CENSORED PAIRWISE LIKELIHOOD-BASED TESTS FOR MIXTURE PARAMETER OF SPATIAL MAX-MIXTURE MODELS

Abdul-Fattah Abu-Awwad* ,* Véronique Maume-Deschamps* and Pierre Riberau*
*Institut Camille Jordan, Université Claude Bernard Lyon 1, Université de Lyon, France

ABSTRACT
Max-Mixture (MM) processes are defined as $Z = \max\{aX, (1-a)Y\}$ with X an asymptotic dependent (AD) process, Y an asymptotic independent (AI) process and $a \in [0,1]$. So that, the mixture parameter a controls the level of the AD part present in the MM process Z. Here we focus on two statistical tests for the mixing parameter a which are based on censored pairwise likelihood estimates. We compare their performance through an extensive simulation study. Monte Carlo simulation are a fundamental tool for asymptotic variance calculations. We apply our tests to daily precipitations from the East of Australia. Limitations and possible developments of the approach are discussed.

KEYWORDS: Composite likelihood; max-stable process; max-mixture models; pairwise likelihood; Monte Carlo simulation.

MSC: 62F03; 62F05; 62M30; 60G70

RESUMEN
Los procesos Max-mixture son definidos por $Z = \max\{aX, (1-a)Y\}$ donde X es un proceso asintóticamente dependiente (AD), Y es un proceso asintóticamente independiente (AI) y $a \in [0,1]$. Asi el parámetro de mezcla a controla el nivel de la parte asintóticamente dependiente en el proceso MM Z. En este articulo, nos interesamos a 2 tests estadísticos sobre el parametro de mezcla a, estos tests están basados sobre estimadores de la verosimilitud bidimensional censurada. Comparamos sus capacidades con un amplio estudio de simulación. Las simulaciones de Monte-Carlo juegan un papel importante en el cálculo de la varianza asintótica. Utilizamos estos tests sobre precipitaciones en el Este de Australia. Concluimos con las limitaciones y los posibles desarrollos de estas herramientas.

PALABRAS CLAVE: Verosimilitud compuesta; Proceso max-estable; Modelos Max-Mixture; Verosimilitud bidimensional; Simulaciones de Monte-Carlo.

1. INTRODUCTION

The rise of risky environmental events leads to renewed interest in the statistical modelling of extremes, for instance, modelling extreme precipitation is pivotal in flood protection. In the last decade, max-stable (MS) models have arised as a common tool for modeling spatial extremes, since they extend the

*abuawwad@math.univ-lyon1.fr

gereralized extreme value (GEV) distribution to the spatial setting. The definitions on MS processes, GEV distributions ... are given in details in Section 2.. MS process models for spatial data were first constructed using the spectral representation by [18]. Several subsequent works on the construction of spatial MS processes have been developed, see e.g., [40, 37, 26, 16]. For a detailed overview of MS processes, we refer the reader to [20].

Within the class of MS models, only two types of dependence structures are possible; either the process is asymptotically dependent (AD) or it is exactly independent [41, 15, 2]. Whereas, in spatial context, especially in the environmental domain, many dependence scenarios could arise, where AD and asymptotic independent (AI) might cohabite. Hence, fitting AD models (MS models) to AI data may lead to over/under estimation of probabilities of extreme joint events, since it is wrongfully assumed that the most extreme marginal events may occur simultaneously [10]. Alternatively, [46] developed a a flexible class of MM models. The basic idea is to mix MS and AI processes: $Z = \max\{aX, (1-a)Y\}$, with X an AD process, Y an AI process and $a \in [0,1]$. So that, these models are able to capture both AD and AI, and the spatial extremal dependencies are varying according to the value of a.

Hypothesis testing is one of the main tools in statistics and crucial in many applications. In the present paper, we propose two statistical tests on the value of the mixing parameter of a MM process. Our objective is to facilitate the modeling of spatial data by a random field with appropriate extremal behaviour. In particular, we expand the use of two classical statistics to hypothesis testing in the area of spatial extremes.

Parametric likelihood inference for MS models is not possible in general, since the full likelihood is not easily computable for a MS vector in dimension greater than 2. For this reason a methodology based on partial specification of the full likelihood is proposed. The composite likelihood which has gained its popularity thanks to its computational manageability and its theoretical properties [30, 43]. In particular, maximum pairwise likelihood estimation for MS models were first suggested by [34] and is now widely used, see e.g., [17, 23]. As with MS models, MM models can be fitted using composite likelihood and compared using the composite likelihood information criterion (CLIC), see, e.g., [46, 2].

The remainder of the paper is organized as follows. Section 2. reviews the definitions and theory of spatial extreme processes. The censored pairwise likelihood approach is presented for the statistical inference in Section 3., our proposed pairwise statistics and their main properties are detailed in Section 4. In Section 5., we show by means of a series of simulation studies the performance of our proposed tests. In Section 6., we illustrate our testing approach by the analysis of daily precipitation from the East of Australia. Concluding remarks and some perspectives are addressed in Section 7.. Some auxiliary results are summarized in an appendix.

2. SPATIAL EXTREMES MODELING

Throughout the paper, $\{X(s)\}_{s\in\mathcal{S}}$, $\mathcal{S} \subset \mathbb{R}^d$ (generally, $d = 2$) is a spatial process, and for the sake of simplicity it will be assumed to be strictly stationary. This means that for any $k \in \mathbb{N}$, $s_1, \ldots, s_k \in \mathcal{S}$ and $s_1 + h, \ldots, s_k + h \in \mathcal{S}$,

$$(X(s_1), \ldots, X(s_k))^t \overset{\mathcal{D}}{=} (X(s_1 + h), \ldots, X(s_k + h))^t,$$

provided that h is any translation vector. In other words, strict stationarity implies that the finite-dimensional distribution is unaffected by the translation of an arbitrary quantity $h \in \mathbb{R}^d$.
We shall assume also the the process is isotropic, which means: for any $\mathbb{R}^d$ isometry m with $m(s_i) \in \mathcal{S}$, $i = 1, \ldots, k$,

$$(X(s_1), \ldots, X(s_k))^t \stackrel{\mathcal{D}}{=} (X(m(s_1)), \ldots, X(m(s_k)))^t .$$

For a detailed description on the fundamentals of spatial stochastic processes, see, e.g. [12, 31].

2.1. Max-stable processes

MS processes form the natural extension of multivariate extreme value distributions to infinite dimensions. We briefly review here spatial MS processes. Suppose that $\{Y_i(s) : s \in \mathcal{S} \subset \mathbb{R}^d\}$, $i = 0, 1, 2, \ldots$, are independent and identically distributed (i.i.d) replicates of a random process $Y(s)$, and that there are sequences of continuous functions $\{a_n(s) > 0\}$ and $\{b_n(s)\}$ such that, the rescaled process of maxima,

$$\bigvee_{i=1}^{n} \frac{Y_i(s) - b_n(s)}{a_n(s)} \xrightarrow{\mathcal{D}} X(s), \quad n \to \infty, \tag{2.1}$$

where $\bigvee$ and $\xrightarrow{\mathcal{D}}$ denote, resp., the max-operator and convergence in distribution. The limiting random process X is assumed to be non-degenerate. From [18, 20] the class of the limiting processes $X(s)$ is the class of MS processes. This definition of MS processes offers a natural choice for modeling spatial extremes. Univariate extreme value theory (EVT) implies that the marginal distributions of $X(s)$ are Generalized Extreme Value (GEV) distributed, i.e.,

$$GEV_{\mu(s),\sigma(s),\xi s)}(x) := \mathbb{P}(X(s) \leq x) = \exp\left\{ -\left(1 + \xi(s)\frac{x - \mu(s)}{\sigma(s)}\right)^{-1/\xi(s)} \right\}, \quad 1 + \xi(s)\frac{x - \mu(s)}{\sigma(s)} > 0, \tag{2.2}$$

for some location $\mu(s) \in \mathbb{R}$, scale $\sigma(s) > 0$, and shape $\xi(s) \in \mathbb{R}$.
MS processes with unit Fréchet margins (ie with distribution function $F(x) = \exp\{-x^{-1}\}$, $x > 0$) are called *simple MS processes* (corresponding to $\mu(s) = 0$, $\sigma(s) = 1$ and $\xi(s) = 1$). They have the following very useful representation (see [18, 37]):

$$X(s) = \max_{k \geq 1} Q_k(s)/P_k, \qquad s \in \mathcal{S}, \tag{2.3}$$

where $Q_k(s)$ are independent replicates of a non-negative stochastic process $Q(s)$ with unit mean at each s, and P_k are points of a unit rate Poisson process on $\mathbb{R}^+$. Given any MS process, margins can be transformed to Fréchet margins. In what follows, we consider only simple MS processes. With this choice the normalizing functions are $a_n = n$ and $b_n = 0$. For a detailed overview on EVT, we refer the reader to [9, 3, 19].
Different choices for the process $Q(s)$ in (2.3) lead to more or less flexible models for spatial maxima. Commonly used models are the Gaussian extreme value model (Smith model) [40], the extremal Gaussian model (Schlather model) [37], the truncated extremal Gaussian model (TEG) originally due to [37] and has been exemplified by [16, 2] (resp. [24]) in the spatial (resp. spatiotemporal) context, the Brown-Resnick model (BR) [26], and the extremal-t model [32].

For $K \in \mathbb{N} \setminus \{0\}$, $s_1, \ldots, s_K \in \mathcal{S}$, and $x_1, \ldots, x_K > 0$, the finite K-dimensional distributions of the process X owing to the representation (2.3) may be written as

$$-\log \mathbb{P}(X(s_1) \leq x_1, \ldots, X(s_K) \leq x_K) = \mathbb{E}\left[\bigvee_{k=1}^{K} \left\{\frac{Q(s_k)}{x_k}\right\}\right] \equiv V^X_{s_1,\ldots,s_K}(x_1, \ldots, x_K), \tag{2.4}$$

where $V^X_{s_1,\ldots,s_K}$ is called the exponent function. It summarizes the extremal dependence structure, and is homogeneous of order -1, i.e.,$V^X_{s_1,\ldots,s_K}(tx_1, \ldots, tx_K) = t^{-1}V^X_{s_1,\ldots,s_K}(x_1, \ldots, x_K)$ for any $t > 0$. Moreover, it satisfies $V^X_{s_1,\ldots,s_K}(\infty, \ldots, x_k, \ldots, \infty) = 1/x_k$ for each $k = 1, \ldots, K$. The coefficient defined as

$$-x \log \mathbb{P}\{X(s_1) \leq x, \ldots, X(s_K) \leq x\} = V^X_{s_1,\ldots,s_K}(1, \ldots, 1) \equiv \theta_{s_1,\ldots,s_K},$$

is known as the extremal coefficient. Complete dependence (resp. complete independence) is achieved when $\theta_{s_1,\ldots,s_K} = 1$ (resp. $\theta_{s_1,\ldots,s_K} = K$). Roughly speaking, $\theta_{s_1,\ldots,s_K}$ corresponds to the effective numbers of independent maxima among K [38]. Due to computational issues, the K dimensional distributions are not easily tracktable. This is why the pairwise distributions are used. The pairwise extremal coefficient $\theta : \mathbb{R}^+ \mapsto [1, 2]$ satisfies

$$\mathbb{P}[\max\{X(s), X(s+h)\} \leq x] = \exp\left(-1/x\right)^{V^X_h(1,1)} := F(x)^{\theta(h)}, \quad x > 0, \tag{2.5}$$

for some spatial lag vector h. In Table 1, we give a brief summary of five well-known MS processes by their bivariate exponent function V^X_h and extremal coefficient θ. Indeed, Smith (resp. Schlather) process is a special case of BR (resp. extremal-t) when $2\gamma(h) = \sqrt{h^T\Sigma^{-1}h}$, for some covariance matrix Σ (resp. degrees of freedom $v = 1$).

Model	Exponent function $V^X_h(x_1, x_2)$	Extremal coefficient $\theta(h)$
Smith	$\frac{1}{x_1}\Phi\left(\frac{\delta(h)}{2} + \frac{\log\left(\frac{x_2}{x_1}\right)}{\delta(h)}\right) + \frac{1}{x_2}\Phi\left(\frac{\delta(h)}{2} + \frac{\log\left(\frac{x_1}{x_2}\right)}{\delta(h)}\right)$ $\delta(h) = \sqrt{h^T\Sigma^{-1}h}$ and $\Phi(.)$ denotes the standard normal distribution.	$2\Phi\left(\frac{\delta(h)}{2}\right)$
Schlather	$\frac{1}{2}\left(\frac{1}{x_1} + \frac{1}{x_2}\right)\left[1 + \sqrt{1 - \frac{2(\rho(h)+1)x_1x_2}{(x_1+x_2)^2}}\right]$ $\rho(h)$ is the correlation coeffient.	$1 + \left(\frac{1-\rho(h)}{2}\right)^{1/2}$
TEG	$\left(\frac{1}{x_1} + \frac{1}{x_2}\right)\left[1 - \frac{\alpha(h)}{2}\left(1 - \sqrt{1 - \frac{2(\rho(h)+1)x_1x_2}{(x_1+x_2)^2}}\right)\right]$ $\alpha(h) = \mathbb{E}[\lvert\mathcal{A} \cap \{h + \mathcal{A}\}\rvert]/\mathbb{E}(\lvert\mathcal{A}\rvert)$, $\lvert.\rvert$: volume of the random set $\mathcal{A}$. If $\mathcal{A}$ is a disk of fixed radius r, $\alpha(h) \simeq (1 - h/2r)\mathbb{1}_{[0,2r]}$.	$2 - \alpha(h)\left[1 - \left(\frac{1-\rho(h)}{2}\right)^{1/2}\right]$
BR	$\frac{1}{x_1}\Phi\left(\sqrt{\frac{\gamma(h)}{2}} + \frac{\log\left(\frac{x_2}{x_1}\right)}{\sqrt{2\gamma(h)}}\right) + \frac{1}{x_2}\Phi\left(\sqrt{\frac{\gamma(h)}{2}} + \frac{\log\left(\frac{x_1}{x_2}\right)}{\sqrt{2\gamma(h)}}\right)$ $\gamma(.)$ denotes the semivariogram.	$2\Phi\left[\left(\frac{\gamma(h)}{2}\right)^{1/2}\right]$
Extremal-t	$\frac{1}{x_1}T_{v+1}\left(-b\rho(h) + b\left(\frac{x_2}{x_1}\right)^{1/v}\right) + \frac{1}{x_2}T_{v+1}\left(-b\rho(h) + b\left(\frac{x_1}{x_2}\right)^{1/v}\right)$ T_v is the cumulative distribution function of a Student random variable with $v \geq 1$ degrees of freedom. $b^2 = \{v+1\}/\{1-\rho^2(h)\}$	$2T_{v+1}\left(\sqrt{(v+1)\frac{1-\rho(h)}{1+\rho(h)}}\right)$

Table 1: Bivariate marginal probability distributions for MS models

The stochastic process $X(s)$ defined in (2.3) has the bivariate density function

$$f_h^X(x_1,x_2) = \exp\{-V_h^X(x_1,x_2)\}\left\{\partial_1 V_h^X(x_1,x_2)\partial_2 V_h^X(x_1,x_2) - \partial_{12}^2 V_h^X(x_1,x_2)\right\},$$

where $\partial_i := \frac{\partial}{\partial x_i}$, $\partial_{12}^2 := \frac{\partial^2}{\partial x_1 \partial x_2}$, and V_h^X is the exponent measure of the MS process $X(s)$.

2.2. Pairwise extremal dependence summary measures

In addition to the extremal coefficient θ described in Section 2.1., we recall here some measures/functions that may be used to describe the extremal dependence behavior of spatial processes.
[10] suggested two model-free diagnostic measures; $\chi_u(h)$ and $\bar{\chi}_u(h)$ in order to identify different types of tail dependence. For a stationary spatial process $\{X(s)\}_{s\in\mathcal{S}}$ with univariate margin F, AD is characterized by $\chi(h) > 0$, with

$$\chi(h) = \lim_{u\to 1^-} \mathbb{P}\{F(X(s)) > u | F(X(s+h)) > u\},\ s, s+h \in \mathcal{S}. \tag{2.6}$$

This means that, for an AD process like MS process, a large event at location $s+h$ leads to a non-zero probability of a similarly large event at location s for some spatial lag vector h. On the other hand, a process is AI if $\chi(h) = 0$ for any h. The best-known AI example is the Gaussian model for which $\rho(h) \neq 1$. This means that the dependence strength between events observed at two distinct spatial locations vanishes as their extremeness increases, see [39]. It is known that MS processes ar AD (or independent). This coefficient is related to the pairwise extremal coefficient θ of a MS process through the relation $\chi(h) = 2 - \theta(h)$. Equivalently, for $u \in [0,1]$

$$\chi_u(h) = 2 - \frac{\log \mathbb{P}\{F(X(s)) < u, F(X(s+h)) < u\}}{\log \mathbb{P}\{F(X(s+h)) < u\}} \quad \text{and} \quad \chi(h) = \lim_{u\to 1^-} \chi_u(h). \tag{2.7}$$

The function $\chi_u(h)$ can be viewed as a measure of the dependence at the u-quantile level. Both dependence functions θ and χ provide simple measures of extremal dependence within the class of AD distributions. An alternative dependence coefficient which measures the strength of dependence for AI processes

$$\bar{\chi}_u(h) = \frac{2\log \mathbb{P}(F(X(s)) > u)}{\log \mathbb{P}(F(X(s)) > u, F(X(s+h)) > u)} - 1 \quad \text{and} \quad \bar{\chi}(h) = \lim_{u\to 1^-} \bar{\chi}_u(h). \tag{2.8}$$

AD (resp. AI) is achieved if $\bar{\chi}(h) = 1$ (resp. $\bar{\chi}(h) < 1$). Moreover, $\bar{\chi}_u(h) \in (0,1)$ (resp. $(-1,0)$) implies positive (resp. negative) association at distance h. Hence $|\bar{\chi}_u(h)|$ usually increases with the dependence. So, in practice, the two indicators χ and $\bar{\chi}$ should be considered together. Nevertheless, inference based on these measures is difficult because few observations are available as u approaches 1, see e.g., [5].
On one other hand, the coefficient of tail dependence $\eta(h)$ has been introduced by [29]. It measures the strength of extremal dependence within the class of AI models. This can be seen by looking at the bivariate joint tail and the conditional upper tail that behave resp. as

$$\mathbb{P}(X(s) > x, X(s+h) > x) \sim \mathcal{L}_h(x) x^{-1/\eta(h)},\ x \to \infty, \tag{2.9}$$

$$\mathbb{P}(X(s) > x | X(s+h) > x) \sim \mathcal{L}_h(x) x^{1-1/\eta(h)},\ x \to \infty, \tag{2.10}$$

where $\mathcal{L}_h(.)$ is a slowly varying function at ∞; that is, $\lim_{y\to\infty}\frac{\mathcal{L}_h(yt)}{\mathcal{L}_h(y)}=1, t>0$, and $\eta(h)\in(0,1]$. Here, AI corresponds to $\eta(h)<1$. $\eta(h)\in\left(\frac{1}{2},1\right)$ $\left(\text{resp. }\left(0,\frac{1}{2}\right)\right)$ implies a negative (resp. positive) association at distance h. In particular, non independent simple MS processes have $\eta(h)=1$ for all h. Moreover, under condition (2.9), one can easily deduce that $\bar{\chi}(h)=2\eta(h)-1$.

2.3. Hybrid models of spatial extremal dependence

For any stationary process X, we denote by $F_h^X(x_1,x_2)$ the bivariate distribution function of the pair $(X(s),X(s+h))$.

Processes derived from Gaussian processes by a non decreasing transformation of the marginals are examples of AI processes. Alternatively, the class of inverted MS processes has been proposed by [46]. It were found to be more flexible than Gaussian derivated processes in some applications [41, 15]. Let $\{X'(s)\}_{s\in\mathcal{S}}$ be a MS process and consider

$$Y(s)=-1/\log\left[1-\exp\left\{-1/X'(s)\right\}\right],\ s\in\mathcal{S}. \tag{2.11}$$

$(Y(s)_{s\in\mathcal{S}})$ is called an *inverted MS process*. For these processes, $\eta(h)=1/\theta(h)$, where $\theta(h)$ is the bivariate extremal coefficient of the MS process $X'(s)$. With this construction, any MS process $X'(s)$ may be transformed into an AI counterpart $Y(s)$. The bivariate cumulative distribution function (c.d.f.) of $Y(s)$ is given by

$$F_h^Y(x_1,x_2)=-1+\exp(-x_1^{-1})+\exp(-x_2^{-1})+\exp\{-V_h^{X'}[t(x_1),t(x_2)]\}, \tag{2.12}$$

where $V_h^{X'}$ is the exponent measure of the bivariate c.d.f. corresponding to $\{X'(s),X'(s+h)\}$ and the transformation $t(x):(0,\infty)\mapsto(0,\infty)$ is given by $t(x)=-1/\log[1-\exp\{-1/x\}]$.

In addition, [46] developed MM models by combining inverted MS and MS models, which add modeling flexibility to spatial extreme analysis and seem able to encompass different degrees of spatial extremal dependence. Let $\{X(s)\}_{s\in\mathcal{S}}$ be a simple MS process with a bivariate c.d.f., F_h^X, and $\{Y(s)\}_{s\in\mathcal{S}}$ be an inverted MS process with a bivariate c.d.f. F_h^Y. Assume X and Y are independent. Then the process

$$Z(s)=\max\{aX(s),(1-a)Y(s)\},\quad 0\le a\le 1, \tag{2.13}$$

is called a MM process. It has unit Fréchet margins. The bivariate c.d.f. is straightforwardly obtained for $0<a<1$.

$$F_h^Z(z_1,z_2)=F_h^X\left(\frac{z_1}{a},\frac{z_2}{a}\right)F_h^Y\left(\frac{z_1}{1-\ a},\frac{z_2}{1-\ a}\right). \tag{2.14}$$

Furthermore, the bivariate conditional upper tail probability has the form (see, [46, 2])

$$\mathbb{P}(Z(s)>z|Z(s+h)>z)\sim a\left\{2-\theta(h)\right\}+(1-a)^{1/\eta(h)}\frac{\mathcal{L}_h\left\{z/(1-a)\right\}}{z^{1/\eta(h)-1}},\ z\to\infty. \tag{2.15}$$

Accordingly, MS models ($a=1$), may be too restrictive in the sense that they have only the first-order term in (2.15), while AI models ($a=0$), may be unreliable since they are left with the second term only. Therefore, MM models seem to provide a good balance between the two classes.

Figure 1 displays four realizations of two MM models over the square $[0,10]^2$ according to different values of the mixing coefficient a. In order to show the role of the mixing coefficient, the plots

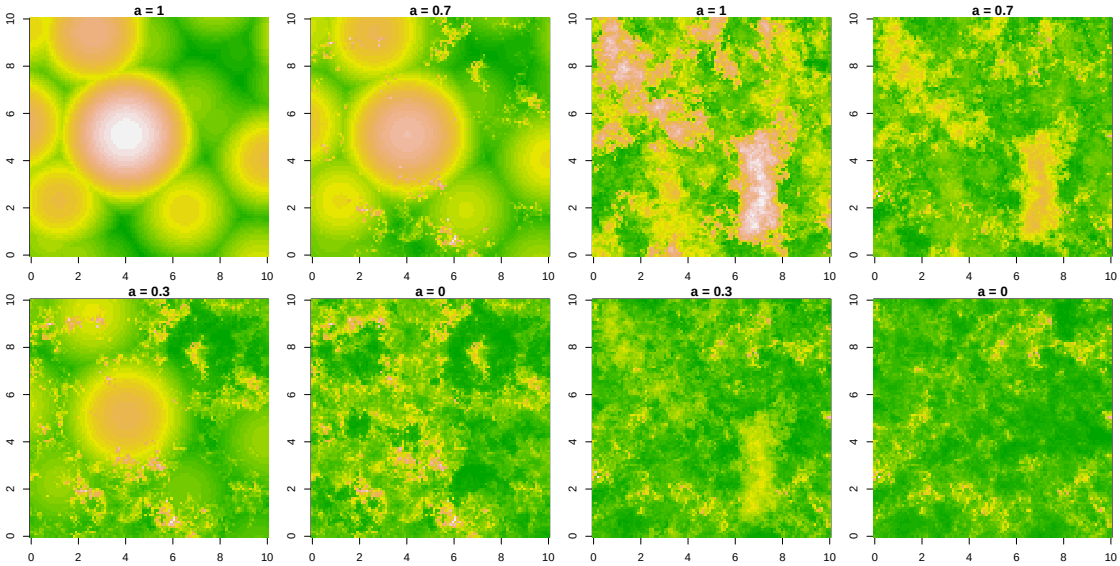


Figure 1: Simulations of the MM model (2.13) on the logarithm scale according to different values of the mixing coefficient $a \in \{1, 0.7, 0.3, 0\}$. Left panel: X is an isotropic Smith process with $\Sigma = \mathbf{I}_2$ (2×2 identity matrix), and Y is an inverted extremal-t process with $v = 2$ degrees of freedom and exponential correlation function $\rho(h) = \exp(-\|h\|)$. Right panel: X is an isotropic extremal-t process with $v_1 = 2$ degrees of freedom and exponential correlation function $\rho_X(h) = \exp(-\|h\|)$, and Y is an inverted extremal-t process with $v_2 = 5$ degrees of freedom and $\rho_Y(h) = \exp(-\|h\|)$.

are obtained by considering values between $a = 1$ (MS process) and $a = 0$ (inverted MS process). Clearly, we observe that a reveals the strength of AD part in the process Z. So that, this model extends classical dependence modeling within the AD class and is appropriate when AD is present at all distances because it allows to capture a second order in the dependence structure which is not possible with a MS model. The simulations have been carried out using the function `rmaxstab` of the `R` package SpatialExtremes [35].

3. INFERENCE FOR MM PROCESSES: CENSORED PAIRWISE LIKELIHOOD APPROACH

In order to propose a testing procedure on the mixing parameter a of a MM process Z (2.13), we use the composite likelihood for inference. Hence, in this section, we briefly describe the composite likelihood estimation procedure for the process Z. As the full likelihood is not generally known for MS models due to the tractability of distribution functions and densities for MS process models, composite likelihoods turn out to be attractive surrogates as the resulting estimator enjoys desirable asymptotic properties such as consistency and asymptotic normality, provided some regularity conditions are met [4, 30, 43, 44]. In particular, the pairwise likelihood estimation has been found useful to estimate parameters in a MS process, see e.g. [34, 16]. Assume that we observe the process Z at D locations $s_1, \ldots, s_D$ and T times $t_1, \ldots, t_T$, where the observations are assumed to be independent in time. We

denote by ϑ the vector gathering the parameters to be estimated using pairwise likelihood. Then the (weighted) pairwise log-likelihood is

$$p\ell(\vartheta) = \sum_{k=1}^{T} \sum_{i=1}^{D-1} \sum_{j>i}^{D} \omega_{ij} \log \mathcal{L}(z_{ik}, z_{jk}; \vartheta), \tag{3.1}$$

where z_{jk} is the kth observation of the jth site, $\mathcal{L}(z_{ik}, z_{jk}; \vartheta)$ is the likelihood of the pair (z_{ik}, z_{jk}), and ω_{ij} are non negative weights specifying the contribution of each pair. A careful choice of the weights ω_{ij} could improve the statistical efficiency as well as the computational one. A simple weighting choice is to let $\omega_{ij} = \mathbb{1}_{\{h \leq \delta\}}$ for some specified value δ, where $\mathbb{1}_{\{\mathcal{B}\}}$ represents the indicator function of $\mathcal{B}$. With this approach, the selection of the threshold δ is critical, see [22]. One possibility would be to consider δ as the q-quantile of the distribution of the distances between pairs of sites, $q \in [0, 1]$, see, e.g. [1, 2]. The idea beyond this approach is that it is expected that neighboring pairs are strongly dependent, thus providing valuable information for estimating the dependence structure.

In order to improve the estimation of the AI parameters of Z, a censored pairwise likelihood estimation approach has been developed by [46] and applied by [2] for MM processes. Let F_h^Z denotes the bivariate c.d.f. of Z (2.14) and $u \in \mathbb{R}$ is a high threshold; the censored pairwise likelihood contribution is defined as follows

$$\mathcal{L}_u(z_{ik}, z_{jk}; \vartheta) = \begin{cases} F_h^Z(u, u; \vartheta), & \text{if } \max(z_{ik}, z_{jk}) \leq u \\ \partial_{12}^2 F_h^Z(z_{ik}, z_{jk}; \vartheta), & \text{if } \max(z_{ik}, z_{jk}) > u. \end{cases} \tag{3.2}$$

Then the maximum censored pairwise likelihood estimator is given by $\widehat{\vartheta}_u = \text{argmax } p\ell_u(\vartheta)$, where $p\ell_u(.)$ is the (weighted) censored pairwise log-likelihood. Asymptotic properties of the maximum censored pairwise likelihood estimator $\widehat{\vartheta}_u$ are available from [46, 2, 24], which we now summarize. For large T and under some regularity conditions, $\widehat{\vartheta}_u$ is asymptotically Gaussian with asymptotic variance

$$\mathcal{G}^{-1}(\vartheta) = \mathcal{H}^{-1}(\vartheta)\mathcal{J}(\vartheta)\mathcal{H}^{-1}(\vartheta),$$

where $\mathcal{G}(\vartheta)$ is the Godambe information matrix, $\mathcal{H}(\vartheta) = \mathbb{E}\{-\nabla^2 p\ell_u(\vartheta)\}$ is called the sensitivity matrix, and $\mathcal{J}(\vartheta) = \mathbb{V}ar\{\nabla p\ell_u(\vartheta)\} = \mathbb{E}\{\nabla p\ell_u(\vartheta)\nabla^t p\ell_u(\vartheta)\}$ is called the variability matrix. Standard errors computation of $\widehat{\vartheta}_u$ requires the estimation of the Godambe matrix and its components. Analytical expressions for $\mathcal{H}(\vartheta)$ and $\mathcal{J}(\vartheta)$ are difficult to obtain in mostly realistic applications. Hence, a Monte Carlo procedure can be applied to estimate the asymptotic variance matrix,

$$\widehat{\mathcal{H}}(\vartheta) = -M^{-1} \sum_{k=1}^{M} \nabla^2 p\ell_u(\widehat{\vartheta}; z^k), \tag{3.3}$$

$$\widehat{\mathcal{J}}(\vartheta) = M^{-1} \sum_{k=1}^{M} \nabla p\ell_u(\widehat{\vartheta}; z^k) \nabla^t p\ell_u(\widehat{\vartheta}; z^k), \tag{3.4}$$

where z^k, $k = 1, ..., M$, denote the kth datasets simulated from the fitted model, see for e.g., [44, 7] and the references therein. Analogously, the above-mentioned asymptotic properties of $\widehat{\vartheta}_u$ are also satisfied for the maximum pairwise likelihood estimator $\widehat{\vartheta} = \text{argmax } p\ell(\vartheta)$, see [27, 30, 11].

Remark 3.1 *Different censoring approaches of the pairwise likelihood have been introduced. In addition to (3.2), the censored pairwise likelihood contribution* $\mathcal{L}_u(z_{ik}, z_{jk}; \vartheta)$ *of a pair* (z_{ik}, z_{jk}) *can be taken as follows*

$$\mathcal{L}_u(z_{ik}, z_{jk}; \vartheta) = \begin{cases} F_h^Z(u, u; \vartheta), & \text{if } max\ (z_{ik}, z_{jk}) \leq u \\ \partial_1 F_h^Z(z_{ik}, u; \vartheta), & \text{if } z_{ik} > u, z_{jk} \leq u \\ \partial_2 F_h^Z(u, z_{jk}; \vartheta), & \text{if } z_{ik} \leq u, z_{jk} > u \\ \partial_{12}^2 F_h^Z(z_{ik}, z_{jk}; \vartheta), & \text{if } min(z_{ik}, z_{jk}) > u. \end{cases} \tag{3.5}$$

This approach has proved to be useful for the statistical inference of spatial extremes, see, e.g. [24, 47, 42, 25].

Similarly to the weights ω_{ij}, the choice of threshold u is crucial. A common choice is to set u corresponding to the empirical q-quantile at each site, provided that q is sufficiently large, see, e.g. [24, 5, 2].

4. CENSORED PAIRWISE LIKELIHOOD STATISTICS FOR TESTING $H_0 : a = a_0$ VERSUS $H_1 : a \neq a_0$

To compare the performance of nested models, a composite likelihood ratio test can be performed. Suppose that the parameters of a MM model $\vartheta \in \mathbb{R}^q$ is partitioned as $\vartheta = (\vartheta^1, \vartheta^2) \in \mathbb{R}^{q_1} \times \mathbb{R}^{q_2}$, with $q_1 + q_2 = q$, and that we want to test whether the null hypothesis $H_0 : \vartheta^1 = \vartheta^{1*}$ holds. In this testing framework, the parameter $\vartheta^1 \in \mathbb{R}^{q_1}$ is the parameter of interest, while $\vartheta^2 \in \mathbb{R}^{q_2}$ acts as a nuisance parameter. Let $\widehat{\vartheta}_u = (\widehat{\vartheta}_u^1, \widehat{\vartheta}_u^2)$ denotes the unrestricted maximum censored pairwise likelihood estimator, and $\widehat{\vartheta}_u^* = (\vartheta_u^{1*}, \widehat{\vartheta}_u^{2*})$ denotes the maximum censored pairwise likelihood estimator under the null hypothesis, i.e., $\widehat{\vartheta}_u^{2*}$ is the maximum censored pairwise likelihood estimator of ϑ^2 when ϑ^1 is held fixed to the value ϑ^{1*}. A two-sided censored pairwise likelihood ratio test may be based on the statistic (see, e.g. [27, 44] and the references therein)

$$LR = 2\{p\ell_u(\widehat{\vartheta}_u) - p\ell_u(\widehat{\vartheta}_u^*\} \xrightarrow{\mathcal{D}} \sum_{j=1}^{q_1} c_j W_j, \tag{4.1}$$

where the W_j's are independent χ_1^2 random variables, and the c_j's are the eigenvalues of the matrix $\{\mathcal{H}^1\}^{-1}\mathcal{G}_1$ evaluated under the null hypothesis, where $\mathcal{H}^1(\vartheta)$ and $\mathcal{G}_1(\vartheta)$ denote, resp., the $q_1 \times q_1$ submatrices of $\{\mathcal{H}(\vartheta)\}^{-1}$ and $\mathcal{G}(\vartheta)$ with elements corresponding to ϑ^1. Various adjustments have been proposed in the literature to recover an asymptotic chi-squared distribution $\chi_{q_1}^2$ when $q_1 > 1$. For a detailed description on these adjustments, we refer the reader to see, e.g., [8, 21, 28, 33, 36]. Nevertheless, simulation-based techniques could be used to calculate the quantiles of the limit $\sum_{j=1}^{q_1} c_j W_j$.

In the present paper, our purpose is to test the hypothesis $H_0 : a = a_0$ versus $H_1 : a \neq a_0$, for some specified value $a_0 \in$[0,1]. Therefore, we propose the following two statistics exploiting the maximum pairwise likelihood as an inferential tool.

- The pairwise likelihood ratio statistic, which can be easily deduced from (4.1) with $\vartheta^1 = a$ and $q_1 = 1$,

$$LR_a = c^{-1} LR \xrightarrow{\mathcal{D}} \chi^2_1, \tag{4.2}$$

where the constant c is computed by the same manner described above, in which the matrices for this special case are of dimensions 1×1.

- The Z-test statistic which is straightforwardly derived from the central limit theorem (CLT) for maximum composite likelihood estimators.

$$Z_a = \frac{\hat{a} - a}{\sqrt{\mathcal{G}^{aa}(\widehat{\vartheta}_u)}} \xrightarrow{\mathcal{D}} N\{0, 1\}, \tag{4.3}$$

where $\mathcal{G}^{aa}(\widehat{\vartheta}_u)$ denotes a 1×1 submatrix of the inverse of $\mathcal{G}(\widehat{\vartheta}_u)$ pertaining to a.

Remark 4.1 *Traditional methods for deriving hypothesis tests at boundary points $H_0 : a = 1$ (AD) or $H_0 : a = 0$ (AI) do not work in this situation due to the presence of additional nuisance parameters which are not identified under the null hypotheses. This problem has been introduced in several studies for various situations, see, e.g. [13, 14]. In that case, still, much more theoretical research has to be undertaken to determine the limiting distribution. Therefore, we apply our LR_a test at points close to the boundaries, i.e., $a_0 = 0.99$ or $a_0 = 0.01$.*

5. SIMULATION STUDY

This section studies the performance of our proposed test statistics LR_a and Z_a via several simulation studies. We consider testing $H_0 : a = a_0$ against $H_1 : a \neq a_0$, where a_0 varies from 0.01 to 0.99 by steps of 0.01.
Throughout this section, we consider the following MM model

- **MM**: is a MM model (2.13) in which X is a TEG process (see Table 1) with $\mathcal{A}_X$ a disk of fixed radius r_X and isotopic exponential correlation function $\rho_X(h) = \exp\{-\|h\|/\phi_X\}$, for some range parameter $\phi_X > 0$. The AI process Y is an inverted TEG process with $\mathcal{A}_Y$ a disk of fixed radius r_Y and isotopic correlation function $\rho_Y(h) = \exp\{-\|h\|/\phi_Y\}, \phi_Y > 0$. Note that we call range the distance beyond which it is considered the correlation cancels out.

By this construction of the model **MM**, the pairs of sites separated by a distance $\|h\|$ smaller than $2r_X$ or greater than $2r_Y$ are asymptotically dependent or exactly independent, respectively. Whereas, at intermediate distances the pairs exhibit AI.
Note that for the MS TEG process X (see Table 1), the short-range dependence is largely determined by the correlation function $\rho(h)$, while the longer-range dependence is regulated by the geometry of the random set $\mathcal{A}$. For the sake of simplicity, we may consider $\mathcal{A}$ as a disk with fixed radius r. With this choice the expected volume of overlap between the random set $\mathcal{A}$ and $\mathcal{A}+h$ can be approximated by $\alpha(h) \simeq (1 - \|h\|/2r)\mathbb{1}_{[0,2r]}$. In such a case, $\chi(h) = 0,\ \|h\| \geq 2r$. In other words, the process X is exactly independent for all $\|h\| \geq 2r$. For more details, see [16].

5.1. Estimation performance

The censored pairwise likelihood approach (3.2) is used for estimation, where the threshold u is taken corresponding to the 0.9 empirical quantile at each site. Equal weights ω_{ij} are considered. Indeed, choosing equal weights in the pairwise likelihood may not be ideal both in terms of computational efficiency and statistical efficiency, see [6]. In order to obtain the estimates of $\mathcal{H}(\vartheta)$ and $\mathcal{J}(\vartheta)$ in the Monte Carlo procedure (3.3) and (3.4), we use $M = 1500$ as a compromise between accuracy and computation time. Furthermore, to reduce computational burden the pairwise likelihood function has been coded in C; the optimization has been parallelized on 20 cores using the R library parallel and carried out using the Nelder-Mead optimization routine in R.

To assess the quality of the censored pairwise likelihood estimation procedure, a simulation study has been carried out. We simulate $T = 1000$ independent copies of the model **MM** at $D = 50$ sites randomly and uniformly distributed in the square $[0, 3]^2$. We consider several mixing parameters $a \in \{0, 0.25, 0, 50, 0.75 \text{ and } 1\}$. The parameters used are $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$. Each experiment was repeated $J = 200$ times to obtain boxplots of the estimated parameters and compute performance metrics, i.e., the mean estimate, the root mean squared error (RMSE), and the mean absolute error (MAE). Denote by $\widehat{\vartheta}_j$ the jth estimation, then

$$\text{RMSE} = \left[J^{-1} \sum_{j=1}^{J} (\widehat{\vartheta}_j - \vartheta)^2 \right]^{1/2} \quad \text{and} \quad \text{MAE} = J^{-1} \sum_{j=1}^{J} |\widehat{\vartheta}_j - \vartheta| \tag{5.1}$$

The boxplots of the errors of the estimated parameters on the J samples are displayed in Figure 2. Table 2 reports the mean estimate, RMSE, and MAE of the estimated parameters. Generally, the estimation procedure appears to work well although the variability of some estimates were relatively large, especially for the AI parameters $\{\phi_Y, r_Y\}$. This probably stems from the fact that asymptotic independence is difficult to estimate [15]. Moreover, we observe that contrary to AI parameters $\{\phi_Y, r_Y\}$, the estimation of AD parameters $\{\phi_X, r_X\}$ becomes more accurate as the mixing parameter value increases (the RMSE and MAE are lower). Using (3.5) instead of (3.2) as censored pairwise likelihood leads to similar results (see Appendix A).

5.2. Testing performance with true a being a non-boundary point in $(0, 1)$

We evaluate the performance of the proposed pairwise likelihood test statistics (4.2) and (4.3). We test whether the null hypothesis $H_0 : a = a_0$ holds for all a_0 values in the set $\{0.01, 0.02, \ldots, 0.99\}$. Here we examine three cases with a true mixing parameter $a \in \{0.25, 0.5, 0.75\}$, based on $J = 150$ simulation replicates from $T = 1000$ independent copies simulated at $D = 50$ randomly and uniformaly sampled locations in the square $[0, 3]^2$ from the model **MM** with parameters $\{\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, \text{and } r_Y = 1.2\}$. We compute the empirical probabilities of rejecting H_0 which we denote by P. In other words, P represents the power of the test when H_0 is false (i.e. the proportion of null hypotheses rejected). Decisions obtained at three significance levels $\alpha \in \{0.01, 0.05, 0.10\}$.

For both statistics LR_a and Z_a, our results are summarized in Figure 3 and Table 3. Overall, the results show a reasonable performance for the two statistics. In particular, both tests seem to be unbiaised (the power is greater than the sensitivity level α). Moreover, the type I errors (probability

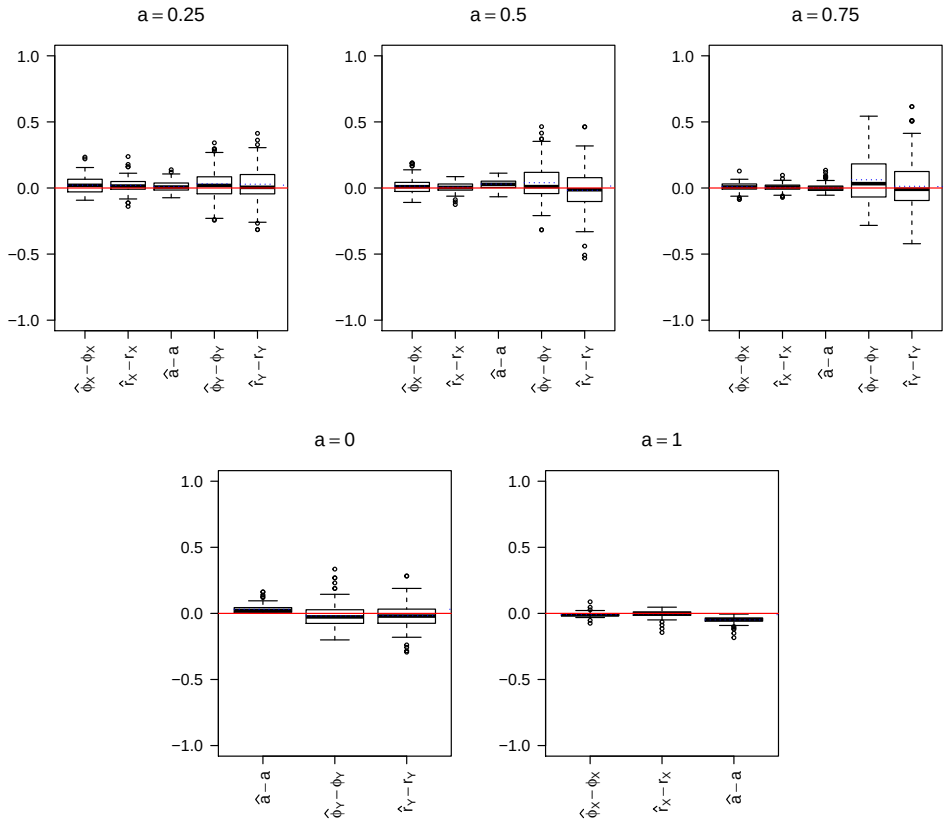


Figure 2: Boxplots display $\widehat{\vartheta} - \vartheta$ of the estimated parameters using censored pairwise likelihood approach (3.2) based on 200 simulation replicates from 1000 independent copies of the model **MM** with parameters $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$, and mixture parameter $a \in \{0, 0.25, 0.5, 0.75, 1\}$. Blue dotted/red horizontal lines show the estimated errors means/0 value.

of rejecting H_0 when H_0 holds true) are close to the nominal level size $\alpha = 0.01, 0.05, 0.1$. The LR_a test seems to be more powerful than the Z_a one as can be seen on Table 3.

5.3. Testing performance with true a being a boundary point (i.e., $a = 0$ or $a = 1$)

The censored pairwise likelihood ratio test statistics LR_a cannot be applied directly for testing $H_0 : a = a_0$ with a_0 being the boundary point (i.e., $a_0 = 0$ or $a_0 = 1$) since there are additional nuisance parameters which are present only under the alternative hypothesis H_1. Thus, we apply our statistics at some points very close to the boundaries, i.e., $a_0 = 0.01$ or $a_0 = 0.99$.

For this purpose, we simulate $T = 1000$ independent observations of a TEG process (AD case)(resp. an inverted TEG process (AI case)) with parameters $\{\phi_X = 0.10, r_X = 0.25\}$ (resp. $\{\phi_Y = 0.75, r_Y = 1.2\}$) at $D = 50$ sites uniformly generated in $[0, 3]^2$. We repeat this experiment $J = 150$ times. For both cases, we test whether the null hypothesis $H_0 : a = a_0$ holds, where a_0 varies from 0.01 to 0.99 by steps of 0.01. Here, the MM model that mixes both processes (TEG and inverted TEG) is used to

True	Performance metrics		
	Mean estimate	RMSE	MAE
$a = 0$	0.031	0.043	0.031
$\phi_Y = 0.75$	0.731	0.092	0.072
$r_Y = 1.2$	1.183	0.095	0.073
$\phi_X = 0.1$	0.121	0.064	0.053
$r_X = 0.25$	0.269	0.053	0.039
$a = 0.25$	0.261	0.037	0.029
$\phi_Y = 0.75$	0.782	0.115	0.088
$r_Y = 1.2$	1.229	0.125	0.091
$\phi_X = 0.1$	0.115	0.058	0.044
$r_X = 0.25$	0.257	0.036	0.028
$a = 0.5$	0.529	0.043	0.032
$\phi_Y = 0.75$	0.789	0.143	0.105
$r_Y = 1.2$	1.183	0.168	0.126
$\phi_X = 0.1$	0.111	0.030	0.023
$r_X = 0.25$	0.255	0.027	0.021
$a = 0.75$	0.751	0.029	0.020
$\phi_Y = 0.75$	0.812	0.186	0.144
$r_Y = 1.2$	1.213	0.203	0.153
$\phi_X = 0.1$	0.093	0.018	0.014
$r_X = 0.25$	0.231	0.024	0.017
$a = 1$	0.951	0.053	0.049

Table 2: Performance of the estimation for 200 simulated **MM** models with parameters $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$, and several mixing coefficient $a \in \{0, 0.25, 0.5, 0.75, 1\}$. The mean estimate, RMSE, and MAE of the estimated paramters.

perform testing for each step on the basis of the simulated data from TEG (AD case) and inverted TEG (AI case). Similarly as before, the empirical probabilities (P) are computed for both cases. Decisions are obtained at three significance levels $\alpha \in \{0.01, 0.05, 0.10\}$.

Figure 4 and Table 4 compare the empirical probabilities (P) obtained for both statistics LR_a and Z_a. We note that the performance of the two statistics is satisfactory. As expected, the power to reject asymptotic dependence, i.e., $H_0 : a = 1$ (resp. asymptotic independence, i.e., $H_0 : a = 0$) improves as $a \to 0$ (resp. $a \to 1$). Moreover, we observe that the P values at a_0 close to the boundaries, i.e., $a_0 = 0.01$ or $a_0 = 0.99$ are close to the nominal level size $\alpha = 0.01, 0.05, 0.1$. Therefore, these pairwise likelihood statistics can control the type I error rate α. So, these statistics can provide a strong indication for both AD and AI. Let us remark that near the boudaries ($a = 0$ or $a = 1$), the LR_a test seems to be less powerful than the Z_a one (contrary to what we observed for a away from the boundaries).

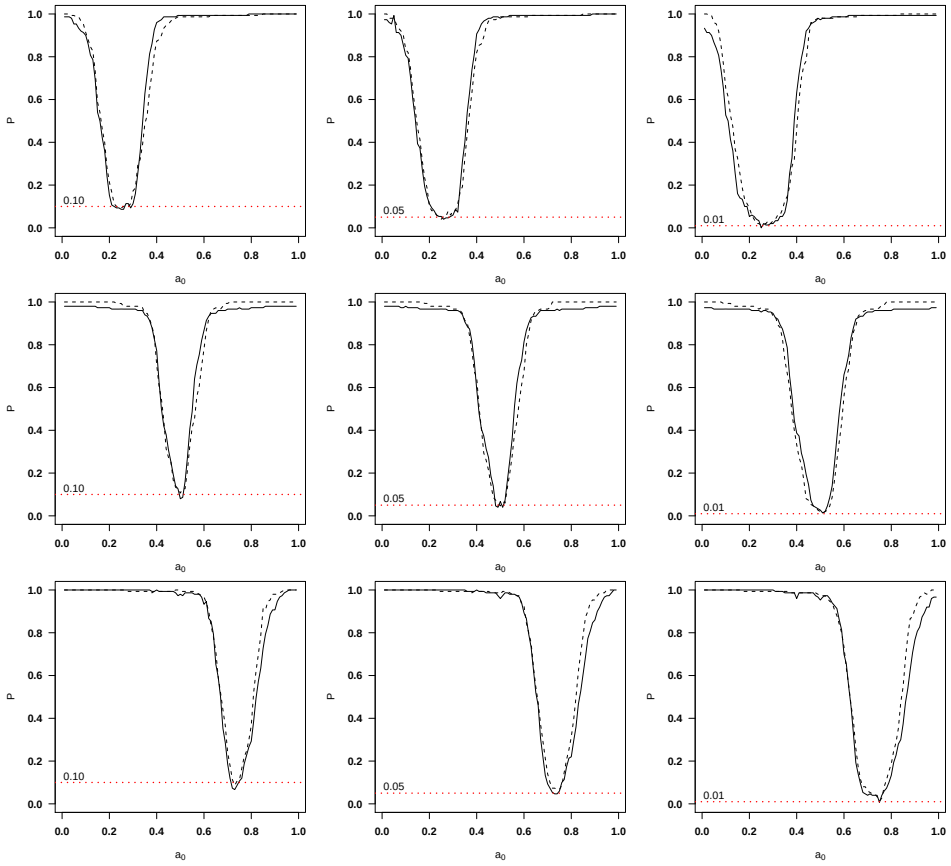


Figure 3: Empirical probabilities (P) based on 150 replicates simulation study of the model **MM** with $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$. The mixing coefficients (top row: $a = 0.25$, middle row: $a = 0.50$ and bottom row: $a = 0.75$). (Solid line: Z_a test and dashed line: LR_a test). Horizontal red dotted lines show the nominal level size $\alpha = 0.1, 0.05, 0.01$.

6. RAINFALL DATA EXAMPLE

Below we discuss a spatial application to illustrate the practical utility of our testing approach. We analyze daily precipitation data collected over the East of Australia.

6.1. Data

The dataset analyzed in this section is composed of daily rainfall measurements (in millimetres) recorded from 1972 to 2014 at 38 monitoring stations on the East of Australia whose locations are shown in Figure 5. Only the winter period (April-September) is considered. The altitude of the stations varying from 4 to 552 meters above mean sea level. The stations are separated by distances ranging from 34 km to 1383 km. This dataset is freely available on http://www.bom.gov.au/climate/data/. More details about data and monitoring stations are in the Supplementary Material, see Appendix B.

	True $a = 0.25$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.05	0.960	0.867	0.980	0.993	0.993	0.953
0.10	0.653	0.527	0.833	0.807	0.913	0.900
0.15	0.327	0.160	0.467	0.393	0.587	0.540
0.20	0.093	0.053	0.100	0.113	0.213	0.153
0.25	0.013	0.000	0.040	0.053	0.093	0.087
0.30	0.033	0.020	0.080	0.060	0.173	0.113
0.35	0.127	0.073	0.327	0.400	0.487	0.627
0.40	0.460	0.633	0.820	0.907	0.873	0.960
0.50	0.987	0.973	0.987	0.980	0.987	0.993
0.80	0.993	0.993	0.993	0.993	0.993	1.000
	True $a = 0.5$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.10	0.993	0.967	1.000	0.980	1.000	0.980
0.25	0.973	0.953	0.980	0.967	0.993	0.967
0.35	0.727	0.833	0.913	0.907	0.947	0.940
0.40	0.333	0.387	0.647	0.593	0.707	0.773
0.45	0.073	0.147	0.213	0.273	0.273	0.307
0.50	0.020	0.027	0.047	0.067	0.107	0.080
0.55	0.087	0.127	0.280	0.360	0.413	0.487
0.60	0.560	0.660	0.673	0.813	0.780	0.867
0.65	0.887	0.920	0.960	0.933	0.973	0.947
0.75	0.973	0.960	1.000	0.960	1.000	0.973
	True $a = 0.75$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.40	0.987	0.960	0.993	1.000	0.993	1.000
0.50	0.973	0.953	0.980	0.967	0.993	0.980
0.60	0.733	0.700	0.927	0.893	0.967	0.933
0.65	0.260	0.193	0.547	0.520	0.713	0.653
0.70	0.060	0.040	0.140	0.093	0.240	0.187
0.75	0.013	0.007	0.067	0.060	0.113	0.107
0.80	0.187	0.133	0.313	0.220	0.367	0.293
0.85	0.547	0.373	0.787	0.607	0.913	0.680
0.90	0.913	0.767	0.953	0.860	0.980	0.907
0.95	0.980	0.920	1.000	0.973	1.000	0.993

Table 3: Empirical probabilities (P) for testing $H_0 : a = a_0$ against $H_1 : a \neq a_0$ based on 150 simulation replicates from 1000 independent copies of the model **MM** with parameters $\phi_X = 0.10, , r_X = 0.25, \phi_Y = 0.75, r_Y = 1.2$ and mixing coefficients $a \in \{0.25, 0.5, 0.75\}$ at three significance levels $\alpha \in \{0.01, 0.05, 0.10\}$.

6.2. Exploratory analysis

For given data, it might be important to check whether the extremes in space have directional dependence. So, a graphical test based on the empirical versions of $\chi_u(h)$ (2.7) and $\bar{\chi}_u(h)$ (2.8) is used to

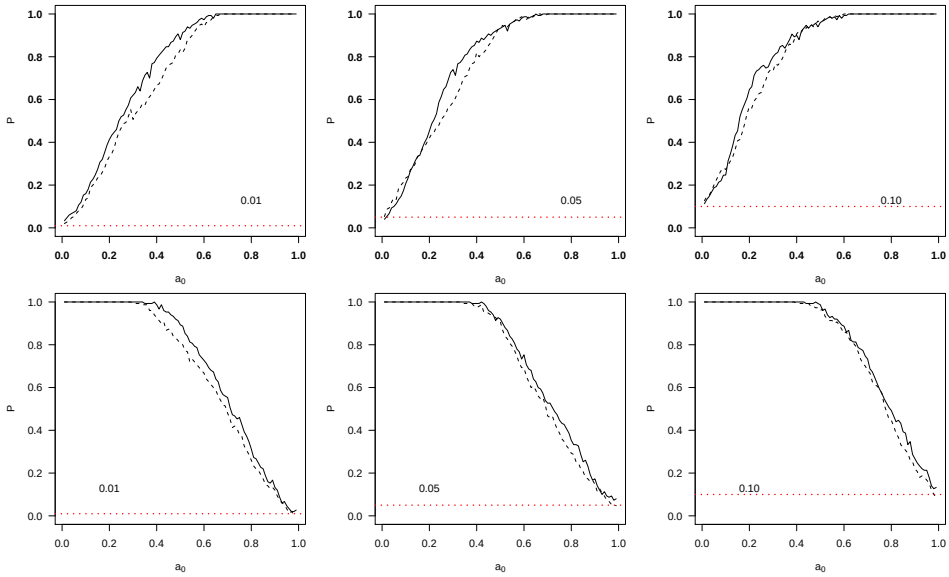


Figure 4: Top row: empirical probabilities (P) for testing $H_0 : a = 0$ (AI case) based on 150 simulation replicates from 1000 independent copies from an inverted TEG process with parameters $\{\phi_Y = 0.75, r_Y = 1.2\}$). Bottom row: empirical probabilities (P) for testing $H_0 : a = 1$ (AD case) based on 150 simulation replicates from 1000 independent copies from a TEG process with parameters $\{\phi_X = 0.1, r_X = 0.25\}$). (Solid line: Z_a test and dashed line: LR_a test). Horizontal red dotted lines show the nominal level size $\alpha = 0.1, 0.05, 0.01$.

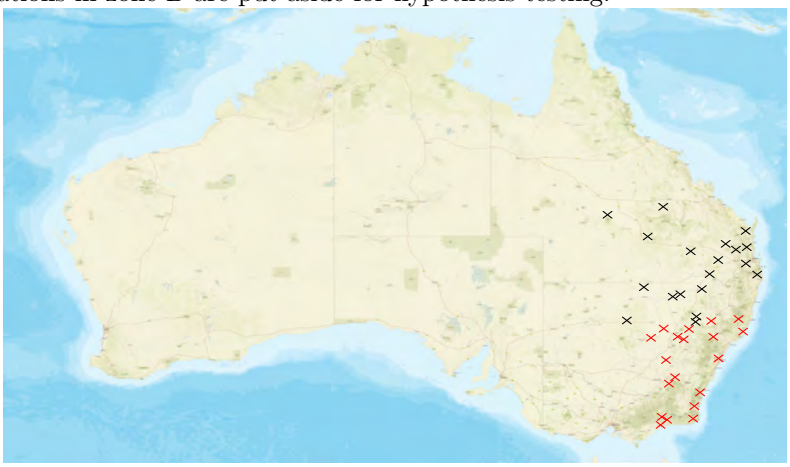
	True $a = 0$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.01	0.020	0.033	0.053	0.040	0.127	0.113
0.05	0.053	0.073	0.140	0.100	0.193	0.187
0.10	0.133	0.160	0.233	0.207	0.273	0.247
0.20	0.333	0.413	0.420	0.453	0.567	0.647
0.30	0.507	0.620	0.587	0.740	0.727	0.800
0.40	0.667	0.793	0.820	0.873	0.907	0.893
0.50	0.833	0.880	0.933	0.940	0.960	0.967
0.60	0.933	0.973	0.980	0.987	1.000	0.993

	True $a = 1$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.30	0.993	1.000	1.000	1.000	1.000	1.000
0.40	0.927	0.987	0.973	0.993	0.933	1.000
0.50	0.807	0.893	0.907	0.920	0.960	0.987
0.60	0.667	0.727	0.693	0.753	0.853	0.887
0.70	0.480	0.553	0.467	0.527	0.680	0.733
0.80	0.260	0.313	0.293	0.353	0.447	0.493
0.90	0.120	0.133	0.147	0.173	0.220	0.267
0.95	0.033	0.047	0.073	0.107	0.167	0.213
0.99	0.000	0.027	0.047	0.080	0.087	0.133

Table 4: Empirical probabilities (P) for testing $H_0 : a = a_0$ against $H_1 : a \neq a_0$ based on 150 simulation replicates from 1000 independent copies of an inverted TEG process (true $a = 0$) with parameters $\{\phi_Y = 0.75, r_Y = 1.2\}$ and a TEG process (true $a = 1$) with parameters $\{\phi_X = 0.10, r_X = 0.25\}$. Significance levels $\alpha \in \{0.01, 0.05, 0.10\}$.

explore possible anisotropy of the spatial dependence. Consider $Z_t, t = 1, \ldots, T$, T copies of a MM process Z with unit Fréchet margin F. It is easy to compute the empirical estimates of $\chi_u(h)$ and

Figure 5: Geographical locations of the 38 meteorological stations located in the East of Australia. Black/Red crosses show the stations in A/B zones. Stations in zone A are used for model inference, and the other stations in zone B are put aside for hypothesis testing.



$\bar{\chi}_u(h)$ from the empirical univariate and bivariate distributions as follows

$$\widehat{\chi}_u(h) = 2 - \frac{\log\left(T^{-1}\sum_{t=1}^{T} \mathbb{1}_{\{U_t(s)<u, U_t(s+h)<u\}}\right)}{\log\left(T^{-1}\sum_{t=1}^{T} \mathbb{1}_{\{U_t(s)<u\}}\right)}, \quad s, s+h \in \mathcal{S}, \tag{6.1}$$

$$\widehat{\bar{\chi}}_u(h) = \frac{2\log\left(T^{-1}\sum_{t=1}^{T} \mathbb{1}_{\{U_t(s)>u\}}\right)}{\log\left(T^{-1}\sum_{t=1}^{T} \mathbb{1}_{\{U_t(s)>u, U_t(s+h)>u\}}\right)} - 1, \quad s, s+h \in \mathcal{S}, \tag{6.2}$$

where $U_t = F(Z_t)$.

Remark 6.1 *When dealing with real data, the marginal laws are usually not unit Fréchet and thus have to be transformed to unit Fréchet margins according to the probability integral transform:* $x \to \frac{-1}{\log(\widehat{F}(x))}$, *with* $\widehat{F}(.)$ *either the empirical distribution function or the estimated generalized extreme value distribution* $GEV_{\hat{\mu}(s),\hat{\sigma}(s),\hat{\xi}(s)}(.)$ *(2.2).*

We divide the dataset according to different directional sectors;
$(-\pi/8, \pi/8]$, $(\pi/8, 3\pi/8]$, $(3\pi/8, 5\pi/8]$, and $(5\pi/8, 7\pi/8]$, where 0 represents the northing direction. On the basis of observed data, we construct the empirical estimates $\widehat{\chi}_u(h)$ and $\widehat{\bar{\chi}}_u(h)$. Figure 6 displays the directional loess smoothing of the empirical estimates $\widehat{\chi}_u(h)$ and $\widehat{\bar{\chi}}_u(h)$ at $u = 0.97$ with respect to h. Based on these estimates there is no clear evidence of anisotropy. On the other hand, as mentioned in Sect. 2.2., the empirical estimates of $\chi_u(h)$ and $\bar{\chi}_u(h)$ can be useful in distinguishing between AD and AI. By means of visual inspection, Figure 6 provides an indication that AD between stations seems to be present up to a distance of 500 km, where $\widehat{\chi}_{0.97}(h) > 0$ for $h \leq 500$ km, whereas AI appears could be conjectured between 500 km and 1000 km, where $\widehat{\bar{\chi}}_{0.97}(h) < 1$ for $h \in [500, 1000]$ km. Moreover, the pairs of sites separated by a distance $h > 1000$ km are (exactly) independent as $\widehat{\bar{\chi}}_{0.97}(h) \simeq 0$ for $h > 1000$ km. Hence, a MM model sounds a strong candidate for quantifying the extreme dependence structure in this rainfall dataset.

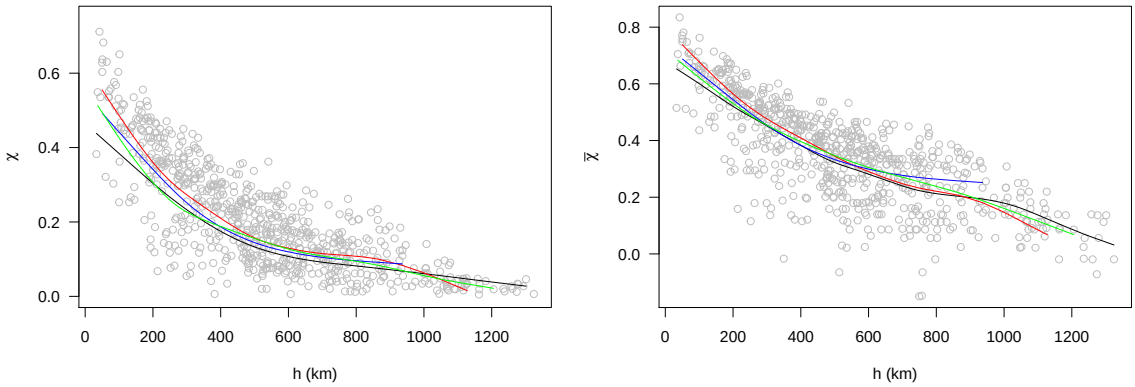


Figure 6: Pairwise empirical estimates of χ (left panel) and $\bar{\chi}$ (right panel) versus distance at threshold $u = 0.970$. Grey points are empirical pairwise estimates for all data pairs. Colored lines are the loess smoothed values of the empirical estimates in different directional sectors: black line $(-\pi/8, \pi/8]$, red line $(\pi/8, 3\pi/8]$, blue line $(3\pi/8, 5\pi/8]$, and green line $(5\pi/8, 7\pi/8]$.

6.3. Testing procedure

The two statistics LR_a and Z_a have emerged as an efficient tools for testing hypothesis on the mixing parameter of spatial MM models by the simulation study. We now describe a testing procedure for the mixing parameter of a MM process Z (2.13) based on the observed rainfall dataset. As the proposed tests are model-based approaches which means that the distribution family has to be specified prior to perform the tests. Hence, we divide the daily rainfall dataset from the 38 sites into two zones; A and B (see Figure 5). The 19 sites in zone A are used for model inference and the other sites in zone B are put aside for testing. Therefore, we analyze the daily rainfall data from zone A using the censored pairwise approach (3.2), where the threshold u is taken as the 0.9 empirical quantile at each site. Equal weights are assumed. For this purpose, we shall consider nine models which belong to the three classes: $\mathbf{M_1} - \mathbf{M_3}$ are flexible MM models, $\mathbf{M_4} - \mathbf{M_6}$ are MS models, and $\mathbf{M_7} - \mathbf{M_9}$ are Inverted MS models.

- $\mathbf{M_1}$: a MM (2.13) model where X is a TEG process with an exponential correlation function $\rho_X(h) = \exp(-\|h\|/\phi_X)$, $\phi_X > 0$. $\mathcal{A}_X$ is a disk of fixed and unknown radius r_X, and Y is an inverted TEG process with exponential correlation function $\rho_Y(h) = \exp(-\|h\|/\phi_Y)$, $\phi_Y > 0$, and $\mathcal{A}_Y$ is a disk with fixed and unknown radius r_Y.

- $\mathbf{M_2}$: a MM model where X is a TEG process with correlation function of type powered exponential defined by $\rho(h) = \exp[-(\|h\|/\phi_X)^{\kappa_X}]$, $\phi_X > 0$ and $0 < \kappa_X < 2$, where ϕ_X and κ_X are the range and the smoothing parameters, respectively. Y is an isotropic inverted BR process with a spherical semivariogram model

$$\gamma_Y(h) = \begin{cases} \alpha_Y\{1.5(\|h\|/\phi_Y) - 0.5(\|h\|/\phi_Y)^3\} & , \|h\| \le \phi_Y \\ \alpha_Y & , \|h\| > \phi_Y, \end{cases}$$

where $\phi_Y > 0$ is the range and $\alpha_Y > 0$ denotes the sill; the value that the semivariogram model attains at the range.

- $\mathbf{M_3}$: a MM model where X is a TEG process as in $\mathbf{M_1}$. Y is an isotropic inverted Smith process with a diagonal covariance matrix $\Sigma_Y = \phi_Y \mathbf{I}_2$, $\mathbf{I}_2$ is 2×2 identity matrix.
- $\mathbf{M_4}$: a MS TEG process described as X in $\mathbf{M_1}$.
- $\mathbf{M_5}$: a MS isotropic BR process with a spherical semivariogram model

$$\gamma_X(h) = \begin{cases} \alpha_X\{1.5(\|h\|/\phi_X) - 0.5(\|h\|/\phi_X)^3\} & , \|h\| \leq \phi_X \\ \alpha_X & , \|h\| > \phi_X. \end{cases}$$

- $\mathbf{M_6}$: a MS isotropic Smith process with $\Sigma_X = \phi_X \mathbf{I}_2$.
- $\mathbf{M_7}$: an inverted TEG process described as Y in $\mathbf{M_1}$.
- $\mathbf{M_8}$: an inverted BR process described as Y in $\mathbf{M_2}$.
- $\mathbf{M_9}$: an inverted Smith process described as Y in $\mathbf{M_3}$.

We fit the $GEV_{\mu(s),\sigma(s),\xi(s)}(x)$ (2.2) separately to each site. Then data are transformed to unit Fréchet margins through the transformation $x \to \frac{-1}{\log(\widehat{F}(x))}$, where $\widehat{F}(.)$ is the estimated GEV cumulative distribution function. The CLIC criterion [45], defined as CLIC$= -2[p\ell(\widehat{\vartheta}) - tr\{\mathcal{J}(\widehat{\vartheta})\mathcal{H}^{-1}(\widehat{\vartheta})\}]$ is used to choose the best fitted model. Lower values of CLIC indicate a better fit. The matrices $\mathcal{H}(\vartheta)$ and $\mathcal{J}(\vartheta)$ and the related quantities (CLIC and standard errors) are obtained by Monte Carlo procedure (3.3) and (3.4) through simulating data with $M = 1500$ independent draws from the fitted model.

Our results are summarized in Table 5. The best-fitting model for zone A, as judged by CLIC, is the hybrid dependence model $\mathbf{M_2}$, for which pairs of sites separated by a distance $\|h\|$ smaller than $2\widehat{r}_X \simeq 720$ km are exactly MS independent. While, the dependence structure is characterized by the AI part of this model for pairs separated by a distance $\|h\|$ greater than 720 km.

Lastly, we consider the best-fitting model in zone A ($\mathbf{M_2}$), we perform the proposed statistical tests LR_a and Z_a to examine whether model $\mathbf{M_2}$ can be an appropriate model to quantify the extremal dependence structures for rainfall data in zone B as in zone A. Simply, on the basis of observed data in zone B, we want to test whether the null hypothesis $H_0 : a = a_0$ holds for $a_0 \in \{0.01, 0.1, 0.2, 0.3, 0.39, 0.5, 0.6, 0.99\}$. To assess the validity of our results, the corresponding CLIC under the null hypothesis is computed. Our results are summarized in Table 6. In summary, both statistics retain the null hypothesis $H_0 : a = 0.39$. Moreover, we observe the agreement between the test findings and CLIC values. Thus, the MM model $\mathbf{M_2}$ can be used to model the dependence structures of daily precipitation in both zones A and B. These conclusions seem to be reasonable and could be expected, since both studied zones are located in the east coast of Australia and may have a homogeneous precipitation patterns. Similar data have also been used in [2], where the authors also concluded that winter rainfall in this region is homogeneous.

Table 5: Parameter estimates of selected dependence models fitted to the daily rainfall data at zone A. The CLIC criterion and standard errors reported between parentheses. (*) indicates to the lower CLIC.

Model	$\widehat{\phi}_X$	κ_X	$\widehat{r}_X$	$\widehat{a}$	$\widehat{\phi}_Y$	$\widehat{r}_Y$	$\widehat{\alpha}_Y$	CLIC
$\mathbf{M_1}$	342.87	-	713.08	$\simeq 1$	2235.51	986.40	-	1952371
	(58.39)	-	(210.74)	(0.03)	(498.65)	(271.34)	-	-
$\mathbf{M_2}$	206.22	1.93	361.71	0.39	1014.54	-	2.46	1952188*
	(81.17)	(1.04)	(199.62)	(0.11)	(306.55)	-	(1.24)	-
$\mathbf{M_3}$	32.19	-	108.55	0.28	1154.96	-	-	1952402
	(13.76)	-	(70.02)	(0.16)	(217.88)	-	-	-

Model	$\widehat{\phi}_X$	$\widehat{r}_X$	$\widehat{\alpha}_X$	CLIC
$\mathbf{M_4}$	311.92	761.84	-	1952378
	(68.52)	(189.10)	-	-
$\mathbf{M_5}$	422.89	-	4.77	1952266
	(78.29)	-	(2.93)	-
$\mathbf{M_6}$	189.59	-	-	1960850
	(41.66)	-	-	-

Model	$\widehat{\phi}_Y$	$\widehat{r}_Y$	$\widehat{\alpha}_Y$	CLIC
$\mathbf{M_7}$	18.05	707.12	-	1960726
	(6.81)	(115.41)	-	-
$\mathbf{M_8}$	533.04	-	5.63	1952411
	(88.15)	-	(2.27)	-
$\mathbf{M_9}$	316.63	-	-	1963215
	(207.85)	-	-	-

Equivalently, an independent two-samples Z-test is performed to compare the mixture parameter in both zones. Denote by a_A (resp. a_B the mixing parameter for the best-fitting model in zone A (resp. zone B). We want to test whether the null hypothesis $H_0 : a_A = a_B$ holds, the statistic

$$Z_a^* = \frac{\widehat{a}_A - \widehat{a}_B}{SE_{\widehat{a}_A - \widehat{a}_B}}, \tag{6.3}$$

where SE stands for the estimated standard error (we found $|Z_a^*| = 0.42$, and p-value $= 0.67$, $\alpha = 0.05$). Similarly, we conclude that there is no significant difference between the mixing parameters in both zones.

By the above testing scheme, our statistics LR_a and Z_a appear to play as an efficient parametric approach for model validation. The latter conclusion can be a big advantage for our testing scheme since the reason for fitting a statistical model to data is to make conclusions about some aspect of the population from which the data were drawn. Such conclusions can be sensitive to the accuracy of the fitted model, so it is necessary to check that the model fits well, especially, in the context of spatial extremes processes such as heavy precipitation, heat waves and windstorms, since when modeling dependence for these processes, imposing a specific type of asymptotic behaviour has important consequences on the estimation of return levels for spatial functionals.

	LR_a		$\lvert Z_a \rvert$		
a_0	statistic	p-value	statistic	p-value	CLIC
0.01	16.07	6.10×10^{-5}	4.42	9.87×10^{-6}	2085837
0.10	10.98	9.21×10^{-4}	3.63	2.83×10^{-4}	1980814
0.20	9.53	2.02×10^{-3}	2.75	5.96×10^{-3}	1955458
0.30	6.34	1.18×10^{-2}	2.19	2.85×10^{-2}	1952356
0.39	1.21	2.71×10^{-1}	0.92	3.58×10^{-1}	1952194
0.50	3.78	5.18×10^{-2}	1.83	6.72×10^{-2}	1952227
0.60	10.14	1.45×10^{-3}	2.68	7.36×10^{-3}	1957503
0.99	21.46	3.61×10^{-6}	4.69	2.73×10^{-6}	2213019

Table 6: Testing results of the null hypothesis $H_0 : a = a_0$, with $a_0 \in \{0.01, 0.1, 0.2, 0.3, 0.39, 0.5, 0.6, 0.99\}$ on the basis of observed data in zone B. $\alpha = 0.05$. The corresponding CLIC values under H_0 are reported.

7. CONCLUDING REMARKS

In summary, we have considered hypothesis testing for the mixture parameter of a MM models using two statistics the Z_a and LR_a. A censored pairwise likelihood is employed for inferential purposes, since composite marginal likelihood appears to be an attractive alternative for modeling complex data, and has received increasing attention in handling high dimensional datasets when the joint distribution is computationally difficult to evaluate, or intractable due to complex structure of dependence, like in MS and MM models.

A simulation study has shown that both statistics perform well, even when we have considered testing at values that are very close to the boundary points. These tests can control the type I error rate α and they seem to be unbiased. The LR_a test seem to be more powerful than the Z_a one, for a away from the boundaries, while for $a \in \{0, 1\}$, the contrary seems to hold. We show by the real data example that our testing procedure could be a performant model validation tool. Also, we find that MM model appears of interest for modeling environmental data since it can handle both AD and AI. Of course, still, much more research both from theoretical and practical point of view has to be done, for testing at boundary points (i.e., $a_0 = 1$ (AD) or $a_0 = 0$ (AI)). The LR statistic could also be useful when comparing nested models. For instance, testing whether it is more appropriate to use a powered exponential correlation function $\rho(h) = \exp\{-(\|h\|/\phi)^\kappa\}$, for $\phi > 0$ and $0 < \kappa < 2$, or an exponential correlation function $\rho(h) = \exp(-\|h\|/\phi)$ in spatial TEG model, more precisely, testing the hypothesis $H_0 : \kappa = 1$ versus $H_0 : \kappa \neq 1$. The ideas sketched in this paper provides initial steps toward testing of hypotheses related to spatial rare events. Our future work will be dedicated to investigate the above-mentioned issues.

8. APPENDIX

APPENDIX A

Estimation using censored pairwise likelihood approach (3.5).

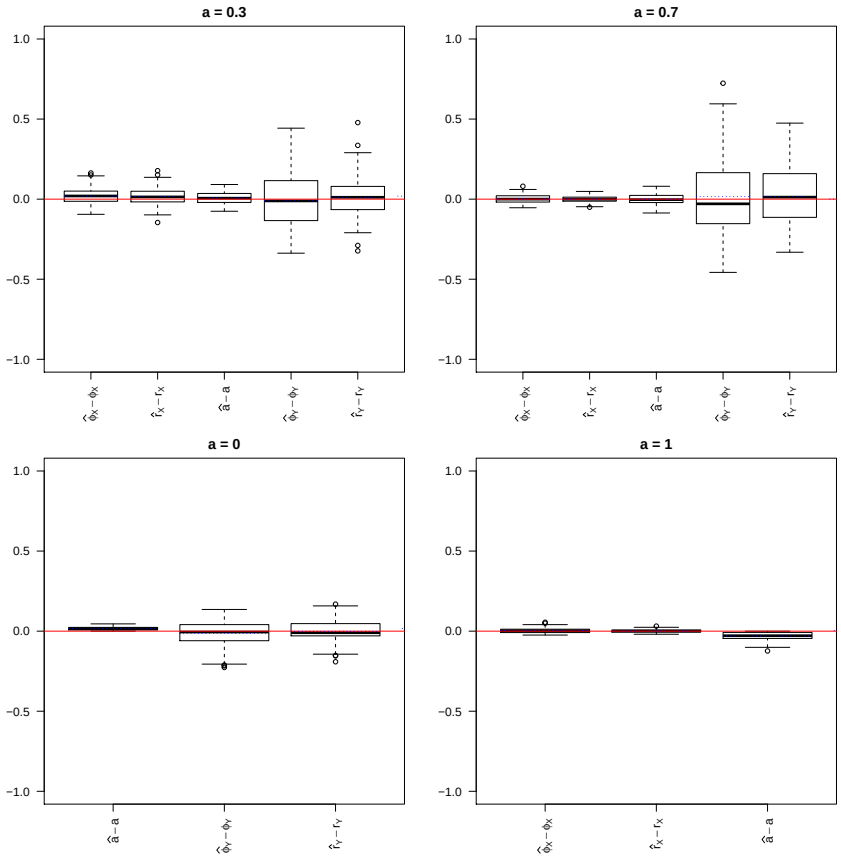


Figure 7: Boxplots display $\widehat{\vartheta} - \vartheta$ of the estimated parameters using censored pairwise likelihood approach (3.5) based on 100 simulation replicates from 1000 independent copies of the model **MM** with parameters $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$, and mixture parameter $a \in \{0, 0.3, 0.7, 1\}$. Blue dotted/red horizontal lines show the estimated errors means/0 value.

APPENDIX B

Supplementary material

The rainfall dataset used in (Sect. 6.) is included as an electronic supplementary material on the website:
http://math.univ-lyon1.fr/homes-www/abuawwad/Rainfalldataset/.

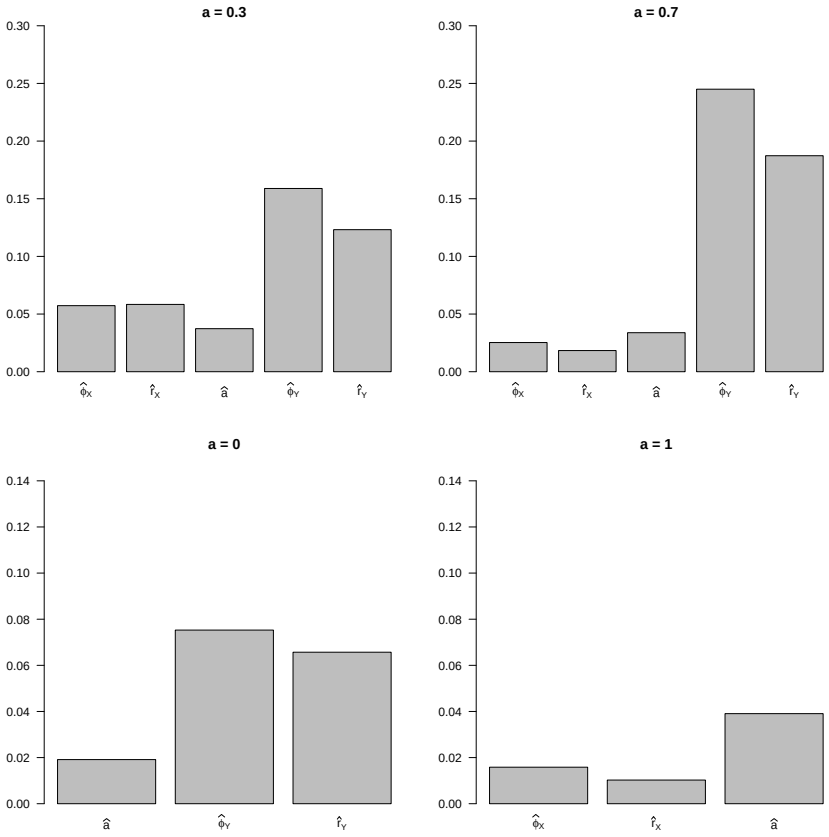


Figure 8: Barplots display RMSE of $\widehat{\vartheta} = \{\widehat{\phi}_X, \widehat{r}_X, \widehat{a}, \widehat{\phi}_Y, \widehat{r}_Y\}$ for each estimated parameter using censored pairwise likelihood approach (3.5) based on 100 simulation replicates from 1000 independent copies of the model **MM** with parameters $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$, and mixture parameter $a \in \{0, 0.3, 0.7, 1\}$.

ACKNOWLEDGMENTS

This work has been supported by the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). It was also partly supported by the CERISE LEFE-INSU projcet. We are very grateful to Jean-Noël Bacro, Carlo Gaetan and Gwladys Toulemonde for providing their estimation C codes which we have developed for computing the statistics of our tests and the related quantities. We also acknowledge the valuable comments from the referees.

RECEIVED: FEBRUARY, 2019.
REVISED: APRIL, 2019.

REFERENCES

[1] Bacro, J.-N. and Gaetan, C. (2014): Estimation of spatial max-stable models using threshold exceedances **Statistics and Computing**, 24(4):651–662.

[2] Bacro, J.-N., Gaetan, C., and Toulemonde, G. (2016): A flexible dependence model for spatial extremes **Journal of Statistical Planning and Inference**, 172:36–52.

[3] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006): **Statistics of extremes: theory and applications** John Wiley & Sons.

[4] Besag, J. (1974): Spatial interaction and the statistical analysis of lattice systems **Journal of the Royal Statistical Society. Series B (Methodological)**, pages 192–236.

[5] Buhl, S., Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2016): Semiparametric estimation for isotropic max-stable space-time processes **arXiv preprint arXiv:1609.04967**.

[6] Castruccio, S., Huser, R., and Genton, M. G. (2016): High-order composite likelihood inference for max-stable distributions and processes **Journal of Computational and Graphical Statistics**, 25(4):1212–1229.

[7] Cattelan, M. and Sartori, N. (2016): Empirical and simulated adjustments of composite likelihood ratio statistics **Journal of Statistical Computation and Simulation**, 86(5):1056–1067.

[8] Chandler, R. E. and Bate, S. (2007): Inference for clustered data using the independence loglikelihood **Biometrika**, pages 167–183.

[9] Coles, S. (2001): **An introduction to statistical modeling of extreme values** Springer.

[10] Coles, S., Heffernan, J., and Tawn, J. (1999): Dependence measures for extreme value analyses **Extremes**, 2(4):339–365.

[11] Cox, D. R. and Reid, N. (2004): A note on pseudolikelihood constructed from marginal densities **Biometrika**, 91(3):729–737.

[12] Cressie, N. (1993): **Statistics for spatial data** New York: Wiley.

[13] Davies, R. B. (1977): Hypothesis testing when a nuisance parameter is present only under the alternative **Biometrika**, 64(2):247–254.

[14] Davies, R. B. (1987): Hypothesis testing when a nuisance parameter is present only under the alternatives **Biometrika**, pages 33–43.

[15] Davison, A., Huser, R., and Thibaud, E. (2013): Geostatistics of dependent and asymptotically independent extremes **Mathematical Geosciences**, 45(5):511–529.

[16] Davison, A. C. and Gholamrezaee, M. M. (2012): Geostatistics of extremes 468:581–608.

[17] Davison, A. C., Padoan, S. A., and Ribatet, M. (2012): Statistical modeling of spatial extremes **Statistical science**, pages 161–186.

[18] De Haan, L. (1984): A spectral representation for max-stable processes **The annals of probability**, pages 1194–1204.

[19] De Haan, L. and Ferreira, A. (2007): **Extreme value theory: an introduction** Springer Science & Business Media.

[20] De Haan, L. and Pereira, T. T. (2006): Spatial extremes: Models for the stationary case **The annals of statistics**, pages 146–168.

[21] Geys, H., Molenberghs, G., and Ryan, L. M. (1999): Pseudolikelihood modeling of multivariate outcomes in developmental toxicology **Journal of the American Statistical Association**, 94(447):734–745.

[22] Heagerty, P. and Lele, S. (1998): A composite likelihood approach to binary spatial data **Journal of the American Statistical Association**, 93(443):1099–1111.

[23] Huser, R. and Davison, A. C. (2013): Composite likelihood estimation for the brown–resnick process **Biometrika**, 100(2):511–518.

[24] Huser, R. and Davison, A. C. (2014): Space–time modelling of extreme events **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, 76(2):439–461.

[25] Huser, R., Davison, A. C., and Genton, M. G. (2016): Likelihood estimators for multivariate extremes **Extremes**, 19(1):79–103.

[26] Kabluchko, Z., Schlather, M., and De Haan, L. (2009): Stationary max-stable fields associated to negative definite functions **The Annals of Probability**, pages 2042–2065.

[27] Kent, J. T. (1982): Robust properties of likelihood ratio tests **Biometrika**, 69(1):19–27.

[28] Kuonen, D. (1999): Miscellanea. saddlepoint approximations for distributions of quadratic forms in normal variables **Biometrika**, 86(4):929–935.

[29] Ledford, A. W. and Tawn, J. A. (1996): Statistics for near independence in multivariate extreme values **Biometrika**, 83(1):169–187.

[30] Lindsay, B. G. (1988): Composite likelihood methods **Contemporary mathematics**, 80(1):221–239.

[31] Montero, J.-M. and Mateu, J. (2015): **Spatial and spatio-temporal geostatistical modeling and kriging**, volume 998 John Wiley & Sons.

[32] Opitz, T. (2013): Extremal t processes: Elliptical domain of attraction and a spectral representation **Journal of Multivariate Analysis**, 122:409–413.

[33] Pace, L., Salvan, A., and Sartori, N. (2011): Adjusting composite likelihood ratio statistics **Statistica Sinica**, pages 129–148.

[34] Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010): Likelihood-based inference for max-stable processes **Journal of the American Statistical Association**, 105(489):263–277.

[35] Ribatet, M. (2015): Spatialextremes: R package, version 2.0–2 **Available at http://spatialextremes.r-forge.rproject.org/**.

[36] Rotnitzky, A. and Jewell, N. P. (1990): Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data **Biometrika**, pages 485–497.

[37] Schlather, M. (2002): Models for stationary max-stable random fields **Extremes**, 5(1):33–44.

[38] Schlather, M. and Twan, J.-A. (2003): A dependence measure for multivariate and spatial extreme values: Properties and inference **Biometrika**, 1(90):195–156.

[39] Sibuya, M. (1960): Bivariate extreme statistics, **Annals of the Institute of Statistical Mathematics**, 11(2):195–210.

[40] Smith, R. L. (1990): Max-stable processes and spatial extremes **Unpublished manuscript**.

[41] Thibaud, E., Mutzner, R., and Davison, A. C. (2013): Threshold modeling of extreme spatial rainfall **Water resources research**, 49(8):4633–4644.

[42] Thibaud, E. and Opitz, T. (2015): Efficient inference and simulation for elliptical pareto processes **Biometrika**, 102(4):855–870.

[43] Varin, C. (2008): On composite marginal likelihoods **AStA Advances in Statistical Analysis**, 92(1):1–28.

[44] Varin, C., Reid, N., and Firth, D. (2011): An overview of composite likelihood methods **Statistica Sinica**, pages 5–42.

[45] Varin, C. and Vidoni, P. (2005): A note on composite likelihood inference and model selection **Biometrika**, pages 519–528.

[46] Wadsworth, J. L. and Tawn, J. A. (2012): Dependence modelling for spatial extremes **Biometrika**, 99(2):253–272.

[47] Wadsworth, J. L. and Tawn, J. A. (2013): Efficient inference for spatial extreme value processes associated to log-gaussian random functions **Biometrika**, 101(1):1–15.