



HAL
open science

Testing the independence of maxima: from bivariate vectors to spatial extreme fields

Jean-Noel Bacro, Liliane Bel, Christian Lantuéjoul

► **To cite this version:**

Jean-Noel Bacro, Liliane Bel, Christian Lantuéjoul. Testing the independence of maxima: from bivariate vectors to spatial extreme fields. *Extremes*, 2010, 13 (2), pp.155-175. 10.1007/s10687-010-0104-8 . hal-00371398

HAL Id: hal-00371398

<https://hal.science/hal-00371398>

Submitted on 27 Mar 2009

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.

Testing the independence of maxima: from bivariate vectors to spatial extreme fields

Asymptotic independence of extremes

Jean-Noël Bacro · Liliane Bel · Christian Lantuéjoul

Version : mars 25, 2009

Abstract Characterizing the behaviour of multivariate or spatial extreme values is of fundamental interest to understand how extreme events tend to occur. In this paper we propose to test for the asymptotic independence of bivariate maxima vectors. Our test statistic is derived from a madogram, a notion classically used in geostatistics to capture spatial structures. The test can be applied to bivariate vectors, and a generalization to the spatial context is proposed. For bivariate vectors, a comparison to the test by Falk and Michel (2006) is conducted through a simulation study. In the spatial case, special attention is paid to pairwise dependence. A multiple test procedure is designed to determine at which lag asymptotic independence takes place. This new procedure is based on the bootstrap distribution of the number of times the null hypothesis is rejected. It is then tested on maxima of three classical spatial models and finally applied to two climate datasets.

Keywords Bivariate extremes · Asymptotic independence · Max-stable random fields · Spatial processes

1 Introduction

For a wide class of environmental and climate studies, spatial extreme values are of fundamental interest since extreme events may have dramatic consequences. If standard geostatistic approaches perform well for statistical inference on the mean

Jean-Noël Bacro
I3M, Université Montpellier II, 4, place Eugène Bataillon, F-34095 Montpellier cedex 5, France.
E-mail: bacro@math.univ-montp2.fr

Liliane Bel
AgroParisTech, 16, rue Claude Bernard, F-75231 Paris cedex 05, France.
E-mail: Liliane.Bel@agroparistech.fr

Christian Lantuéjoul
MinesParisTech, 35 rue Saint-Honoré, F-77305 Fontainebleau, France.
E-mail: Christian.Lantuejoul@ensmp.fr

behaviour of spatial processes, it is well known that these ones are of poor interest when dealing with extreme realizations. As in the univariate case, the mean and the extremes of a spatial process are essentially of different nature and spatial dependencies related to mean values and to extreme ones can be very different. As a consequence, specific approaches and models have to be developed to infer on the extremal behaviour of spatial processes.

Problems concerning environmental extremes are multivariate in character because data sets are more often spatialized. The multivariate extreme value theory (MEVT) offers various notions to capture the main characteristics of the underlying dependence structure (see for example Beirlant et al. (2004) and the references therein). A useful one is asymptotic independence. Roughly speaking, a random vector has its components asymptotically independent if the componentwise maxima are ultimately independent. In practice, asymptotic independence is difficult to detect and may lead to erroneous conclusions concerning the prediction of extreme events if not addressed properly. Typically, fitting a model under a wrong dependence assumption in extreme values may result in a large extrapolation error.

In this paper, we are concerned with testing for pairwise independence of maxima from spatial data. At first, a new and simple test for the asymptotic independence of bivariate vectors of maxima is proposed. An extension to the spatial case is then derived, leading to a test for asymptotic independence on several classes of distance. It involves a standard geostatistical tool that is called a madogram (Matheron 1989). The paper is organized as follows. Bivariate vectors of maxima are first considered. In section 2, the notion of asymptotic independence for bivariate vectors is reminded and classical statistical tests of asymptotic independence are presented. The test we propose is introduced in section 3. Its performances are compared with that of Falk and Michel (2006) using simulations. The spatial framework is then considered. A test for asymptotic independence applicable to random fields of maxima is proposed in section 4. Two cases are distinguished depending on whether a single or multiple realizations are available. Because rejecting the null hypothesis of asymptotic independence on several classes of distances requires a multi-test approach on (spatially) correlated data, a bootstrap procedure controlling the False Discovery Rate (FDR) is set up. This procedure is finally applied, firstly in section 4 on simulated random fields with known asymptotic dependence behaviours, secondly in section 5 on two real data sets (annual maximal temperatures of 29 French towns and annual maximal precipitations over a 30 years period in the French region of Burgundy). Conclusions are drawn and some perspectives for future work are discussed in section 6.

2 Bivariate extreme distributions and asymptotic independence

Let $(X_1, Y_1), (X_2, Y_2) \dots$ be independent copies of a bivariate random vector (X, Y) with distribution function K and marginals F_X and F_Y . The classical extreme bivariate theory is concerned with the limit behaviour of

$$(M_n(X), M_n(Y)) \equiv \left(\max_{i=1, \dots, n} X_i, \max_{i=1, \dots, n} Y_i \right)$$

as $n \rightarrow +\infty$. Because of the definition, the marginals of $(M_n(X), M_n(Y))$ belong to the generalized extreme value (GEV) distribution family. The general form of a GEV distribution is $GEV_{\mu, \sigma, \xi}(x) = \exp\left(-\left[1 + \xi \frac{x-\mu}{\sigma}\right]^{-1/\xi}\right)$ with $\mu \in \mathbb{R}$, $\sigma > 0$, $\xi \in \mathbb{R}$ (Coles 2001).

From now onward, it is assumed without loss of generality that $F_X \equiv F_Y \equiv F$, where $F(\cdot)$ is the unit Fréchet distribution $F(z) = \exp(-1/z)$, $z > 0$. Hence, the limit distribution of $M_n(X)$ and $M_n(Y)$ is also Fréchet.

The following theorem (de Haan and Resnick 1977) characterizes the limit joint distribution of $(M_n(X), M_n(Y))$.

Theorem 1 *If $P(M_n(X) \leq nx, M_n(Y) \leq ny) \xrightarrow{n \rightarrow \infty} G(x, y)$, where G is a non-degenerate distribution function, then $G(\cdot, \cdot)$ takes the form $G(x, y) = \exp(-V(x, y))$ with*

$$V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega)$$

and H is a distribution on $[0, 1]$ with mean $\frac{1}{2}$ (spectral measure).

Note that the marginal distributions of G are unit Fréchet. In the particular case where $H = (\delta_0 + \delta_1)/2$, then $G(x, y) = F(x)F(y)$. We then say that X and Y are *asymptotically independent*. It has been established by Joe (1993) that (X, Y) are asymptotically independent if and only if

$$\chi \equiv \lim_{z \rightarrow \infty} \frac{P(X > z, Y > z)}{1 - F(z)} = \lim_{z \rightarrow \infty} P(Y > z | X > z) = 0$$

Inference on χ is difficult because few observations are available as $z \rightarrow \infty$. Using the properties of V , it can be shown that χ can be written as $2 - \theta$ where $1 \leq \theta \leq 2$ is a coefficient satisfying $G(x, x) = \exp(-\theta/x) \equiv F(x)^\theta$. θ is called the *extremal coefficient* (see Schlather and Tawn (2003) for details). Asymptotic independence corresponds to the case $\theta = 2$.

3 Testing for bivariate asymptotic independence

3.1 Two usual tests

Several methods to test for the bivariate asymptotic independence have been proposed in the literature. Only two of them are discussed here, namely a graphical test by de Haan and de Ronde (1998) and a statistical test by Falk and Michel (2006).

3.1.1 A graphical test

Suppose that G is non-degenerate, and consider the function

$$\ell(x, y) = \lim_{n \rightarrow \infty} nP\left(1 - F(X) < \frac{x}{n} \text{ or } 1 - F(Y) < \frac{y}{n}\right) \quad x, y > 0$$

Its level sets $\mathcal{Q}_c = \{(x, y) \mid \ell(x, y) = c\}$ satisfy a number of interesting properties established by de Haan and de Ronde (1998). In particular, the curve \mathcal{Q}_c along with the axes $x = 0$ and $y = 0$ delimits a convex domain D_c such that $c \leq c' \Rightarrow D_c \subset D_{c'}$. Moreover the graph of this curve is closely related to the strength of the dependence between X and Y . \mathcal{Q}_c is equal to $\{(x, y) \mid x + y = c\}$ if X and Y are asymptotically independent, and $\{(x, y) \mid \max(x, y) = c\}$ if X and Y are fully dependent. A graphical test based on \mathcal{Q}_c curves with different c values can be used: asymptotic independence comes out as a straight line between $(0, c)$ and $(c, 0)$. Such an approach is only a graphical one but can be helpful in practical situations.

3.1.2 A statistical test

Originally set up for bivariate distributions with reverse exponential margins, the test by Falk and Michel (2006) can be adapted to unit Fréchet margins. More specifically, let $\varepsilon > 0$ and $t \in [0, 1]$. When ε tends to 0, the conditional distribution function

$$K_\varepsilon(t) \equiv P\{X^{-1} + Y^{-1} < \varepsilon t \mid X^{-1} + Y^{-1} < \varepsilon\}$$

tends to t^2 if X and Y are asymptotically independent, and t otherwise. This result can be used to test for the asymptotic independence of X and Y using classical goodness-of-fit tests such as the Kolmogorov-Smirnov or the likelihood ratio ones as well as the chi-square test.

Note that Frick et al. (2007) recently proposed a generalization of Falk and Michel's work, based on a second order differential expansion of the spectral decomposition of G . They focused on the case of a null hypothesis of tail dependence against a composite alternative representing the various degrees of tail independence. Since these hypothesis differ notably from the ones we consider, we will restrict our comparison to the Falk and Michel approach.

3.2 A statistical test based on a madogram

In this section, the test we propose for bivariate vectors is presented. Simulations are then used to compare its performances with those of the two aforementioned tests.

3.2.1 A new statistical test for bivariate vectors

Suppose again that G is non-degenerate with spectral measure H , and consider the random variable

$$W = \frac{1}{2} |F(X) - F(Y)|$$

Because $F(X)$ and $F(Y)$ are uniformly distributed on $[0, 1[$, their dependence relationships specify the distribution of W . If $X = Y$ almost surely, then the distribution of W is a Dirac distribution at 0. If X and Y are independent, then W admits the p.d.f. $f_W(z) = 4 - 8z$ on $[0, \frac{1}{2}[$. Of course many intermediary situations are possible between perfect dependence and full independence. In other words, the distribution of W provides information about the asymptotic dependence between X and Y . Cooley

et al. (2006) have shown that the mean of W is related to the extremal coefficient θ by the formula

$$\mathbb{E}(W) = \frac{1}{2} \frac{\theta - 1}{\theta + 1} \equiv v_W. \quad (1)$$

This formula and some variations (Bel et al. 2008), can be used for estimating θ . Regarding the variance of W , a proof is given in the appendix that

$$\sigma_W^2 = \frac{1}{6} - \frac{1}{4} \left(\frac{\theta - 1}{\theta + 1} \right)^2 - \frac{1}{2} \int_0^1 \frac{dt}{[1 + A(t)]^2} \quad (2)$$

where A is the classical Pickands dependence function

$$A(t) = 2 \int_0^1 \max(\omega(1-t), (1-\omega)t) dH(\omega)$$

In the case of asymptotic independence, $A(t) \equiv 1$ and $\sigma_W^2 = \frac{1}{72}$.

Formula (1) shows that v_W is a strictly monotonic increasing function of θ with $v_W = \frac{1}{6}$ if and only if $\theta = 2$. Accordingly, the asymptotic independence of X and Y can be checked by testing the null hypothesis $H_0 : v_W = \frac{1}{6}$ against $H_1 : v_W < \frac{1}{6}$. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sequence of independent copies of (X, Y) . Then a natural estimator for v_W is

$$\widehat{v}_W = \frac{1}{2n} \sum_{i=1}^n |\widehat{F}(X_i) - \widehat{F}(Y_i)|$$

where \widehat{F} is the empirical distribution function as provided by the X_i 's and the Y_i 's.

It can be shown (Fermanian et al. (2004)) that

$$\sqrt{n} \frac{\widehat{v}_W - \frac{1}{6}}{\widehat{\sigma}_W} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad (3)$$

as $n \rightarrow \infty$ under H_0 , providing a straightforward test for the asymptotic independence of bivariate random vectors.

As the expectation v_W is related to the well-known madogram used in spatial statistics, we call this test the madogram test.

3.2.2 Simulation study

Samples of independent bivariate vectors of maxima are simulated and the null hypothesis *asymptotic independence* is tested against *non asymptotic independence*. Four bivariate distributions $G(\cdot, \cdot)$ are considered. The first three ones are bivariate max-stable distributions whereas the last one is an empirical distribution of bivariate gaussian maxima. The max-stable distributions considered are the following ones:

1. Logistic distribution (Tawn 1988)

Defined as

$$G(x, y) = \exp \left\{ - \left(x^{-1/\alpha} + y^{-1/\alpha} \right)^\alpha \right\} \quad x, y > 0,$$

this bivariate distribution depends on one parameter $0 < \alpha \leq 1$. When $\alpha = 1$, we have $G(x, y) = \exp\{-(x^{-1} + y^{-1})\}$, which is the independence case. When $\alpha \rightarrow 0$, then we get $G(x, y) \rightarrow \exp\{-\max(x^{-1}, y^{-1})\}$, which is the perfect dependence case.

2. Asymmetric logistic distribution (Tawn 1988)

This generalization of the logistic distribution allows asymmetry and nonexchangeability. Its general form is given by

$$G(x, y) = \exp \left\{ - \frac{1 - \psi_1}{x} - \frac{1 - \psi_2}{y} - \left(\left(\frac{\psi_1}{x} \right)^{1/\alpha} + \left(\frac{\psi_2}{y} \right)^{1/\alpha} \right)^\alpha \right\} \quad x, y > 0$$

It depends on 3 parameters, namely $0 < \alpha \leq 1$ and $0 \leq \psi_1, \psi_2 \leq 1$. Independence corresponds to $\alpha = 1$ or $\psi_1 \psi_2 = 0$, whereas perfect dependence is obtained when $\psi_1 \psi_2 = 1$ and $\alpha \rightarrow 0$.

3. Hüsler-Reiss distribution (1989)

This bivariate distribution is defined as

$$G(x, y) = \exp \left\{ - \frac{1}{x} \Phi \left(\frac{1}{\alpha} + \frac{\alpha}{2} \log \frac{y}{x} \right) - \frac{1}{y} \Phi \left(\frac{1}{\alpha} + \frac{\alpha}{2} \log \frac{x}{y} \right) \right\} \quad x, y > 0,$$

where $\phi(\cdot)$ is the standard normal distribution function. It depends on one parameter $\alpha > 0$. Independence and perfect dependence are limit cases respectively obtained when $\alpha \rightarrow 0$ and $\alpha \rightarrow \infty$.

500 samples of size 500 are simulated per model. For the first three models, the samples are drawn using the *R*-package *evd* Stephenson (2002) with specified parameter values. Regarding the gaussian model, each sample is generated by taking 500 times the componentwise maxima over 2000 realizations of a bigaussian vector.

Figure 1 shows how the graphical test performs on those 500×500 samples. It can be checked that the fourth model produces straight contour lines, which is not the case for the three max-stable models.

Asymptotic independence is now tested using 4 different statistics, namely the madogram statistic and three other statistics derived from the Falk and Michel (2006) test procedure (Neyman-Pearson, Kolmogorov-Smirnov and Chi-square statistics). This procedure requires a constant ε that has been chosen as the 90% quantile of the distribution of $X + Y$.

The power function, that is $P(\text{rejecting } H_0 \mid \alpha)$, is calculated on the three max-stable models using various parameter values. For the logistic and asymmetric logistic models, α varies from 0.1 to 0.9 by steps of 0.1, and from 0.90 to 0.99 by steps of

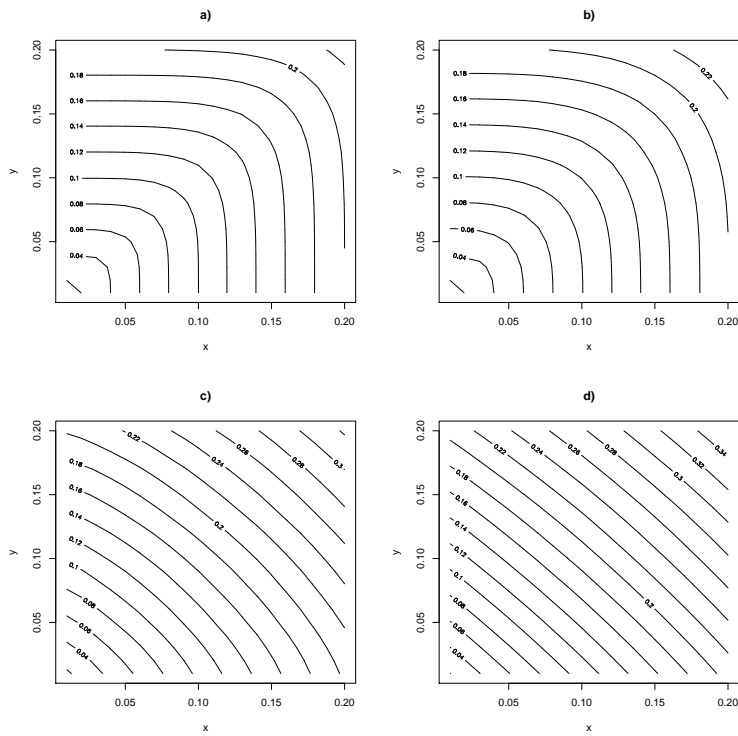


Fig. 1 de Haan graphical test for models a) logistic ($\alpha = 0.2$), b) asymmetric logistic ($\psi_1 = \psi_2 = 0.3$), c) Hüsler-Reiss ($\alpha = 0.9$), d) Gaussian ($\rho = 0.5$)

0.01. The two parameters ψ_1 and ψ_2 of the asymmetric model are set to 0.4. For the Hüsler-Reiss model, α varies from 0.01 to 0.09, from 0.1 to 1.0 and from 1. to 3. by respective steps of 0.01, 0.1 and 1.

The type I error, that is $P(\text{accepting } H_0 \mid \rho)$, is calculated on a gaussian simulation because the maxima of bigaussian vectors are well-known to be asymptotically independent. The correlation ρ varies from -0.9 to 0.9 by steps of 0.1 and from 0.90 to 0.99 by steps of 0.01.

Results for the power functions are presented in Figures 2, 3 and 4. Despite its simplicity, the madogram test based appears often as the most powerful. The Falk and Michel procedure using the Neyman-Pearson goodness of fit test is slightly less powerful meanwhile Chi-square and Kolmogorov-Smirnov tests are not powerful at all. Regarding the gaussian case, Figure 5 shows that when ρ is growing the type I error increases, and this more pronounced for the madogram test and the Neyman-Pearson version of the Falk and Michel test.

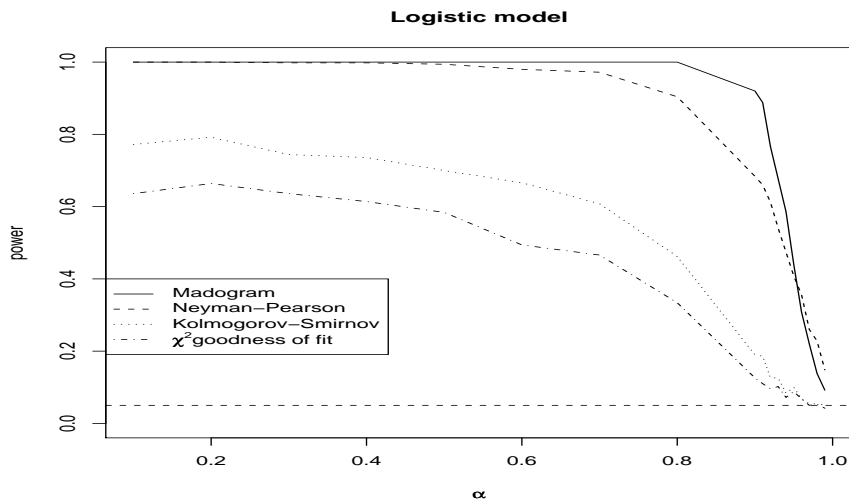


Fig. 2 Power function for the logistic model. The parameter α is related to the strength of the dependence

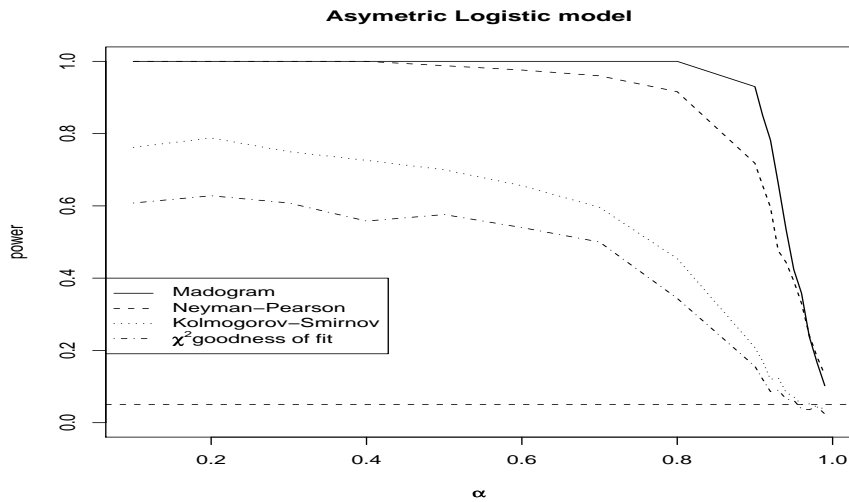


Fig. 3 Power function for the asymmetric logistic model. The parameter α is related to the strength of the dependence and asymmetry parameters are fixed to 0.4.

4 The spatial framework

4.1 A test for spatial asymptotic independence

Let $Z(\cdot)$ be a stationary, isotropic random field with unit Fréchet marginal distribution. In geostatistics, it is usual to characterize spatial dependence using the semi-variogram $\gamma(h) = \frac{1}{2}E(Z(s+h) - Z(s))^2$. However second order moments may not be

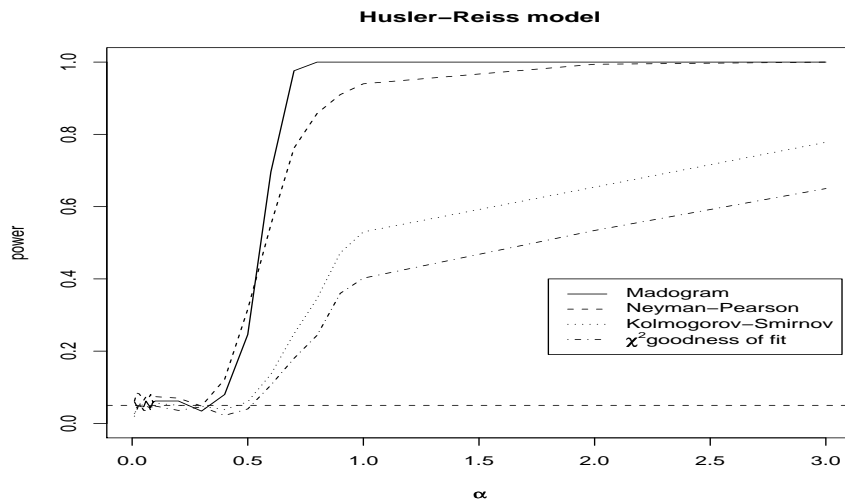


Fig. 4 Power function for the Hüsler-Reiss model. The parameter α is related to the strength of the dependence.

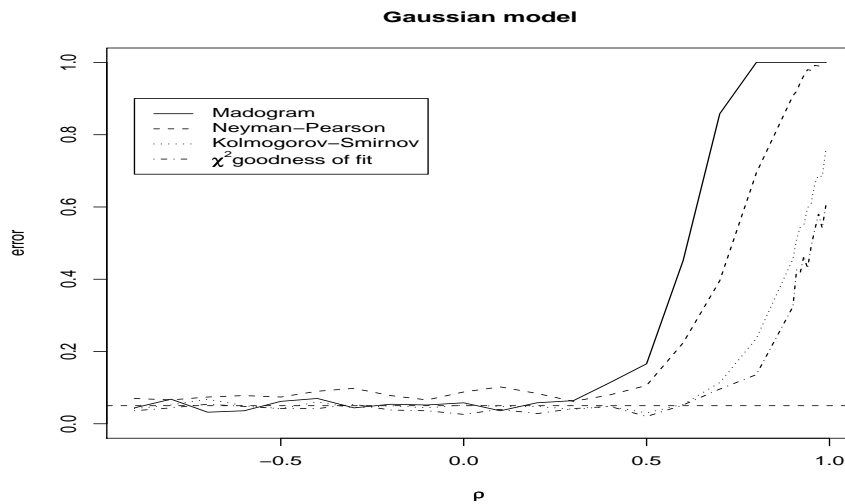


Fig. 5 Type I error function for the maxima of Gaussian model. Parameter ρ is the correlation.

adapted for extremes and in the sequel the madogram $v_Z(h) = \frac{1}{2}E | Z(s+h) - Z(s) |$ is considered instead. In order to determine whether asymptotic independence takes place at lag h , the previous results on bivariate random vectors (section 3) can be applied to all pairs $(Z(s_k), Z(s_\ell))$ of samples such that $\|s_k - s_\ell\| = h$ or more generally $h - \varepsilon \leq \|s_k - s_\ell\| \leq h + \varepsilon$ for some prespecified $\varepsilon > 0$. Two different situations are considered depending on whether a single realization or more than one indepen-

dent realizations of the random field are available at each (sampled) site because the procedure for testing asymptotic independence is not the same.

4.1.1 Case of multiple realizations

Let C be one class of distances. In order to test asymptotic independence (AI) on C , we are faced with a problem of multiple hypothesis testing that needs to be carefully handled. Controlling a global type I error, say α , under multiple hypothesis testing is known to be difficult, and the false discovery rate (FDR) approach (Benjamini and Hochberg 1995) appears as one of the most relevant to cope with it.

The FDR is defined as the expected proportion of the rejected null hypotheses among the rejected hypotheses. Starting from the p -values associated to all pairs of sites of class C , the FDR approach makes it possible to determine which p -values are significant while controlling the fixed global type I error. In order to decide whether the AI hypothesis is admissible or not, we need to assess how many pairs would be wrongly rejected under the null hypothesis of global AI. A decision rule could be, for instance, that there are less than a fixed proportion α of significant p -values. In fact, this procedure would be justified if the p -values distribution was uniform under the null hypothesis. In our case, the test statistics are spatially dependent, so that the p -value distribution is unknown under the null hypothesis. A way to cope with this problem is to resort to a bootstrap procedure. The test statistics that we consider satisfy the PRDS property¹, so that the standard Benjamini-Hochberg procedure for determining significant p -values controls the FDR as in an independent framework (Benjamini and Yekutieli 2001). Let R be the (data based) number of individual hypothesis rejected using this approach. The following procedure aims at simulating the distribution of R under the null hypothesis of AI while preserving the spatial dependence structure. In this procedure, S_ℓ stands for the set of all sites whose distance to a given site s_ℓ belongs to class C . Its cardinality is denoted by m_ℓ .

1. Let $\mathbf{Z}(s_\ell) = (Z^i(s_\ell), i = 1, \dots, n)$ be the vector of realizations at s_ℓ and denote by $\mathbf{Z}^*(s_\ell) = (Z^{i*}(s_\ell), i = 1, \dots, n)$ the resampling vector at s_ℓ obtained from a standard bootstrap on $\mathbf{Z}(s_\ell)$;
2. for each site $s_j \in S_\ell$ compute $\widehat{\mathbf{v}}_{F(\mathbf{Z})}^*(s_\ell, s_j) = \frac{1}{2n} \sum_{i=1}^n \left| \widehat{F}(Z^{i*}(s_\ell)) - \widehat{F}(Z^i(s_j)) \right|$;
3. compute the vector of the associated p -values $(p_1^*, \dots, p_{m_\ell}^*)$;
4. assign successively the role of s_ℓ to each sampled site and repeat (1) to (3). For each particular site s_ℓ , pairs of sites which have been already taken into account along the procedure are removed;
5. consider the complete sample of p -values obtained at the end of (4) and use the FDR approach to count the number of rejected null hypotheses R^* among all the performed tests;

¹ PRDS stands for *Positive Regression Dependency on Subset*. Let \leq be a partial order on \mathbb{R}^m . A subset D of \mathbb{R}^m is called an upper set if $x \in D$ and $y \geq x$ implies $y \in D$. Then the random vector $\mathbf{X} = (X_1, \dots, X_m)$ is said to satisfy the PRDS property on $I_0 \subset \{1, \dots, m\}$ if for each upper set D and for each $i \in I_0$, $P(\mathbf{X} \in D \mid X_i = x)$ is a non-decreasing function in x .

6. repeat (1) to (5) a large number of times and derive the empirical distribution of R^* .

Resampling values of one of the terms in the pair ensures the asymptotic independence, while not resampling the other term will let the p -values being spatially dependent. Hence this procedure gives an empirical distribution of the number R^* of rejected hypotheses under the assumption of global asymptotic independence in a spatial dependent framework. It then suffices to compute empirically $P(R^* \geq R)$ to conclude about the validity of this assumption.

4.1.2 Case of a single realization

When independent copies of the maxima field are not available, the stationarity property of the field is to be used in order to calculate the estimators. Margin law is given by $\widehat{F}(z) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{\{Z(s_i) \leq z\}}$ and the madogram is estimated in each distance class C_k by

$$h \in C_k \quad \widehat{v}_F(h) = \frac{1}{2n_k} \sum_{d(s_{\ell_1}, s_{\ell_2}) \in C_k} |\widehat{F}(Z(s_{\ell_1})) - \widehat{F}(Z(s_{\ell_2}))|$$

The asymptotic independence on each class can then be tested by the standard Benjamini-Hochberg procedure, but clearly the relevance of (3) will be affected because of spatial dependencies.

4.2 Simulation study

Three random field models are considered, namely the gaussian model, the storm model and the gaussian extremal model. Note that only the last two models are *max-stable*. This prompted us to also consider *random fields of maxima* that are the maxima of 100 independent copies of a given random field. Independent copies of such random fields of maxima are obviously obtained by repeating this procedure.

Gaussian model:

Let Z_1, Z_2, \dots, Z_n be n independent copies of a 2-dimensional, stationary gaussian random field with autocorrelation function ρ , and let $M_n = \frac{1}{n} \max(Z_1, Z_2, \dots, Z_n)$. For each site $s \in \mathbb{R}^2$, the bivariate distribution of $(M_n(s), M_n(s+h))$ is asymptotically independent. Accordingly $\theta(h) = 2$ as soon as $h \neq 0$.

Storm model:

Introduced by Smith (1990), this model is defined by $Z(s) = \sup_{j=1,2,\dots} \zeta_j g(x_j - s)$, $s \in \mathbb{R}^2$, where $(\zeta_j, x_j)_j$ are the points of a Poisson process on $]0, \infty[\times \mathbb{R}^2$ with intensity measure $d\Lambda(\zeta, x) = \zeta^{-2} d\zeta dx$. The process Z is max-stable with unit Fréchet margins. Here the deterministic function g is chosen as a bigaussian pdf with covariance matrix M . Let $M_n = \frac{1}{n} \max(Z_1, Z_2, \dots, Z_n)$ as before. Then the random vector $(M_n(s), M_n(s+h))$ is Hüsler-Reiss distributed with extremal coefficient function $\theta(h) = 2\Phi(\frac{1}{2}\sqrt{h'M^{-1}h})$, where Φ denotes the standard gaussian distribution function. As a consequence, asymptotic independence is expected as soon as h has its modulus large enough.

Extremal gaussian model:

Let Y be a stationary standard gaussian random field with correlation function ρ , and let Π be a Poisson point process on $]0, \infty[$ with intensity measure $d\Lambda(\zeta) = \sqrt{2\pi}\zeta^{-2}d\zeta$. The extremal gaussian process proposed by Schlather (2002) is defined as $Z(s) = \max_{\zeta \in \Pi} \zeta Y_{\zeta}(s)$, $s \in \mathbb{R}^2$. It is asymptotically dependent with extremal function $\theta(h) = 1 + \sqrt{1 - \rho(h)}/\sqrt{2}$. It should be pointed out that $\lim_{|h| \rightarrow \infty} \theta(h) = 1 + 1/\sqrt{2} < 2$, even if $\rho(h)$ tends to 0 as the modulus $|h|$ of h tends to ∞ . Therefore asymptotic independence does not occur at any lag.

4.2.1 Case of multiple realizations

Results are presented on Figures 6 to 8. For the gaussian random field, the observed value R of the number of rejected hypothesis fully agrees with the bootstrap distribution of R^* , and the global asymptotic independence is accepted at all lags. For the gaussian storm process, the null hypothesis is rejected up to when $|h| = 2.5$ and accepted beyond that distance. It can be mentioned that $\theta(2.5) = 1.79$ with the correlation matrix $M = Id$ considered for this exercise. For the extremal gaussian model with correlation function $\rho(h) = \exp(-|h|)$, the null hypothesis is always rejected, whatever the class of distance considered, which was the expected conclusion as this random field does not display asymptotic independence at any lag.

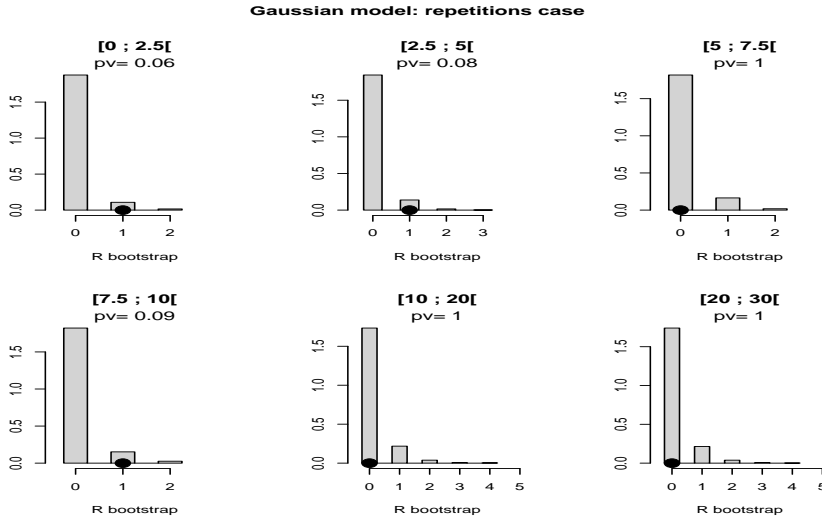


Fig. 6 Gaussian model, distribution of R^* , number of rejected hypothesis under the null hypothesis for some classes of distances. Bullet: observed R , $p_v = P(R^* \geq R)$.

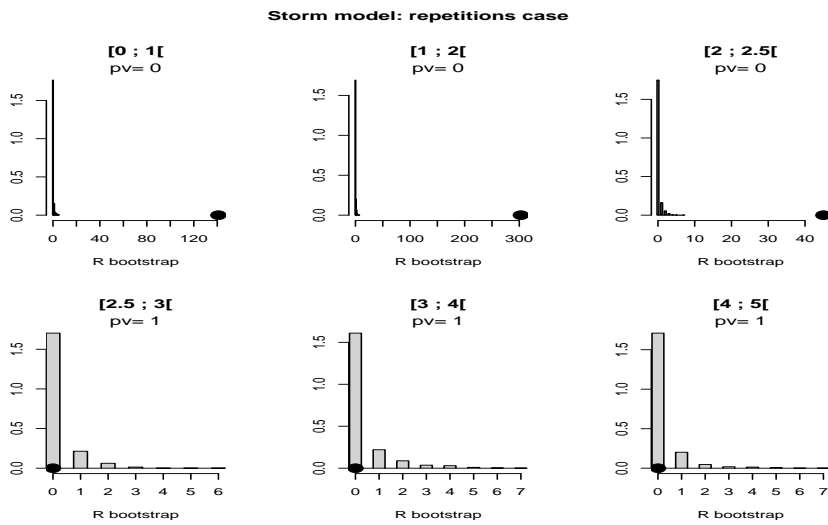


Fig. 7 Storm model, distribution of R^* , number of rejected hypothesis under the null hypothesis for some classes of distances. Bullet: observed R , $p_v = P(R^* \geq R)$.

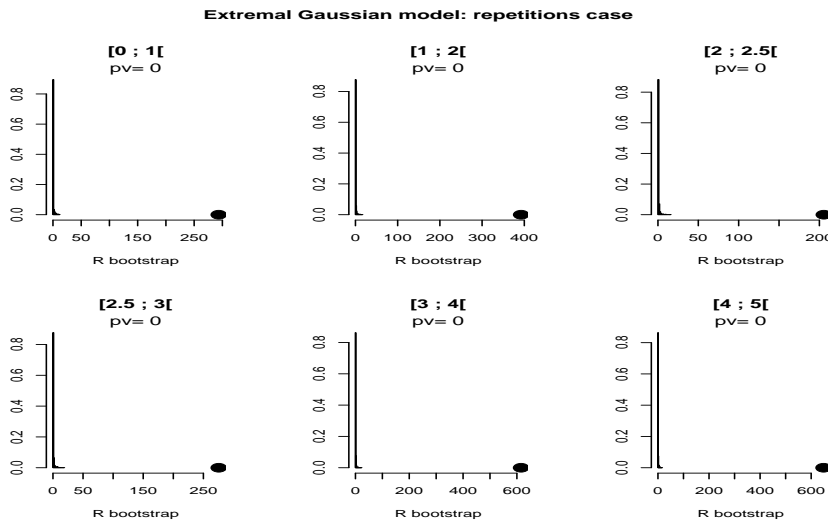


Fig. 8 Extremal gaussian model, distribution of R^* , number of rejected hypothesis under the null hypothesis for some classes of distances. Bullet: observed R , $p_v = P(R^* \geq R)$.

4.2.2 Case of a single realization

For each random field of maxima, a p -value based on the madogram is computed at each distance class. Then standard FDR procedure is applied to determine which p -values are significant. The same operation is repeated 100 times. Tables 1 to 3 give the number of significant and non significant p -values obtained on those realizations.

Table 1 Gaussian model, number of significant and non significant p -values for several classes of distances in the spatial case.

Classes	[0 ; 2.5[[2.5 ; 5[[5 ; 7.5[[7.5 ; 10[[10 ; 20[[20 ; 30[
significant	100	100	18	15	0	0
non significant	0	0	82	85	100	100

Table 2 Storm model, number of significant and non significant p -values for several classes of distances in the spatial case.

Classes	[0 ; 1[[1 ; 2[[2 ; 2.5[[2.5 ; 3[[3 ; 4[[4 ; 5[
significant	100	84	33	17	12	8
non significant	0	16	67	83	88	92

Table 3 Extremal gaussian model, number of significant and non significant p -values for several classes of distances in the spatial case.

Classes	[0 ; 1[[1 ; 2[[2 ; 2.5[[2.5 ; 3[[3 ; 4[[4 ; 5[
significant	100	92	69	49	44	26
non significant	0	8	31	51	56	74

For the gaussian random field and the first class of distances ($h < 0.05$), the p -values are given as significant for almost all copies. The correlation at small distances is too high for asymptotic independence to be detected. For the storm model, the number of non significant p -values becomes larger than that of significant p -values at distance $|h| = 2.0$, which is a bit smaller than what was obtained in the case of multiple realizations. For the gaussian extremal model, the number of non significant p -values becomes larger than that of significant p -values at a distance $h = 0.3$, which should be never the case for this random field.

5 Applications: temperatures and precipitations

5.1 Temperature data

In this section, the spatial asymptotic independence of temperatures in France is investigated. The data consist of daily maxima temperature recorded at 29 locations for more than 50 years, some of them up to 100 years. The closest sites are 40 km apart. In order to have the data compatible with a stationary model, a spatiotemporal trend was removed by calculating an annual profile $\bar{X}(t, s)$ at each site. The annual maxima process considered for analysis is then the yearly block maximum

$$M_y(s) = \max_{t=1, \dots, 365} \{X(t, y, s) - \bar{X}(t, s)\}, \quad y = 1, \dots, Y$$

where Y is the total number of years available for location s .

Figure 9 shows the distribution of R^* produced by the bootstrap procedure under the null hypothesis, together with the value of R observed on the data. The number of rejected hypothesis is always very high whatever the class of distance (actually almost all the p -values are significant even at large distances). We must conclude that the field of temperature exhibits a very strong asymptotic dependence at all distances.

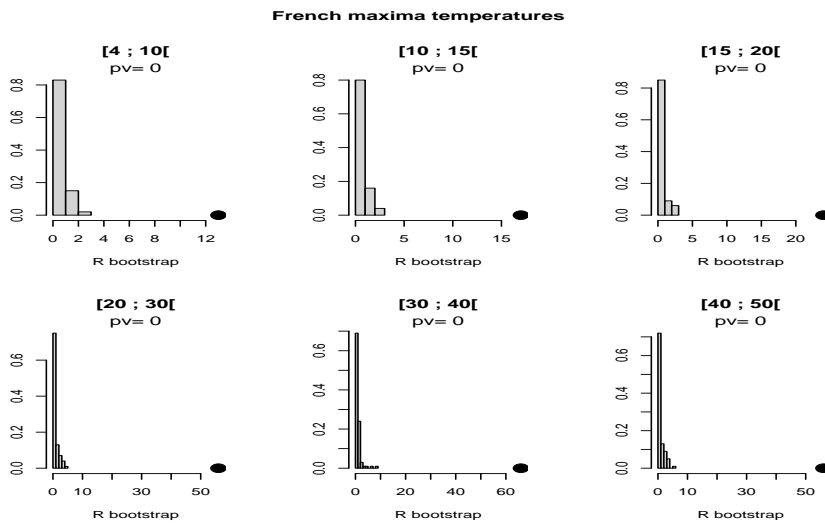


Fig. 9 Maxima French temperatures, distribution of R^* , number of rejected hypothesis under the null hypothesis for some classes of distances. Bullet: observed R , $p_v = P(R^* \geq R)$

5.2 Precipitations data

The second example deals with the precipitations in the French region of Burgundy. The basic data consist of maximum daily precipitations recorded at 146 locations during 30 years. They were preprocessed by Meteo France research laboratory in order to make them compatible with a stationary and isotropic random field. As a result, only the maxima of the resultant field over the whole period were available, which precluded an analysis based on several realizations.

A p -value is computed for each class of distances based on the estimated madogram. Then the FDR procedure is performed to determine which p -values are significant, whereby one can derive the distance up to which the process shows asymptotic dependence. Table 4 shows that this distance is about $h = 50$ km.

Table 4 Burgundy maxima precipitations, p -values function of distance in testing asymptotic independence. Significant p -values are in bold fonts.

distance (km)	0–7.5	7.5–15	15–22.5	22.5–30	30–40	40–60	60–80	80–100
p -values	0.0	0.0	0.049	0.0	0.0	0.05	0.10	0.63

6 Conclusion

Prediction of spatial extreme events appears as a major challenge for various research communities such as environmental ones. A possible way to predict spatial extreme

events is to consider scenarios of extreme realizations using a simulation approach. Such an approach needs a model to be specified, so that any information about the type of extreme dependence at each lag is of fundamental interest for spatial modeling. The test procedure that has been proposed allows us to discriminate between asymptotic independence and asymptotic dependence on the basis of realizations of maxima. It constitutes a first and helpful tool toward that goal.

APPENDIX

Let (X, Y) be a bivariate random vector with distribution $G(\cdot, \cdot)$ and margins Fréchet $F(\cdot)$. Denote by $A(\cdot)$ the related Pickands dependence function. Then,

$$G(x, y) = \exp\left(\left(-\frac{1}{x} - \frac{1}{y}\right)A\left(\frac{x}{x+y}\right)\right)$$

Let $U = F(X) = e^{-\frac{1}{X}}$ and $V = F(Y) = e^{-\frac{1}{Y}}$.

$$\begin{aligned} E(UV) &= E\left(\int_0^1 \int_0^1 \mathbf{1}_{u < U} \mathbf{1}_{v < V} dudv\right) \\ &= \int_0^1 \int_0^1 P\left(X < \frac{-1}{\log u}, Y < \frac{-1}{\log v}\right) dudv \\ &= \int_0^1 \int_0^1 \exp\left((\log u + \log v)A\left(\frac{\log v}{\log u + \log v}\right)\right) dudv \end{aligned}$$

The change of variables $u = e^z$, $v = e^{\frac{z}{1-t}}$ gives

$$\begin{aligned} E(UV) &= -\int_0^1 \int_{-\infty}^0 \exp\left(\frac{z}{(1-t)}A(t)\right) e^{\frac{z}{1-t}} \frac{z}{(1-t)^2} dz dt \\ &= \int_0^1 \frac{1}{(1+A(t))^2} dt \end{aligned}$$

leading to equation (2).

References

- Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). Statistics of extremes. Theory and applications. With contributions from Daniel De Waal and Chris Ferro. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.
- Bel, L., Bacro, J., and Lantuéjoul, C. (2008). Assessing extremal dependence of environmental spatial fields. *Environmetrics*, 19(2):163–182.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B*, 57:289–300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29:1165–1188.
- Coles, S. (2001). An introduction to statistical modeling of extreme values. Springer, London.

-
- Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Doukhan, P., et Soulier, P., editors, Dependence in Probability and Statistics, Springer Lecture Notes in Statistics. Springer, New York.
- de Haan, L. and de Ronde, J. (1998). Sea and Wind: Multivariate Extreme at Work. Extremes, 40:7–45.
- de Haan, L. and Resnick, S. (1977). Limit theory for multivariate sample extremes. Z. Wahrscheinlichkeitstheorie, 1:317–337.
- Falk, M. and Michel, R. (2006). Testing for Tail Independence in Extreme Value models. Annals of the Institute of Statistical Mathematics, 58:261–290.
- Fermanian, J., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. Bernoulli, 10(5):847–860.
- Frick, M., Kaufmann, E., and Reiss, R. (2007). Testing the tail-dependent based on the radial component. Extremes, 10:109–128.
- Joe, H. (1993). Parametric family of multivariate distributions with given margins. J. Multivariate Anal., 46:262–282.
- Matheron, G. (1989). The internal consistency of models in geostatistics. In Armstrong, M., editor, Geostatistics, pages 21–38. Kluwer Academic Publishers, Dordrecht.
- Schlather, M. (2002). Models for stationary max-stable random fields. Extremes, 5:33–44.
- Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. Biometrika, 90:139–156.
- Smith, R. (1990). Max-stable processes and spatial extremes. Technical report, University of Surrey.
- Stephenson, A. (2002). EVD: extreme value distributions. R News, 2-2:31–32.
- Tawn, J. (1988). Bivariate extreme value theory: models and estimation. Biometrika, 75:397–415.