equiv \frac{1}{2}\mathrm{E} \mid F(Z(s+h)) - F(Z(s)) \mid$. Then $\nu_F(h) = \frac{\theta(h)-1}{2(\theta(h)+1)}$, leading to

$$\widehat{\theta}(h) = \frac{1 + 2\hat{\nu_F}(h)}{1 - 2\hat{\nu_F}(h)} \equiv \widehat{\theta}_{PCN}(h)$$

We propose a generalization of $\widehat{\theta}_{PCN}(h)$. Let λ be a strictly positive real and consider the madogram function $\nu_F^{(\lambda)}(h) \equiv \frac{1}{2}\mathrm{E} \mid F^{(\lambda)}(Z(s+h)) - F^{(\lambda)}(Z(s)) \mid$. Then, $\nu_F^{(\lambda)}(h) = \frac{\lambda(\theta(h)-1)}{(\theta(h)+\lambda)(\lambda+1)}$ and $\theta(h)$ can be estimated by a non linear least squares minimization:

$$\widehat{\theta}_{MC}(h) = argmin_\theta \sum_{\lambda=1}^{k}\left(\frac{\lambda+1}{\lambda}\nu_F^{(\lambda)}(h) - g_{\theta,\lambda}(h)\right)^2$$

where $g_{\theta,\lambda}(h) = \frac{\theta(h)-1}{\theta(h)+\lambda}$.

2.2 Behaviour of the estimators

The statistical study of the three estimators $\widehat{\theta}_{ST}, \widehat{\theta}_{CPN}$ and $\widehat{\theta}_{MC}$ can be made in two different ways :

1. using the n independent repetitions of each couple ${}^t(Z(s), Z(s+h))$, coming from $Z^{(1)}$, $Z^{(2)}$, $\ldots$, $Z^{(n)}$;

2. using the stationarity of the process: $\theta(h)$ is estimated from a single realization of the process $Z(\cdot)$, using the ${}^t(Z(s), Z(u))$ values where s and u are observation sites separated by a distance h (or belonging to the same class of distance h). In other words, the independence hypothesis is no longer assumed and the independent realizations of the process are used to calculate empirical statistics for the three estimators. The idea is to illustrate how the estimators are accurate when the stationarity hypothesis is fully exploited. Such an approach is usual in geostatistic: the madogram function is estimated on the basis of an unique sample of m realizations of the underlying process at m spatial sites.

Two spatial models will be considered:

(a) a *storm* model (Smith (1990), Schlather (2002)). For such a model, extreme values are dependent;

(b) a gaussian model, with an exponential covariance function. For gaussian models, it is well known that extremes are asymptotically independent.

The theoretical expression of the $\theta(h)$ functions is known in both cases (a) and (b) and it is possible to compare estimations and true values for fixed h values.

3 Estimation of the tail dependence coefficient function $\eta(\cdot)$

Using the same simulation approachs, the performance of the four usual estimators of the tail dependence coefficient function $\eta(\cdot)$ (Ledford-Tawn (1996), Hill (1975), Peng (1999) and Draisma *et al.* (2001)) can be compared. For small h value, similar problems as for the estimation of the $\theta(h)$ function occur.

4 Application on African rainfalls

We have data on Sahelian rainfalls measured in about 400 stations since 1950. Taking the annual maximum in each station we estimate the extremal coefficient and show that this process exhibit a strong dependence behaviour on its maxima.

References

Ancona-Navarrete, M.A. and Tawn, J.A. (2002). "Diagnostics for pairwise extremal dependence in spatial processes", *Extremes*, 5, 271-285.

Draisma, G., Drees, H., Ferreira, A., and L. de Haan (2001). "Tail dependence in independence". Technical report, EURANDOM, Eindhoven.

Hill, B. M. (1975)."A simple general approach to inference about the tail of a distribution". *Ann. Statist.*, 3, 1163-74.

Ledford, A.W., and Tawn, J.A. (1996). "Statistics for near independence in multivariate extreme values", *Biometrika*, 83, 169-187.

Peng, L. (1999). "Estimation of the coefficient of tail dependence in bivariate extremes". *Statist. Probab. Lett.*, 43, 399-409.

Poncet, P., Cooley, D. and P. Naveau (2005). Variograms for spatial max-stable random fields. *Submitted for publication.*

Schlather, M. (2002). "Models for stationary max-stable random fields". *Extremes*, 5, 33-44.

Schlather, M., and Tawn, J.A. (2003). "A dependance measure for multivariate and spatial extreme values: Properties and inference", *Biometrika*, 90, 139-156.

Smith, R.L. (1990). "Max-stable processes and spatial extremes". Technical report, University of Surrey.