The spectrogram: A threshold-based inferential tool for extremes of stochastic processes[*]

Thomas Opitz

Biostatistics and Spatial Processes Unit (BioSP), French National Institute of Agronomic Research (INRA), Avignon, France
e-mail: Thomas.Opitz@PACA.INRA.fr

Jean-Noël Bacro

Institute of Mathematics and Modeling (I3M), University of Montpellier, Montpellier, France
e-mail: jean-noel.bacro@univ-montp2.fr

and

Pierre Ribereau

Institute of Actuarial Science and Finance (ISFA), University Claude Bernard, Lyon, France
e-mail: pierre.ribereau@univ-lyon1.fr

Abstract: We extend the statistical toolbox for inferring the extreme value behavior of stochastic processes by defining the spectrogram. It is based on a pseudo-polar representation of bivariate data and represents a collection of spectral measures that characterize the bivariate extremal dependence structure in terms of a univariate distribution. Based on threshold exceedances in the original event data, estimation of the spectrogram is flexible and efficient. The virtues of the spectrogram for exploratory studies and for parametric inference are highlighted. We propose a variance reduction technique that can be applied to the estimation of summary statistics like the extremogram. Parametric inference based on distance measures between the empirical and the model spectrogram, as for instance pairwise likelihoods or least squares distances, is developed. Simulated data illustrate gains in parametric estimation efficiency compared to a standard estimation approach. An application to precipitation data collected in the Cévennes region in France shows the practical utility of the introduced notions.

Keywords and phrases: Max-stable processes, pairwise inference, spatial extremes, spectral measure, variance reduction.

Contents

1. Introduction

The statistical analysis of extreme outcomes of stochastic processes plays an important role in many disciplines, including environmental and climatological research, finance and actuarial science. Extreme events in such processes can stretch across space and time and can inflict considerable damage. Extreme value analysis provides a coherent methodology to stochastically model extremes and to quantify extreme risks.

The last decade has seen a boom in the development of workable models and statistical inference for spatial and spatio-temporal extremes in the asymptotically dependent case; see Davison, Padoan and Ribatet [17], Davison and Gholamrezaee [15] who provide an overview of the field of geostatistics of extremes. The commonly used models are based on max-stable processes [19]. Max-stability arises naturally as an asymptotic property for maxima and has its counterpart for threshold exceedances in the peaks-over-threshold stability of generalized Pareto processes [26, 20, 2]. Estimation approaches based on threshold exceedances [7, 28, 3, 48, for instance] currently receive increasing attention

due to their potential for more efficient and more flexible estimation in comparison to an often used composite likelihood approach with block maxima [34].

In this paper, we exploit peaks-over-threshold stability in the bivariate distributions: in adequately defined pseudo-polar coordinate systems, it corresponds to limiting independence between the angular and the radial component as the radial level increases towards infinity. The spectral measure describes the distribution of angles and characterizes the bivariate extremal dependence. It presents the advantage of being a univariate measure, which simplifies statistical procedures. Throughout the following presentation, we employ the term "space-time" when referring to time, space or space augmented by the time dimension.

Based on the assumption of stationarity, we define the spectrogram as an inferential tool for pairwise behavior with respect to the lag in space-time. It is constituted from the spectral measures for point pairs in space-time and provides a generalized and unified view on dependence characterizations like the extremal coefficient function [42] or the tail correlation function, a variant of the extremogram [14]. We propose novel and flexible estimation procedures for such summary functionals and parametric extreme value models.

The remainder of this paper is organized as follows: At the beginning of Section 2, we first present some necessary background on extreme value theory and max-stable processes. Then, it is adapted to our purposes by defining the notions of radial aggregation and spectral measures closely related to well-known summaries and characterizations of extreme value dependence. This approach generalizes the usual norm-based definition of radial aggregations in the literature. The spectrogram and its inferential toolbox are subject of the central Section 3, which features novel theoretical results for a variance reduction technique in empirical estimation and for parametric inference in model-based approaches. We substantiate the utility of the introduced notions: first numerically, by means of a simulation study in Section 4 that shows to what extent our parametric estimation approach can improve estimation efficiency in comparison to a standard approach, then in the context of an application to daily precipitation data over the French Cévennes region in Section 5. The concluding remarks in Section 6 address some remaining issues and perspectives. A recall of useful dependence structures for spatial extremes is deferred to the appendix. All computations in this paper were carried out with the statistical software library R [47].

Notations

Operations on vectors shall be interpreted componentwise, as in $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_m + y_m)$ or $\mathbf{x}/\mathbf{y} = (x_1/y_1, \ldots, x_m/y_m)$. We further use the constant vectors $\mathbf{0} = (0, \ldots, 0)^T$ and $\mathbf{1} = (1, \ldots, 1)^T$. The components of an indexed vector $\mathbf{x_i}$ are given as x_{ij}, $j = 1, \ldots, m$. Rectangular bounded or unbounded sets are given with vectorized notations like $[\mathbf{y}, \mathbf{z}] = [y_1, z_1] \times \cdots \times [y_m, z_m]$. The complementary set of $B \subset \mathbb{R}^m$ is denoted by B^C, and its boundary is written ∂B. The operator $(\cdot)_+ = \max(\,\cdot\,, 0)$ truncates at 0. The indicator function of a set B is written χ_B. For a Dirac function we write $\chi_{\mathbf{x_0}}(\cdot)$ with $\chi_{\mathbf{x_0}}(\mathbf{x}) = 1$ if $\mathbf{x} = \mathbf{x_0}$ and $\chi_{\mathbf{x_0}}(\mathbf{x}) = 0$ otherwise.

2. Extreme value theory

2.1. Domain of attraction

The fundamental limit theorem of extreme value theory states that convergence of a process of rescaled pointwise maxima leads to a max-stable limit process [19]: for n independent copies $\{X_i(s), s \in D\}$ of a process $\mathbf{X}$ defined on a non-empty measurable set $D \subset \mathbb{R}^d$, we assume that

$$\left\{ \max_{i=1,\ldots,n} a_n(s)^{-1}(X_i(s) - b_n(s)) \right\} \to \{Z(s)\}, \quad n \to \infty, \tag{2.1}$$

where $a_n(s) > 0$ and $b_n(s)$ are adequately chosen normalizing sequences and $\mathbf{Z} = \{Z(s)\}$ is a non-degenerate limit process. Then $\mathbf{Z}$ is max-stable, and we say that $\mathbf{X}$ is in the domain of attraction of $\mathbf{Z}$. For continuous stochastic processes, the domain of attraction condition (2.1) has recently been related to exceedances over a high threshold. The limiting Generalized Pareto (GP) processes have been studied by Ferreira and de Haan [26] and Dombry and Ribatet [20]; an adaptation to the flexible extremal-t dependence model was developed in [48]. When continuity is assured on a compact domain D, each max-stable process has its Pareto equivalent and vice versa.

A general constructive approach to generate max-stable models goes back to the seminal paper of de Haan [19]: Let $\{V_i\}$ be the points of a Poisson process on $(0, \infty)$ with intensity $v^{-2}\mathrm{d}v$, and let independently $Q_i(s)$ for $s \in D$ and $i = 1, 2, \ldots$ be independent replicates of a random function $\mathbf{Q}$ with $\mathbb{E}Q_+(s) = 1$. Then

$$Z(s) = \max_{i=1,2,\ldots} V_i Q_i(s), \quad s \in D, \tag{2.2}$$

defines a max-stable field on D with unit Fréchet margins. This construction was interpreted by Smith [44] to introduce a class of models sometimes referred to as *storm profile processes*: a uniform Poisson process $\{S_i\}$ on $\mathbb{R}^d$ is used to define $Q_i(s) = f(s, S_i)$ for some non-negative kernel function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$, $\int f(s_1, s_2)\mathrm{d}s_2 = 1$ for all s_1, which describes the deterministic shape of storms. The *Gaussian extreme value process* is obtained for the Gaussian density f. Random storm shapes $\mathbf{Q}$ were proposed by Schlather [41] and define the class of *extremal Gaussian processes* if $\mathbf{Q}$ is a Gaussian random field $\mathbf{Q}_{\mathrm{gauss}}$. A more general class are the extremal t processes which arise from $\mathbf{Q} = c_\eta(\mathbf{Q}_{\mathrm{gauss}}{}^+)^\eta$ with $\eta > 0$ and a normalizing constant c_η [32]. The *Brown–Resnick process* [6, 29] arises from $\mathbf{Q} = \exp(\tilde{\mathbf{Q}} - 0.5\,\mathrm{var}(\tilde{\mathbf{Q}}))$ with $\tilde{\mathbf{Q}}$ defined as a fractional Brownian motion [31]. We refer to Appendix A for the bivariate expressions and spectral measures of these models.

Max-stable processes are characterized by finite-dimensional max-stable distributions, also known as multivariate extreme value distributions (MEVD) G. Max-stability means that vector sequences $\boldsymbol{\alpha_n} > 0$ and $\boldsymbol{\beta_n}$ exist satisfying $G^n(\boldsymbol{\alpha_n}\mathbf{z} + \boldsymbol{\beta_n}) = G(\mathbf{z})$. In particular, univariate marginal distributions G_j of a max-stable process are generalized extreme value distributions (GEV):

$$G_j(z) = \mathrm{GEV}_{\xi_j, \nu_j, \sigma_j}(z) = \exp\left[-\left\{ 1 + \xi_j \sigma_j^{-1}(z - \nu_j) \right\}_+^{-1/\xi_j} \right], \tag{2.3}$$

with three parameters: ξ_j for shape, ν_j for position and $\sigma_j > 0$ for scale. The case $\xi_j = 0$ corresponds to $G_j(z) = \exp[-\exp\{-(z - \nu_j)/\sigma_j\}]$. It is convenient to assume univariate marginal distributions given by the unit Fréchet distribution $G_j^*(z) = \exp(-1/z)\chi_{(0,\infty)}(z)$. The transformation $-1/\log G_j(Z_j)$ can be applied to all margins Z_j of $\mathbf{Z}$, which yields a multivariate extreme value distribution G^* with unit Fréchet marginal distributions G_j^*.

Balkema and Resnick [4] showed that any MEVD G can be represented through an exponent measure μ as $G(\mathbf{z}) = \exp(-\mu(-\boldsymbol{\infty}, \mathbf{z}]^C)$. For G^*, the exponent measure μ^* can be defined on the punctured cone $E = [\mathbf{0}, \boldsymbol{\infty}) \setminus \{\mathbf{0}\} \subset \mathbb{R}^m$ and satisfies the following property of homogeneity of order -1:

$$\mu^*(tB) = t^{-1}\mu^*(B), \quad t > 0, \tag{2.4}$$

for Borel sets $B \subset \overline{E} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$ [36]. The marginal tail mass of μ^* is $\mu^*\{\mathbf{z} : z_j > z_0\} = 1/z_0$ for $z_0 > 0$. The exponent function V is often used to write a MEVD $G^*(\mathbf{z})$ as $\exp(-V(\mathbf{z}))$, where $V(\mathbf{z}) = \mu^*([\mathbf{0}, \mathbf{z}]^C)$. We can standardize the margins X_j of $\mathbf{X}$ to assure convergence of pointwise maxima towards G^*. Therefore, we use a probability integral transform to transform X_j to a nonnegative variable X_j^* with standard Pareto tail behavior such that $x\, P(X_j^* > x) \to 1$ as x tends to infinity. Then (2.1) is true for $\mathbf{X}^*$ with $a_n = n$, $b_n = 0$ and limit distribution G^*. There is no unique choice for the *-transformation; common choices in the literature are to use either $X_j^* = -1/\log F_j(X_j)$ or $X_j^* = 1/(1 - F_j(X_j))$, leading respectively to unit Fréchet or standard Pareto target distributions when F_j is continuous.

The convergence and estimation of univariate extreme value behavior has been extensively studied, see the introductory presentation in Coles [11], for instance. After standardizing the marginal distributions with the *-transformation, it is convenient to focus on the dependence structure. The representation through the exponent measure μ^* allows us to formulate convergence not in terms of maxima, but in a more general way based on the original data through the concept of multivariate regular variation, see Chapter 6 in Resnick [37]. The following vague convergence property is satisfied on $\overline{E}$:

$$\mu_n^*(\cdot) = n\,\mathrm{pr}(n^{-1}\mathbf{X}^* \in \cdot) \to \mu^*(\cdot), \quad n \to \infty. \tag{2.5}$$

Vague convergence is equivalent to the convergence $\mu_n^*(K) \to \mu^*(K)$ for all sets K relatively compact in $\overline{E}$ for which $\mu^*(\partial K) = 0$.

2.2. Pseudo-polar representations and spectral measures

We now introduce the notion of radial aggregation and the resulting spectral measures. A sizeable amount of literature on spectral measures exists; we refer the reader to Section 8.2.3 in [5] for a presentation of the usual approach with the aggregation of a vector defined as a norm of the vector. In the following, this framework is slightly generalized to 1-homogeneous functions, allowing us to apply useful aggregation functions that are not norms.

2.2.1. Radial aggregation

The homogeneity (2.4) allows us to factorize μ^* in pseudo-polar coordinate systems with a radial and an angular component. If the radial component is typically chosen as a norm of a vector $\mathbf{x} \in E$ in the literature, other choices will be meaningful for our purposes. Therefore, we propose the more general concept of an aggregation function, which aggregates a vector $\mathbf{x} \in E$ into a scalar value.

Definition 2.1 (Aggregation function)**.** A non-null and measurable function

$$\mathrm{rad} : [\mathbf{0}, \infty) \to [0, \infty) \tag{2.6}$$

is called *aggregation function* if it is continuous and verifies the following 1-homogeneity condition:

$$\mathrm{rad}(t\mathbf{x}) = t\,\mathrm{rad}(\mathbf{x}), \quad t > 0.$$

An aggregation function defines the *radius* $r = \mathrm{rad}(\mathbf{x})$ of $\mathbf{x} \in E$.

Due to the continuity and the homogeneity of rad, the unit sphere $\mathcal{S}_{\mathrm{rad}} = \{\mathbf{x} \in [\mathbf{0}, \infty) \mid \mathrm{rad}(\mathbf{x}) = 1\}$ associated to rad is relatively compact in the space $\overline{E}$. To introduce a pseudo-polar transformation T, we choose a norm $\|\cdot\|$ on $\mathbb{R}^m$, called *angular norm*, and write $\mathcal{S} = \{\mathbf{x} \in [\mathbf{0}, \infty) \mid \|\mathbf{x}\| = 1\}$ for the intersection of the first quadrant and the unit sphere with respect to $\|\cdot\|$. We define

$$T : E \to [0, \infty) \times \mathcal{S}, \quad (r, \mathbf{w}) = T(\mathbf{x}) = (\mathrm{rad}(\mathbf{x}), \mathbf{x}/\|\mathbf{x}\|) . \tag{2.7}$$

The mth coordinate of the angle $\mathbf{w}$ is determined by the remaining ones and can thus be suppressed. We can now represent the limit measure μ^* and the vague convergence (2.5) in pseudo-polar coordinates. In general, difficulties in proving convergence with respect to pseudo-polar coordinates arise from the fact that rad and T are not defined for the values at infinity, whereas the vague convergence in (2.5) is stated with respect to the "compactified" space $[\mathbf{0}, \infty]$ punctured at the origin (cf. the proof of Theorem 6.1 in Resnick [37], for instance).

The spectral measure ρ_{rad} associated with T and μ^* is defined on $\mathcal{S}$ by $\rho_{\mathrm{rad}}(B) = \mu^*(\{\mathbf{x} \in E \mid \mathrm{rad}(\mathbf{x}) \geq 1, \; \mathbf{x}/\|\mathbf{x}\| \in B\})$ for measurable sets $B \subset \mathcal{S}$ and gives rise to a product representation of μ^*:

$$\mu^* \circ T^{-1}(\mathrm{d}(r, \mathbf{w})) = r^{-2}\mathrm{d}r \times \rho_{\mathrm{rad}}(\mathrm{d}\mathbf{w}) \tag{2.8}$$

on $(0, \infty) \times \{\mathbf{w} \in \mathcal{S} : \mathrm{rad}(\mathbf{w}) > 0\}$, with an invariant radial component showing a standard Pareto tail and the spectral measure ρ_{rad} as the angular component that captures the extremal dependence structure. In the case where $\rho_{\mathrm{rad}}(\mathcal{S}) > 0$, the normalized measure $\rho_{\mathrm{rad}}(\cdot)/\rho_{\mathrm{rad}}(\mathcal{S})$ is called *spectral distribution*. When there is no risk of confusion, we will use the simpler notation ρ instead of ρ_{rad}.

Proposition 2.1 (Convergence in pseudo-polar coordinates)**.** *If the random vector* $\mathbf{X}^*$ *with standardized margins verifies the convergence* (2.5)*, then* $R = \mathrm{rad}(\mathbf{X}^*)$ *is univariate regularly varying such that*

$$n\,\mathrm{pr}(n^{-1}R \geq r) \to \rho(\mathcal{S})\,r^{-1}, \quad n \to \infty. \tag{2.9}$$

More generally, vague convergence holds in pseudo-polar coordinates:

$$n \operatorname{pr}(T(\mathbf{X}^*/n) \in \cdot) \to \mu^* \circ T^{-1}(\cdot), \quad n \to \infty, \tag{2.10}$$

on $(0, \infty] \times \mathcal{S}$.

To simplify the proof of Proposition 2.1, we first provide a useful lemma.

Lemma 2.1. *Set* $A_{r_1, r_2} = \{\mathbf{x} \in E \mid r_1 \leq \operatorname{rad}(\mathbf{x}) \leq r_2\}$ *for* $0 < r_1 \leq r_2 \leq \infty$, *and denote by* $\partial A_{r_1, r_2}$ *its boundary with respect to the space* $\overline{E}$. *Then*

- A_{r_1, r_2} *is relatively compact in* $\overline{E}$ *and*
- $\mu^*(\partial A_{r_1, r_2}) = 0$.

Proof of Lemma 2.1. First, we remark that $\mu^*(r\mathcal{S}_{\mathrm{rad}}) = 0$ for $0 < r < \infty$. Otherwise, the homogeneity property of μ^* would lead to a contradiction concerning the Radon property of μ^*. For $r_2 < \infty$, the set $A_{r_1, r_2} = \{r\mathcal{S}_{\mathrm{rad}} \mid r_1 \leq r \leq r_2\}$ is compact in $\overline{E}$ due to the compactness of $\mathcal{S}_{\mathrm{rad}}$. Then $\partial A_{r_1, r_2} = r_1 \mathcal{S}_{\mathrm{rad}} \cup r_2 \mathcal{S}_{\mathrm{rad}}$ such that $\mu^*(\partial A_{r_1, r_2}) = 0$.

For $r_2 = \infty$, it is easy to see that $A_{r_1, \infty} \cup \partial A_{r_1, \infty}$ is closed in $[\mathbf{0}, \infty]$. Due to the continuity of rad at $\mathbf{0}$ and $\operatorname{rad}(\mathbf{0}) = 0$ we have $\mathbf{0} \notin \partial A_{r_1, \infty}$. From Proposition 6.1 of [37], we conclude that $A_{r_1, \infty}$ is relatively compact in E. Since $\partial A_{r_1, \infty} = r_1 \mathcal{S}_{\mathrm{rad}}$, we observe $\mu^*(\partial A_{r_1, \infty}) = 0$, which finishes the proof. □

Proof of Proposition 2.1. Using the technical properties concerning the region of radial exceedances detailed in Lemma 2.1, the proof is straightforward. The convergences (2.9) and (2.10) are obtained from the convergence (2.5) by using certain relatively compact sets. □

With $(R, \mathbf{W}) = T(\mathbf{X}^*)$, the inferential approaches worked out in the following rely upon the approximation

$$\operatorname{pr}(R \geq r, \ \mathbf{W} \in B) \approx r^{-1} \rho(B) \tag{2.11}$$

for a high radial threshold r.

2.2.2. Properties of the spectral measure

The choice of angular norm is a choice of convenience since convergence results do not depend on it. In the following, we focus on the angular L^1-norm $\|\mathbf{x}\| = \sum_{j=1}^m x_j$ on E. In the bivariate case, the first angle component $w = w_1$ designates an angle $(w_1, 1 - w_1)$, so we can identify $\mathcal{S}$ with $[0, 1]$. If $\operatorname{rad}(\mathbf{w}) > 0$ for all $\mathbf{w} \in \mathcal{S}$, then μ^* is uniquely determined by the spectral measure ρ which is only subject to the moment constraints

$$\int_{\mathcal{S}} \frac{w_j}{\operatorname{rad}(\mathbf{w})} \rho(d\mathbf{w}) = 1, \quad j = 1, \dots, m. \tag{2.12}$$

Denote by $\mathbf{e_j}$ the canonical vector with j-th component 1 and 0 elsewhere. If $\operatorname{rad}(\cdot) = \| \cdot \|_{\mathrm{rad}}$ is a norm such that $\|\mathbf{e_j}\|_{\mathrm{rad}} = 1$ for $j = 1, \dots, m$, then $1 \leq \rho(\mathcal{S}) \leq m$. With extremal independence, we observe $\rho(\{\mathbf{e_j}\}) = 1$; with full

extremal dependence, we observe $\rho(\mathcal{S}) = \rho(\{\mathbf{1}/\|\mathbf{1}\|\}) = \|\mathbf{1}\|_{\mathrm{rad}}$. We can switch between two spectral measures ρ_1 and ρ_2 associated to aggregation functions rad_1 and rad_2 respectively, given that $\{\mathbf{w} \in \mathcal{S} : \mathrm{rad}_2(\mathbf{w}) > 0\} \subset \{\mathbf{w} \in \mathcal{S} : \mathrm{rad}_1(\mathbf{w}) > 0\}$, via the transformation formula

$$\rho_2(\mathrm{d}\mathbf{w}) = \frac{\mathrm{rad}_2(\mathbf{w})}{\mathrm{rad}_1(\mathbf{w})}\rho_1(\mathrm{d}\mathbf{w}), \tag{2.13}$$

which holds on $\{\mathbf{w} \in \mathcal{S} : \mathrm{rad}_1(\mathbf{w}) > 0\}$. In Section 3.1, we discuss a variance reduction technique based on (2.13).

2.2.3. Empirical estimation of spectral measures

Given data vectors $\mathbf{x_i}$, $i = 1, \ldots, n$, we can define an empirical spectral measure based on the tail approximation (2.11). The set of atoms of the empirical measure is given by the angles that are associated to radii exceeding a high radial threshold r_0 (see Beirlant et al. [5], Section 9.4.1). In practice, we first have to apply a probability integral transform for standardizing $\mathbf{x_i}$ to $\mathbf{x_i^*}$. This requires knowledge or estimation of marginal data distributions F_j and the choice of a marginal target distribution with support in $[0, \infty)$ and tail asymptotics according to the standard Pareto distribution. We can estimate marginal distributions either nonparametrically via the empirical distribution function [22, 23] or via some other adequate marginal distribution, for instance with its tail behavior estimated according to the parametric tail model (2.3) from univariate extreme value theory [21]. The commonly used target is the unit Fréchet distribution. Then we fix a high radial threshold $r_0 > 0$ and define the empirical spectral measure

$$\hat{\rho}_n(\cdot) = r_0 n^{-1} \sum_{i=1}^{n} \chi_{[r_0,\infty)}(\mathrm{rad}(\mathbf{x_i^*}))\, \chi_{\|\mathbf{x_i^*}\|^{-1}\mathbf{x_i^*}}(\cdot). \tag{2.14}$$

When the number of radial exceedances $N = \sum_{i=1}^{n} \chi_{[r_0,\infty)}(\mathrm{rad}(\mathbf{x_i^*}))$ is at least 1, we can directly estimate the empirical spectral distribution $\rho_{p,n}$ by replacing the weight $r_0 n^{-1}$ in (2.14) by N^{-1}. In some cases, the enforcement of the moment constraints (2.12) upon the empirical spectral measure may be desirable for reasons of theoretical coherence. Therefore, we can wiggle atom weights by applying either an empirical likelihood procedure [23] or a Euclidean likelihood procedure [18]. The choice of r_0 is important, yet rarely straightforward. It should take into account bias and variance according to convergence in the domain of attraction setting. In practice, it is based on exploratory analyses, see also the remarks in Section 3.1 and in the application in Section 5.

The convergence rate in data towards the asymptotic distribution may suggest an aggregation function rad to use for the empirical estimator. Aggregations like the maximum, which retain many of the extreme observations outside the joint bivariate tail, might be advised against in case of slow convergence. On the other hand, aggregations like the minimum use only a limited amount of information about extreme value behavior by focusing on observations falling

into the joint tail. Convenient choices are the mean $\text{rad}(x_1, x_2) = 0.5(x_1 + x_2)$ or one of the marginal components such that $\text{rad}(x_1, x_2) = x_1$ or $\text{rad}(x_1, x_2) = x_2$, leading to $\rho(\mathcal{S}) = 1$ being invariant to the dependence structure; in such cases, the number of exceedances N carries no information about ρ_{rad}.

2.3. Related bivariate dependence measures

Alternative characterizations of bivariate extremal dependence are obtained as transformations of the spectral measure. Since the exponent function $V(\boldsymbol{u})$ is defined as $\mu^*([0, \boldsymbol{u}]^C)$ for $\boldsymbol{u} > \boldsymbol{0}$, the choice $\text{rad}(x_1, x_2) = \max(x_1/u_1, x_2/u_2)$ yields $\rho_{\text{rad}}(\mathcal{S}) = V(\boldsymbol{u})$. The extremal coefficient, given as $V(1, 1)$, is a commonly used tool to quantify the extremal dependence. It satisfies $V(1, 1) \in [1, 2]$ with 1 corresponding to full asymptotic dependence and 2 corresponding to asymptotic independence. It is equal to $\rho_{\text{rad}}(\mathcal{S})$ with $\text{rad} = \max$. The two following approaches are also often used.

2.3.1. Pickands dependence function

Pickands [35] showed that a bivariate distribution G^* is a MEVD if and only if there exists a convex function $A(\cdot)$ such that

$$G^*(x, y) = \exp\left(-(1/x + 1/y)A(x/(x + y))\right), \tag{2.15}$$

with $A(\cdot) : [0, 1] \to [1/2, 1]$ satisfying the following property: $\min(t, 1 - t) \leq A(t) \leq 1$ for $t \in [0, 1]$. The lower and upper bounds correspond to complete dependence and independence, respectively. The Pickands dependence function $A(\cdot)$ can be related to the spectral measure ρ_{rad} [5]:

$$A(t) = \int_{[0,1]} \frac{\max\left\{(1 - t)w, t(1 - w)\right\}}{\text{rad}(w, 1 - w)} \rho_{\text{rad}}(\mathrm{d}w). \tag{2.16}$$

2.3.2. Extremogram function

When (X_1, X_2) is in a maximum domain of attraction, the (upper) tail coefficient λ, defined as the conditional limit $\lim_{u \to \infty} \text{pr}(X_2^* > u \mid X_1^* > u) = \lim_{u \to \infty} \text{pr}(X_1^* > u \mid X_2^* > u)$, is often used to measure the extremal dependence. It is related to the extremal coefficient, $V(1, 1) = 2 - \lambda$. This coefficient λ can be defined as $\mu^*([\boldsymbol{1}, \boldsymbol{\infty}])$ and is equal to $\rho_{\text{rad}}(\mathcal{S})$ for $\text{rad} = \min$.

For stationary stochastic processes, functional dependence summaries based on the extremal coefficient or the tail coefficient can be considered. Given a space-time lag Δs and the aggregation function $\text{rad} = \max$, the extremal coefficient function is $\Delta s \mapsto V(1, 1; \Delta s) = \rho_{\max}(\mathcal{S}; \Delta s)$. For the tail coefficient, we obtain the *tail correlation function* (see Strokorb, Ballani and Schlather [46, 45], Fiebig, Strokorb and Schlather [27] for some theoretical characterizations)

$$\Delta s \mapsto \rho_{\min}(\mathcal{S}; \Delta s), \tag{2.17}$$

which is a variant of the so-called *spatial extremogram* [10] originally defined in the time series context [14]. The extremogram (2.17) has a simple interpretation as a correlation function. More generally, [14] have coined the term "extremogram" for not necessarily standardized data and expressions like $\mu(A \times B)$ or $\mu(A \times B)/\mu(A)$ with limit measures μ that are $(-\alpha)$-homogeneous with some $\alpha > 0$, where A and B are Borel sets that must be bounded away from 0. In our space-time setting with a univariate random field and marginally standardized data, it is most useful to specify $A = B = [u, \infty]$ with $u > 0$, leading to $\rho_{\min}(\mathcal{S}; \Delta s) = \mu^*(A \times B)/\mu^*(A) = u\,\mu^*([(u, u), (\infty, \infty)]) = \mu^*([(1, 1), (\infty, \infty)])$ as in the definition (2.17) that we will use in what follows.

3. The spectrogram

In this section, the extremal dependence structure of a stationarity process in a maximum domain of attraction is considered, leading to invariance with respect to space-time shifts. We apply the angular L^1-norm such that we can identify $\mathcal{S} = \{(w, 1 - w) \mid w \in [0, 1]\}$ with the unit interval $[0, 1]$. Given a space-time lag Δs and an aggregation function rad, we denote by $\rho_{\mathrm{rad}}(\,\cdot\,; \Delta s)$ the bivariate spectral measure of the limit process in (2.1).

Definition 3.1 (Spectrogram). We call *spectrogram* the collection of spectral measures $\{\rho_{\mathrm{rad}}(dw; \Delta s) \mid \Delta s \in \Delta K\}$ for point pairs in K, with $\Delta K = \{x - y \mid x, y \in K\}$. Using the notation $\rho_{\mathrm{rad}}(w; \Delta s) = \rho_{\mathrm{rad}}([0, w]; \Delta s)$ for the cumulative spectral measure function, we define the *cumulative spectrogram* as the set $\{\rho_{\mathrm{rad}}(w; \Delta s) \mid w \in \mathcal{S}, \Delta s \in \Delta K\}$.

Examples of isotropic spatial spectrograms are shown in Figure 1 for the mean aggregation $\mathrm{rad}(x_1, x_2) = 0.5(x_1 + x_2)$. For small distances, densities concentrate close to 0.5 since x_1 and x_2 tend to be close. At distance 0, there is discrete mass 1 in 0.5. If there is asymptotic independence when distances increase to infinity,
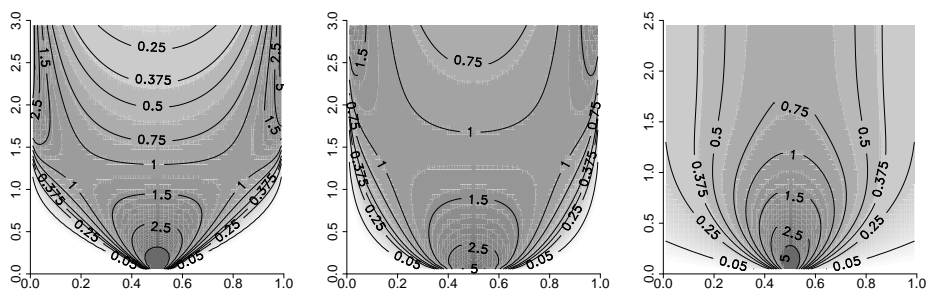
FIG 1. *Spectrogram densities. With respect to angle (horizontal axis) and distance (vertical axis) in isotropic models for the mean aggregation function* $\mathrm{rad}(x_1, x_2) = 0.5(x_1 + x_2)$*; from left to right: Gaussian extreme value model* $(\sigma_{11} = \sigma_{22} = 1, \sigma_{12} = 0)$*; Brown–Resnick model* $(\mathrm{shape} = 1, \mathrm{scale} = 1)$*; the continuous part of the extremal Gaussian model (stable correlation,* $\mathrm{shape} = 2, \mathrm{scale} = 1)$*.*

a very high value arises only in either x_1 or x_2, whereas the other component remains relatively small. Then, the mass of the spectral measures concentrates closer and closer to 0 or 1. The extremal Gaussian process stands apart since it is not mixing and always exhibits long-range dependence. Moreover, in addition to the density displayed in Fig. 1, discrete mass arises in the spectral measure for the angles 0 and 1: $\rho(\{0\}; \Delta s) = 0.25 (1 - \text{Cor}(\Delta s))$, where Cor is the correlation function of the Gaussian field $\mathbf{Q}$ in the construction (2.2), cf. Appendix A.

3.1. Empirical estimators and variance reduction

In the sequel, we will shortly write ρ instead of ρ_{rad} when there is no confusion. Given a sample of observations of a space-time process, an *empirical spectrogram* consists of a collection of empirical spectral measures for a selection of lags in space-time obtained according to the estimators presented in Section 2.2.3. Typically, we assume temporal stationarity of marginal distributions, which may require some pretransformation. Using a mixing assumption [28], we can then apply the empirical transform to standardize margins according to the $*$-transformation. Moreover, we can estimate univariate marginal tail parameters with procedures for dependent samples as for instance the composite likelihood of univariate margins.

Statements on the consistency and asymptotic normality of cumulative empirical spectral measures in the bivariate case have been deduced in the literature mentioned in Section 2.2.3 [21, 22, 23, 18], but they depend on the nature of marginal data transformations and the rate of convergence towards the asymptotic dependence structure. When margins are standardized based on the marginal empirical distribution functions, consistency is obtained when (2.5) holds and both r_0 and $r_0^{-1}n$ tend to infinity. Strong consistency further necessitates that $r_0^{-1}n/(\log \log n)$ tends to infinity [22, Theorem 1]. For asymptotic normality, more involved conditions are necessary to control the second-order behavior in convergence (2.5) [22, Theorem 2].

For an assessment of some important aspects of estimation variance, we here propose to work in a simplified setting where marginal distributions are assumed to be standardized and the limiting independence of the angular and the radial component is satisfied at finite levels. In practice, the standardization of margins based on some estimator of marginal distribution functions could induce additional bias and variance. In practical modeling, we must fix r_0 to define radial exceedance observations. When we do not have equality in the tail approximation (2.11) for a finite threshold r_0, we must consider how bias and variance depend on r_0. In the following, we look at variance from a more theoretical stance. Providing a general theory for bias terms is more involved since it would need assumptions on the second-order behavior in convergence (2.10). In practice, it is useful to check if estimated dependence structures change in a systematic way when we move r_0 to more or less extreme levels. Denote by $(X_i^*(s_0), X_i^*(s_0 + \Delta s))^T$ $(i = 1, \ldots, n)$ the normalized sample for a space-time lag Δs obtained from the marginal standardization of data. Assume that the corresponding pseudo-polar sample (R_i, W_i) with $R_i = \text{rad}(X_i^*(s_0), X_i^*(s_0 + \Delta s))$

and $W_i = X_i^*(s_0)/(X_i^*(s_0) + X_i^*(s_0 + \Delta s))$ contains $0 \leq N \leq n$ radial exceedances, indexed for $N \geq 1$ without loss of generality by $1, \ldots, N$ such that $R_i \geq r_0$ ($i = 1, \ldots, N$). We assign the dummy value $W_i = -1$ to the censored observations below the radial threshold with $R_i < r_0$, $i = N + 1, \ldots, n$. We make two assumptions:

(A1) The angles $\{W_i, \ i = 1, \ldots, n\}$ are independent and identically distributed according to a density $\tilde{f}(w; \Delta s)$.

(A2) The homogeneity property (2.8) holds exactly above the fixed radial threshold $r_0 > 0$ such that

$$\text{pr}((R_1, W_1) \in \text{d}(r, w)) = r^{-2}\text{d}r \times \rho(\text{d}w; \Delta s) \quad \text{for } r \geq r_0. \quad (3.1)$$

Hence we can write

$$\tilde{f}(w; \Delta s) = (1 - r_0^{-1}\rho(\mathcal{S}; \Delta s))\chi_{\{-1\}}(w) + r_0^{-1}f(w; \Delta s)\chi_{[0,1]}(w) \quad (3.2)$$

with $f(\,\cdot\,; \Delta s)$ the density of $\rho(\,\cdot\,; \Delta s)$ and $r_0^{-1}\rho(\mathcal{S}; \Delta s)$ the exceedance probability for the threshold r_0. Then $N \sim \text{Bin}(n, r_0^{-1}\rho(\mathcal{S}; \Delta s))$. For convenience of notation, we suppress the index Δs in the following.

Definition 3.2 (Empirical spectral measures)**.** Define the *empirical (cumulative) spectral measure*

$$\hat{\rho}(w) = \hat{\rho}(w; \text{rad}, r_0, n) = n^{-1}r_0 \sum_{i=1}^{N} \chi_{[0,w]}(w_i) \quad (3.3)$$

according to an aggregation function rad (cf. (2.14)). Further denote f' the density of a spectral measure ρ' according to another aggregation rad$'$. If $\{\mathbf{w} \in \mathcal{S} : \text{rad}'(\mathbf{w}) > 0\} \subset \{\mathbf{w} \in \mathcal{S} : \text{rad}(\mathbf{w}) > 0\}$, we use the transformation formula (2.13) to define an empirical spectral measure according to rad$'$ by

$$\breve{\rho}'(w) = \breve{\rho}(w; \text{rad}, \text{rad}', r_0, n) = n^{-1}r_0 \sum_{i=1}^{N} \text{rad}(w_i)^{-1}\text{rad}'(w_i)\chi_{[0,w]}(w_i). \quad (3.4)$$

We remark that $\breve{\rho}'$ uses inhomogeneous angle weights

$$n^{-1}r_0\frac{\text{rad}'(w_i)}{\text{rad}(w_i)}, \quad i = 1, \ldots, N.$$

We further define $N(w) = \text{card}\{W_i \in [0, w]\}$ for $w \in [0, 1]$ such that $\hat{\rho}(w) = n^{-1}r_0N(w)$.

Proposition 3.1 (Estimator properties)**.** *Under Assumptions (A1) and (A2), we have $N(w) \sim \text{Bin}(n, r_0^{-1}\rho(w))$. The standard estimator $\hat{\rho}$ in (3.3) satisfies*

$$\mathbb{E}\hat{\rho}(w) = \rho(w), \quad \text{var}(\hat{\rho}(w)) = n^{-1}[r_0\rho(w) - \rho(w)^2]. \quad (3.5)$$

The transformation estimator $\breve{\rho}'$ in (3.4) *satisfies* $\mathbb{E}\breve{\rho}'(w) = \rho'(w)$ *and*

$$\text{var}(\breve{\rho}'(w)) = n^{-1}\left(r_0\int_0^w [\text{rad}(v)^{-1}\text{rad}'(v)]^2 f(v)\text{d}v - \rho'(w)^2\right)$$

$$= n^{-1}\left(r_0\int_0^w \text{rad}(v)^{-1}\text{rad}'(v)\rho'(\text{d}v) - \rho'(w)^2\right). \qquad (3.6)$$

Due to Proposition 3.1, the estimators $\hat{\rho}$ and $\hat{\rho}'$ are (weakly) consistent.

Proof. The properties (3.5) of the standard estimator follow from those of the binomial distribution of $N(w)$. The transformation estimator (3.4) is directly related to the idea of importance sampling: based on the proposal distribution with density $\tilde{f}$, we estimate the moment $\rho'(w) = \int_0^1 g(v)dv$ with $g(v) = \text{rad}(v)^{-1}\text{rad}'(v)\chi_{[0,w]}(v)f(v)$. The unbiasedness of $\breve{\rho}'$ and the variance formula (3.6) follow from the standard results of the related theory, cf. Section 3.3 of [38]. $\qquad\square$

In comparison to the variance of the standard estimator $\hat{\rho}'$ of ρ' in (3.5), the variance of $\breve{\rho}'(B)$ for a measurable set $B \subset \mathcal{S}$ can increase if $\text{rad}(v)^{-1}\text{rad}'(v) \geq 1$ for $v \in B$ or can decrease if $\text{rad}(v)^{-1}\text{rad}'(v) \leq 1$ for $v \in B$. The latter case allows for variance reduction. For instance, we can estimate the extremogram associated to $\text{rad}' = \min$ by transforming empirical estimators obtained for the mean $\text{rad}(x_1, x_2) = 0.5(x_1 + x_2)$ or for some other appropriate aggregation. Moreover, the well-known Capéraà-Fougères estimator of Pickands' dependence function A in (2.16) can be represented as a transformation estimator $\breve{\rho}'$.

Example 3.1 (Capéraà-Fougères estimator of A). *We can write* $A(t) = t(1-t)V(t, 1-t) = V(1/(1-t), 1/t)$. *Hence by defining* $\text{rad}'(w) = \max((1-t)w, t(1-w))$ *and using the mean aggregation for* rad, *we obtain a transformation estimator of* A *given by*

$$\hat{A}(t) = \breve{\rho}'(1) = 2r_0 n^{-1}\sum_{i=1}^N \max\{(1-t)W_i, t(1-W_i)\}.$$

The estimator $\hat{A}(t)$ *satisfies convexity and is known as the Capéraà-Fougères estimator (Capéraà and Fougères [8], Beirlant et al. [5], p. 331).*

When the spectral measure ρ is equal to the spectral distribution ρ_p, i.e. $\rho([0,1]) = 1$, the distribution of N is independent of the dependence structure in ρ. This is the case for the mean aggregation and the marginal exceedance aggregation. It is then convenient to condition on the value of N and work with the density $\tilde{f}(w) = f(w)$ for W_i ($i = 1, \ldots, N$) and replace $n^{-1}r_0$ by N^{-1} in the direct estimator (3.3) and the subsequent calculations. This yields $\hat{\rho}(w) = N^{-1}N(w)$ with $N(w) \sim \text{Bin}(N, \rho(w))$. The properties of the binomial distribution yield $\mathbb{E}\hat{\rho}(w) = \rho(w)$ and $\text{Var}(\hat{\rho}(w)) = N^{-1}(\rho(w) - \rho(w)^2)$. The corresponding variance of the transformation estimator (3.6) becomes

$$\text{var}(\breve{\rho}'(w)) = N^{-1}\left(\int_0^w \left[\text{rad}(v)^{-1}\text{rad}'(v)\right]^2 f(v)\text{d}v - \rho'(w)^2\right)$$

$$= N^{-1} \left(\int_0^w \mathrm{rad}(v)^{-1}\mathrm{rad}'(v)\rho'(\mathrm{d}v) - \rho'(w)^2 \right). \qquad (3.7)$$

Since actually $N \sim \mathrm{Bin}(n, r_0^{-1})$ under the standing assumptions, we get

$$\mathbb{E}\mathrm{var}\left(\breve{\rho}'(w) \right) = \mathrm{var}\left(\breve{\rho}(w) \right) - n^{-1}(r_0 - 1)\rho'(w)^2.$$

Example 3.2 (Extremogram estimation)**.** *Estimation of the extremogram defined in* (2.17) *requires estimates of* $\rho_{\min}(\mathcal{S})$*. The estimator given by the standard empirical spectral measure* $\hat{\rho}_{\min}$ *has variance* $n^{-1}r_0\rho_{\min}(\mathcal{S})(1 - r_0^{-1}\rho_{\min}(\mathcal{S}))$*. Similar to replacing* nr_0^{-1} *by the binomial variable* N *when* $\rho(\mathcal{S}) = 1$*, we can improve this estimator by replacing* nr_0^{-1} *by the binomial variable* N_{marg}*, the number of marginal exceedances above* r_0 *such that* $N_{\mathrm{marg}} = \sum_{i=1}^n \chi_{(r_0,\infty)}(X^*(s_0))$ *and* $\mathbb{E}N_{\mathrm{marg}} = nr_0^{-1}$*. The resulting new estimator, written* $\overline{\rho}(\mathcal{S})$*, corresponds to a slightly adapted version of the empirical extremogram as proposed by Davis and Mikosch* [14]*, Section 3.3. Conditional on* N_{marg}*, its variance under the standing assumptions is* $\mathrm{Var}(\hat{\rho}(w)) = N_{\mathrm{marg}}^{-1}\rho_{\min}(\mathcal{S})(1 - \rho_{\min}(\mathcal{S}))$*.*

Using the transformation approach, we can define an interesting alternative based on $\mathrm{rad}(x_1, x_2) = 0.5(x_1 + x_2)$ *and* $\mathrm{rad}'(x_1, x_2) = \min(x_1, x_2)$*. Then, conditional on* N*, the variance of the transformation estimator is*

$$\mathrm{var}\breve{\rho}_{\min}(\mathcal{S}) = N^{-1} \left(\int_0^1 2\min(v, 1 - v)\rho_{\min}(\mathrm{d}v) - \rho_{\min}(\mathcal{S})^2 \right). \qquad (3.8)$$

We remark that $2\min(v, 1 - v) < 1$ *almost everywhere, such that the integral in* (3.8) *is smaller than* $\rho_{\min}$ *when* $\rho_{\min} \not\equiv 0$ *has a continuous density.*

3.2. Parametric minimum composite distance estimation

Current standard methods for inferring models in geostatistics of extremes focus on the information from the observed point pairs. In many cases, either analytical expressions of the likelihood cannot be calculated for higher dimensions or the application of multivariate estimation methods is impeded by the curse of dimensionality. We propose to perform minimum distance estimation of parameters by minimizing a distance measure between the empirical spectrogram and a family of theoretical spectrograms according to a parametric dependence structure of a max-stable process. Such distance measures can also serve for model selection and goodness-of-fit procedures, see the application in Section 5 for an example.

Definition 3.3 (Composite distance)**.** For $\Delta_k s$, $k = 1, \ldots, k_0$, a collection of observed lags with $n_k \geq 1$ observed data pairs respectively, we use some distance metric $\mathrm{dist}(\rho_{k,n_k}, \rho_{k,\boldsymbol{\theta}})$ between the empirical spectral measure and a and parametric family of spectral measures to define a *composite distance* $D(\boldsymbol{\theta}) = \sum_k \mathrm{dist}(\rho_{k,n_k}, \rho_{k,\boldsymbol{\theta}})$.

The vector of unknown parameters $\boldsymbol{\theta_0}$ characterizes the true asymptotic extreme value dependence structure in the data. Appendix A recalls some common

extreme value dependence models. Minimizing $D(\boldsymbol{\theta})$ yields the parameter estimate $\hat{\boldsymbol{\theta}}$. Given a fixed radial threshold $r_{0,k}$ leading to $0 \leq N_k \leq n_k$ radial exceedances among the n_k data pairs, we calculate an empirical spectral measure $\rho_{k,n_k}(\cdot) = r_{0,k} n_k^{-1} \sum_{\ell=1}^{N_k} \omega_{k,\ell} \chi_{W_{k,\ell}}(\cdot)$ from the observed angles $W_{k,\ell}$ with possibly inhomogeneous weights $\omega_{k,\ell}$. Inhomogeneous weights could be used to improve estimation efficiency and to reduce the computational burden of minimizing $D(\boldsymbol{\theta})$, for instance, by giving weight 0 to distant site pairs ("tapering", see Sang and Genton [39]). The exceedance probability $p_{k,\boldsymbol{\theta}} = r_{0,k}^{-1} \rho_{k,\boldsymbol{\theta}}(\mathcal{S})$ is associated to the threshold $r_{0,k}$.

3.2.1. Examples of composite distances

Example 3.3 (Pairwise likelihood). *The pairwise censored log-likelihood* $-D(\boldsymbol{\theta})$ *obtained from*

$$\text{dist}(\rho_{k,n_k}, \rho_{k,\boldsymbol{\theta}}) = -(n_k - N_k)\log(1 - p_{k,\boldsymbol{\theta}}) - \sum_{\ell=1}^{N_k} \omega_{k,\ell} \log(\rho_{k,\boldsymbol{\theta}}(\mathrm{d}W_{k,\ell})) \quad (3.9)$$

can be interpreted as the result of applying the Kullback-Leibler divergence for dist. When parameters to estimate are identifiable from pairwise information, pairwike likelihood estimators have asymptotic properties similar to those for the full likelihood under the usual regularity assumptions. An overview of asymptotic consistency and normality, tests and information criteria in the pairwise likelihood approach is given in [49] and the references therein. In Section 3.2.2, the expression of asymptotic variance and its estimation will be clarified.

The pairwise likelihood approach has been used for the special case of mean aggregation and the Brown–Resnick model by Engelke et al. [24].

In geostatistics of extremes, relatively simple yet robust estimation approaches based on an L^2-distance between a theoretical and an empirical version of summary statistics like the extremogram (or equivalently, the extremal coefficient function) have been applied and can be interpreted as a composite distance in our sense. Since such summary estimators of dependence are statistics that are not sufficient, these approaches use only partial information of the data. Instead, we here propose an L^2-distance between the cumulative measures ρ_{k,n_k} and $\rho_{k,\boldsymbol{\theta}}$ using the full information from observed angles.

Example 3.4 (Pairwise L^2-distance). *Assume that $\rho_{k,\boldsymbol{\theta}}(\mathcal{S}) = 1$ is invariant with respect to the dependence structure and that $\sum_{\ell=1}^{N_k} \omega_{k,\ell} = 1$. Failing this, if $\min_k N_k \geq 1$ and $\min_k \rho_{k,\boldsymbol{\theta}}(\mathcal{S}) > 0$, we could instead use the spectral distribution $\rho_{k,\boldsymbol{\theta}}(\cdot)/\rho_{k,\boldsymbol{\theta}}(\mathcal{S})$ and substitute*

$$\left(\sum_{\ell=1}^{N_k} \omega_{k,\ell}\right)^{-1} \sum_{\ell=1}^{N_k} \omega_{k,\ell} \chi_{W_{k,\ell}}(\cdot)$$

for $\rho_{k,n_k}(\cdot)$. We apply the statistic

$$N_k \int_0^1 (\rho_{k,n_k}(w) - \rho_{k,\boldsymbol{\theta}}(w))^2 \rho_{k,n_k}(\mathrm{d}w) = \sum_{\ell=1}^{N_k} \omega_{k,\ell} \left[\rho_{k,n_k}(W_{k,\ell}) - \rho_{k,\boldsymbol{\theta}}(W_{k,\ell}) \right]^2$$

$$(3.10)$$

for $\mathrm{dist}(\rho_{k,n_k}, \rho_{k,\boldsymbol{\theta}})$.

Whereas the likelihood-based distance (3.9) is related to the density representation of the spectrogram, the focus of the L^2-distance (3.10) is on the cumulative spectrogram.

3.2.2. *Asymptotic properties of composite distance estimators*

For the asymptotic variance and consistency in the general context of so-called extremum estimators, here obtained by minimizing the composite distance, we refer to Chapter 4 in Amemiya [1]. The following proposition states sufficient conditions for weak consistency of the minimum composite distance estimator. We denote n the sample size, for instance given as the number of observed temporal replications on a fixed configuration of spatial sites, and $D_n(\boldsymbol{\theta})$ the composite distance for a sample of size n.

Proposition 3.2 (Weak consistency)**.** *Let Θ be the compact parameter space, and assume that the asymptotic dependence structure of the data distribution is characterized by the true parameter vector $\boldsymbol{\theta}_0 \in \Theta$. Assume that there exists a sequence $d(n)$ such that $n^{-1}d(n)D_n(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta} \in \Theta$ and converges in probability to a non-stochastic limit $D_\infty(\boldsymbol{\theta})$, uniformly for $\boldsymbol{\theta} \in \Theta$. If $D_\infty(\boldsymbol{\theta}_0) < D_\infty(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$ different from the actual value $\boldsymbol{\theta}_0$, then any estimator $\hat{\boldsymbol{\theta}}_n$ satisfying $D_n(\hat{\boldsymbol{\theta}}_n) = \min_{\boldsymbol{\theta} \in \Theta} D_n(\boldsymbol{\theta})$ is weakly consistent, i.e. converges to $\boldsymbol{\theta}_0$ in probability.*

Proof. The assumptions made in Proposition 3.2 allow us to directly apply Theorem 4.1.1 of Amemiya [1]. □

Example 3.5 (Weak consistency of the L^2 estimator)**.** *Assume that we have n independent and identically distributed temporal replications of spatial observation vectors $\boldsymbol{x}_i$, $i = 1, \ldots, n$, for a fixed spatial design with m sites $s_1, \ldots, s_m$. We can fix a radial aggregation and use a standard empirical estimator of the pairwise spectral measures, indexed by $k = 1, \ldots, m(m-1)/2$ and defined as in (3.3). The pairwise estimators (3.3) are weakly consistent if the domain of attraction condition (2.5) holds and $d(n) = r_0(n) \to \infty$ as $n \to \infty$ (cf. Section 3.1). Then, $r_0(n)N_k/n \to \rho_{k,\boldsymbol{\theta}}(\mathcal{S})$ in probability and*

$$D_\infty(\boldsymbol{\theta}) = \sum_k \rho_{k,\boldsymbol{\theta}}(\mathcal{S}) \int_0^1 (\rho_{k,\boldsymbol{\theta}}(w) - \rho_{k,\boldsymbol{\theta}_0}(w))^2 \; \rho_{k,\boldsymbol{\theta}}(\mathrm{d}w).$$

If $\rho_{k,\boldsymbol{\theta}}(\mathcal{S}) \neq 0$ and $\rho_{k,\boldsymbol{\theta}}(w) \neq \rho_{k,\boldsymbol{\theta}_0}(w)$ for at least one k and values w in a neighborhood $(w_0 - \varepsilon, w_0 + \varepsilon)$ of $w_0 \in (0,1)$ with $\varepsilon > 0$, then Proposition 3.2 establishes the weak consistency of the estimator $\hat{\boldsymbol{\theta}}_n$.

Moreover, under regularity conditions necessary for asymptotic normality (which are not made explicit here due to their quite technical nature) the asymptotic variance is given by the information sandwich $\mathbf{M}_{\boldsymbol{\theta_0}} = \mathbf{I}_{\boldsymbol{\theta_0}}^{-1} \mathbf{J}_{\boldsymbol{\theta_0}} \mathbf{I}_{\boldsymbol{\theta_0}}^{-1}$ [1, Theorem 4.1.3]. The information matrix $\mathbf{I}_{\boldsymbol{\theta_0}}$ can be estimated by the numerically observed Hessian. The matrix $\mathbf{J}_{\boldsymbol{\theta_0}} = \mathbb{E}\mathbf{U}(\boldsymbol{\theta_0})\mathbf{U}(\boldsymbol{\theta_0})^T$ captures the (co)variance structure of the score vector $\mathbf{U} = (\partial D(\boldsymbol{\theta})/\partial \theta_j)_j$. Since $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ for the estimation $\hat{\boldsymbol{\theta}}$ of the true parameter $\boldsymbol{\theta_0}$, estimating $\mathbf{J}_{\boldsymbol{\theta_0}}$ is more intricate. Based on the assumption of temporal mixing, we can compute an estimate of the variance $\mathbf{J}_{\boldsymbol{\theta_0}}$ or directly of $\mathbf{M}_{\boldsymbol{\theta_0}}$ with subsampling techniques like the jackknife or the bootstrap (see Shao and Tu [43], Davison and Hinkley [16], Carlstein et al. [9]).

4. Simulation results for parametric estimation

We illustrate numerically the estimation efficiency of the parametric estimation approaches from Section 3.2 with respect to the number of exceedances when the excursion stability (3.1) is satisfied above the fixed radial threshold r_0. The radial mean aggregation $\mathrm{rad}(x_1, x_2) = 0.5(x_1 + x_2)$ is applied. We focus on the Gaussian extreme value process as an easily interpretable benchmark process, for which exact simulation of an excursion-stable version is possible (cf. Ferreira and de Haan [26]): for the sample region $K = [-3.5, a + 3.5]^2$ with $a > 0$, we simulate realizations from the field $(a + 2 \times 3.5)^2 Y f((\cdot) - S)$ with standard Pareto distributed Y, storm centers $S \sim \mathrm{Unif}(K)$ and the isotropic identity covariance matrix $\boldsymbol{\Sigma} = \mathrm{diag}(1,1)$ of the Gaussian density f. Neglecting storm centers outside K leads to edge effects which we reduce by using sample points s_j $(j = 1, \ldots, 25)$ from a uniform distribution on $[0, a]^2$ for estimation. For the sake of simpler exposition, we investigate only the case of simulating an isotropic model, and we study estimates $\hat{\sigma}_{11}$ of $\sigma_{11} = 1$ and $\hat{\sigma}_{12}$ of $\sigma_{12} = 0$. Results for $\hat{\sigma}_{22}$ are not reported since they are similar to those for $\hat{\sigma}_{11}$. Unreported results show that estimator performances in the case of geometric anisotropy allow us to draw the same conclusions on comparative estimator performance.

We choose $a \in \{1, 3, 5\}$ corresponding to strong, middle and weak spatial dependence and $nr_0^{-1} \in \{10, 20, 40, 200\}$ expected radial exceedances for bivariate spectral measures. We use the estimator $\hat{\rho}(w) = N^{-1} \sum_{j=1}^{N} \chi_{[0,w]}(W_j)$ with $\hat{\rho}(\mathcal{S}) = 1$ for estimating the empirical spectral measures and then apply the composite log-likelihood distance (CL) from (3.9) and the composite L^2-distance (L2) from (3.10). We further study the behavior of jackknife estimates of the standard errors based on 20 blocks each consisting of $n/20$ realizations of the process, i.e. *s.e.* $_{\mathrm{jackknife}} = \sqrt{19/20 \times \sum_{i=1}^{20} (\hat{\sigma}_{-i} - \hat{\sigma})^2}$ with $\sigma = \sigma_{11}$ or $\sigma = \sigma_{12}$ and $\hat{\sigma}_{-i}$ the estimator obtained from removing block i.

For comparative purposes, we also report estimates from a standard estimation procedure consisting of the (composite) bivariate partially censored log-likelihood (CPCL): for a standard-scale threshold u and bivariate standard-scale data (x_1, x_2), we use the bivariate likelihood contribution calculated from the

TABLE 1

Mean and standard errors of composite distance estimators (CL – likelihood distance (3.9)*, L2 – L^2-distance* (3.10)*, CPCL – composite partially censored likelihood* (4.1)*) for $\sigma_{11} = 1$ in the Gaussian extreme value model, based on* 500 *simulations; for CL and L^2 with mean and standard deviation of the jackknife estimate of the standard estimation errors calculated with* 20 *blocks*

nr_0^{-1}	a	CL	L2	CPCL
10	1	$1.32(0.72; 0.73(0.48))$	$1.06(0.65; 0.70(0.52))$	$1.28(0.62)$
	3	$1.07(0.24; 0.24(0.16))$	$0.90(0.20; 0.22(0.14))$	$1.11(0.33)$
	5	$1.02(0.12; 0.12(0.07))$	$0.91(0.12; 0.12(0.07))$	$1.03(0.21)$
20	1	$1.13(0.31; 0.34(0.25))$	$1.12(0.37; 0.39(0.28))$	$1.10(0.36)$
	3	$1.04(0.15; 0.15(0.10))$	$0.95(0.15; 0.16(0.10))$	$1.07(0.22)$
	5	$1.02(0.08; 0.08(0.05))$	$1.01(0.09; 0.08(0.05))$	$1.02(0.15)$
40	1	$1.04(0.21; 0.22(0.15))$	$0.98(0.25; 0.26(0.18))$	$1.05(0.22)$
	3	$1.02(0.10; 0.10(0.07))$	$0.97(0.10; 0.11(0.07))$	$1.02(0.14)$
	5	$1.01(0.06; 0.05(0.03))$	$0.98(0.06; 0.06(0.03))$	$1.02(0.10)$
200	1	$1.01(0.09; 0.08(0.06))$	$0.99(0.10; 0.10(0.07))$	$1.02(0.09)$
	3	$1.00(0.04; 0.04(0.03))$	$1.00(0.05; 0.05(0.03))$	$1.01(0.06)$
	5	$1.00(0.02; 0.02(0.02))$	$1.00(0.03; 0.03(0.02))$	$1.01(0.05)$

bivariate exponent function V as follows,

$$\begin{cases} 1 - V(u, u) & \text{if } \max(x_1, x_2) \leq u, \\ -\frac{\partial}{\partial x_1} V(x_1, u) & \text{if } x_1 > u, x_2 \leq u, \\ -\frac{\partial}{\partial x_2} V(u, x_2) & \text{if } x_1 \leq u, x_2 > u, \\ -\frac{\partial^2}{\partial x_1 \partial x_2} V(x_1, x_2) & \text{if } \min(x_1, x_2) > u, \end{cases} \tag{4.1}$$

see, for instance, [30] and the references therein and [15]. The latter approach focuses on the behavior of exceedances with respect to a constant thresholding field u and uses only partial information for simple exceedances outside the joint tail. We set $u = r_0$.

The results in Table 1 for σ_{11} and in Table 2 for σ_{12} show a slight tendency of CL to overestimate and of L2 to underestimate spatial dependence. The estimators appear robust over the range of chosen parameters. Bias and variance tend to 0 when the number of exceedances increases ($nr_0^{-1} : 10 \to 200$) or when spatial mixing improves ($a : 1 \to 5$). The jackknife estimates of the standard error behave similarly for both of the approaches CL and L2. In general, they are close to the actually observed estimation error and appear reliable. Throughout, estimation via CL or L2 tends to be (slightly) more efficient than via CPCL, particularly for moderate spatial dependence with $a = 3$ and $a = 5$, where fewer bivariate observations fall into the joint tail region.

5. Application to precipitation in the Cévennes region

We analyze spatial dependence among extremes in daily precipitation data for 21 gauge stations in the French Cévennes region with observations from 1959

TABLE 2

Mean and standard error of composite distance estimators (CL – likelihood distance (3.9),
$L2 – L^2$-distance (3.10), CPCL – composite partially censored likelihood (4.1)) for $\sigma_{12} = 0$
in the Gaussian extreme value model, based on 500 *simulations; for CL and L2 with mean
and standard deviation of the jackknife estimate of standard estimation errors calculated
with* 20 *blocks*

| nr_0^{-1}	a	CL	L2	CPCL
10	1	0.02(0.43; 0.71(0.36))	0.01(0.37; 0.70(0.43))	0.02(0.48)
	3	−0.01(0.15; 0.11(0.08))	0.00(0.12; 0.09(0.05))	0.00(0.23)
	5	0.00(0.07; 0.04(0.02))	0.00(0.06; 0.04(0.03))	0.00(0.13)
20	1	0.00(0.21; 0.17(0.12))	0.00(0.26; 0.22(0.12))	0.00(0.24)
	3	−0.01(0.10; 0.06(0.03))	−0.01(0.09; 0.05(0.03))	0.00(0.15)
	5	0.00(0.05; 0.02(0.01))	0.00(0.05; 0.02(0.01))	0.00(0.09)
40	1	0.00(0.14; 0.07(0.05))	0.00(0.17; 0.10(0.05))	0.01(0.15)
	3	0.00(0.07; 0.03(0.02))	0.00(0.07; 0.03(0.02))	0.00(0.10)
	5	0.00(0.04; 0.01(0.01))	0.00(0.03; 0.01(0.01))	0.00(0.07)
200	1	0.00(0.06; 0.02(0.01))	0.00(0.07; 0.02(0.01))	0.00(0.06)
	3	0.00(0.03; 0.01(0.01))	0.00(0.03; 0.01(0.01))	0.00(0.04)
	5	0.00(0.02; 0.00(0.00))	0.00(0.02; 0.00(0.00))	0.00(0.03)
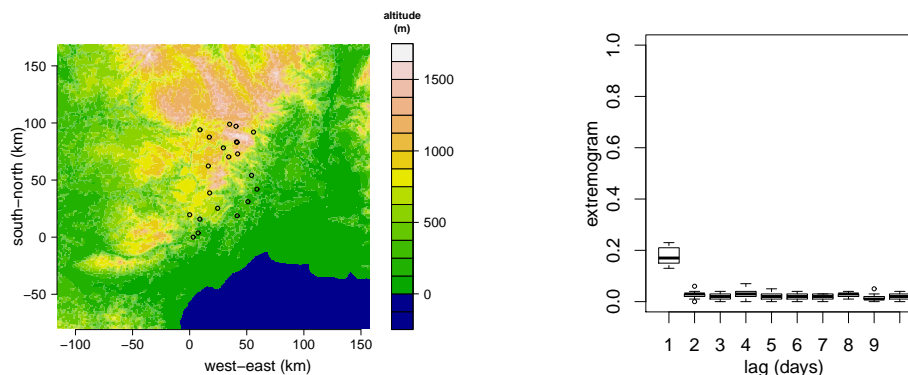
FIG 2. *The geographic setup of Cévennes gauge stations and temporal extreme value depen-
dence in the Cévennes precipitaton data. Left: Elevation map with Cévennes stations; right:
boxplots of extremogram values for the time series of the* 21 *stations with respect to temporal
lag and* $nr_0^{-1} = 100$ *expected marginal exceedances.*

to 2003 (data are available from Météo France). The elevation map on the
left in Fig. 2 shows the spatial setup of stations and the surrounding region
which is close to the Mediterranean sea. The Cévennes region is known for vi-
olent thunderstorms during autumn which often lead to inundations. We limit
modeling to the spatial extremal dependence structure between the intra-day
cumulated precipitation amounts at each site for the period from September 1
to January 31, leaving marginal behavior and temporal clustering out of con-
sideration and assuming stationary seasonal behavior. Removing days without
precipitation at any of the stations and days with missing data records, we
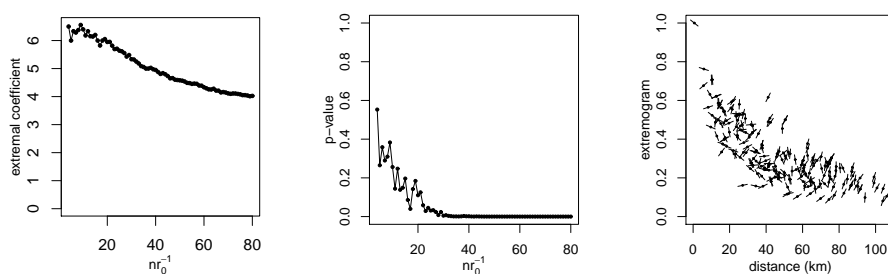keep $n = 4193$ days. On average, 45.3% of the sites are affected by precipi-

FIG 3. *Exploratory plots for modeling spatial extremes in the Cévennes precipitation data. Left: estimated extremal coefficient for the 21 stations with respect to number of expected marginal exceedances; middle: p-values of a homogeneity test with respect to number of expected marginal exceedances; right: the extremogram, transformed from the mean aggregation spectrogram with $nr_0^{-1} = 20$, with pins indicating the orientation of the space lag.*

tation. In order to take into account the presence or absence of precipitation, we modify the standard marginal empirical probability transform to the unit Fréchet distribution, leading to atomic mass at 0. Denote by $\hat{f}_0$ the observed proportion of 0-values at site s_j and write $n_{\neq 0} = (1 - \hat{f}_0)n$ for the number of nonzero observations. We transform a nonzero observation $X_i(s_j)$ $(i = 1, \ldots, n$ and $j = 1, \ldots, 21)$ with rank ran_i among all $n_{\neq 0}$ nonzero observations at s_j to $X_i^*(s_j) = -(1 - \hat{f}_0)^{-1}\{\log[\text{ran}_i/(n_{\neq 0} + 1)]\}^{-1}$. The factor $(1 - \hat{f}_0)^{-1}$ stretches the transformed nonzero data distribution such that $x\,\text{pr}(X^*(s_j) > x) \to 1$ as x tends to infinity. In $X_i^*(s_j)$, we keep the value 0 for the absence of precipitation. On the right in Fig. 2, the extremogram values with respect to temporal lags for the time series of each of the stations indicate little temporal extremal dependence. Estimation was achieved by the transformation estimator $\breve{\rho}_{\min}$ with $\text{rad}' = \min$ and the mean aggregation $\text{rad}(x_1, x_2) = 0.5(x_1 + x_2)$, as proposed in Example 3.2.

We apply a check for the homogeneity property (2.11), based on the tail behavior of the maximum component of a multivariate random vector, which has been proposed by [25] and [2]. If the domain of attraction condition (2.1) is satisfied, we observe $x\,\text{pr}(\max_{j=1,\ldots,21} X^*(s_j) > x) \approx V_{s_1,\ldots,s_{21}}(\mathbf{1})$ for large enough x, where $V_{s_1,\ldots,s_{21}}(\mathbf{1})$ is the extremal coefficient for all 21 sites. We therefore have to check if $x\,\text{pr}(\max_{j=1,\ldots,21}(X^*(s_j) > x)$ is approximately constant for large x. We find that the homogeneity assumption is realistic for our data only at very extreme levels. For illustration, the display on the left of Fig. 3 shows estimates $n^{-1}r_0 \times \text{card}\{\tilde{R}_i \mid \tilde{R}_i > r_0\}$ with $\tilde{R}_i = \max_{j=1,\ldots,21} X_i^*(s_j)$ $(i = 1, \ldots, n)$ of $V_{s_1,\ldots,s_{21}}(\mathbf{1})$ with respect to the expected number of marginal exceedances nr_0^{-1}. As r_0^{-1} tends to 0 in Fig. 3, the estimates seem to stabilize only very close to 0. The p-values reported on the middle display in Fig. 3 for the test procedure in [2] confirm this finding.

We now focus on the pairwise mean aggregation with $\text{rad}(x_1, x_2) = 0.5(x_1 + x_2)$. Based on the preceding findings, we fix the radial threshold r_0 corresponding to $nr_0^{-1} = 20$ expected radial exceedances. The right-hand display in Fig. 3

TABLE 3

Comparison of estimated extremal dependence models for the Cévennes precipitation data.
BR – Brown–Resnick model with power variogram; evt – extremal-t model with either the
stable or the Cauchy correlation model; CL – likelihood distance (3.9); L2 – L^2-distance
(3.10); standard error in parentheses (estimated by the jackknife technique); within-sample
and cross-validation mean absolute errors (err$_1$ and errcv$_1$) and mean squared errors (err$_2$
and errcv$_2$) for the extremogram

	BR		evt (stable)		evt (Cauchy)	
	CL	L2	CL	L2	CL	L2
m_{22}	0.45(0.12)	0.43(0.09)	0.43(0.10)	0.42(0.08)	0.47(0.13)	0.46(0.10)
m_{12}	$-0.37(0.05)$	$-0.28(0.06)$	$-0.32(0.05)$	$-0.28(0.06)$	$-0.3(0.05)$	$-0.28(0.07)$
η	— (—)	— (—)	9.6(2.2)	10(3.5)	9.5(2.2)	16(9.1)
scale	5.8(0.79)	7.1(1.1)	230(87)	210(72)	8.5(1.1)	10(1.4)
shape	0.80(0.05)	0.90(0.05)	0.88(0.05)	0.91(0.05)	0.07(0.02)	0.05(0.03)
err$_1$	231	220	239	**217**	248	225
errcv$_1$	241	233	248	**229**	257	237
err$_2$	404	370	418	**358**	444	395
errcv$_2$	437	411	449	**394**	482	440

shows the empirical extremogram, again estimated by the transformation estimator from Example 3.2. The display further presents the orientation of the spatial lag for each of the station pairs, suggesting an anisotropic dependence structure with vertically oriented lags taking higher values. Therefore we replace distances $\sqrt{(\Delta s)^T(\Delta s)}$ by $\sqrt{(\Delta s)^T \mathbf{M}(\Delta s)}$, where $\mathbf{M}$ is a symmetric matrix with diagonal elements 1, m_{22} and non-diagonal element m_{12}. We apply the minimum composite distance estimators to fit three parametric dependence structures, notably the Brown–Resnick model with power semi-variogram ("BR"; for the composite likelihood distance (3.9) and the L^2-distance(3.10)) and the extremal-t model with correlation of the stable or the Cauchy type ("evt"; for the composite likelihood (3.9) and the L^2-distance); see Appendix A for the description of correlation models. Since 0-observations arise with probability 0 in the Brown–Resnick model, we transform them to $[\log(n+1)]^{-1} > 0$ for the composite likelihood estimation of this model via (3.9); i.e., they are assigned rank 1. As in the simulation study, we use the estimator $\hat{\rho}(w) = N^{-1}\sum_{j=1}^{N} 1(W_j \leq w)$ with $\hat{\rho}(\mathcal{S}) = 1$ for estimating empirical spectral measures. Estimates of the parameter vector $\boldsymbol{\theta}$ (see Appendix A for details on the parametrization) are summarized in Table 3. The reported jackknife estimates of the standard errors are based on a reorganisation of the data into 41 blocks that we consider independent. Finally, we assess the goodness-of-fit by measuring the mean error between the empirical extremogram $\breve{\rho}_{\min}(\mathcal{S}; \Delta s)$ obtained from the mean aggregation empirical spectrogram and the fitted extremogram $\rho_{\min,\hat{\boldsymbol{\theta}}}(\mathcal{S}; \Delta s)$. Therefore, define

$$\text{err}_i = \sum_{\Delta s} \left[\left| \breve{\rho}_{\min}(\mathcal{S}; \Delta s) - \rho_{\min,\hat{\boldsymbol{\theta}}}(\mathcal{S}) \right| / \sqrt{\text{var}(\breve{\rho}_{\min}(\mathcal{S}; \Delta s))} \right]^i \quad (i = 1, 2), \quad (5.1)$$

with the variance formula $\text{var}(\breve{\rho}_{\min}(\mathcal{S}; \Delta s))$ from (3.8) to take into account the variability of $\breve{\rho}_{\min}(\mathcal{S}; \Delta s)$. Then err$_i$ $(i = 1, 2)$ are composite distances as de-
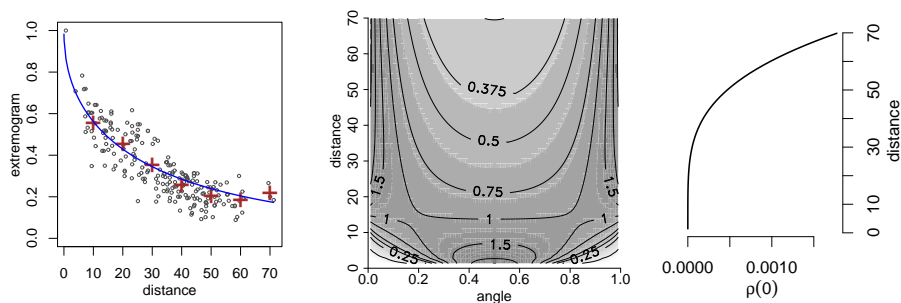
Fig 4. *The selected extremal-t model with stable correlation. Left: empirical and fitted extremogram ("+" indicates a binned empirical extremogram with bin size 10 km); middle: fitted spectrogram density for mean aggregation; right: fitted discrete mass $\rho(\{0\}; \Delta s)$. The distances are corrected for anisotropy by using $\sqrt{(\Delta s)' M (\Delta s)}$ instead of $\sqrt{(\Delta s)'(\Delta s)}$.*

scribed in Section 3.2. They quantify the mean absolute error ($i = 1$) or the mean squared error ($i = 2$) and further allow for a cross-validation procedure to check the predictive quality of a model: for each site, we remove the corresponding observations from the data set to create the training set consisting of 20 sites, we fit the model to the training set and we use the fitted model to predict the sum terms in (5.1) associated to distances Δs between the removed site and the 20 other sites. Finally, we sum up all these terms to obtain the corresponding cross-validation error errcv$_i$. The resulting error statistics reported in Table 3 are unanimously in favor of the extremal-t model with the stable correlation, estimated by minimizing the L^2-distance; see Figure 4 for a visual presentation of the fitted model. Fitted discrete masses $\rho_{\mathrm{mean}}(\{0\}; \Delta s) = \rho_{\mathrm{mean}}(\{1\}; \Delta s)$ are small, which adequately reflects the situation for the data where less than 0.7% of angles take one of the values 0 or 1 corresponding to extreme rainfall in one site and no rainfall in the other site.

6. Conclusion and perspectives

Representing multivariate limit distributions for threshold exceedances with the spectral measure opens a versatile toolbox for statistical analysis with both formal and visual tools. Further investigation could be dedicated to how we best include temporal dependence, which plays a more prominent role when switching from block maxima to exceedances in the original process. Some max-stable space-time process models are introduced in [13]. Moreover, current research efforts are directed towards estimation methods capable to exploit more than only bivariate information.

Acknowledgements

manuscript. The Cévennes precipitation data from Météo France were made available within the scope of a convention between Montpellier University and Météo France. The authors were partially funded by the ANR McSim and Gicc Miraccle projects.

Appendix A: Bivariate spectral measures for a selection of models

A general fomula to calculate spectral measure densities from the exponent function V was given by Coles and Tawn [12]. We use their Theorem 1 or a link to Pickands' dependence function [35] to calculate the expressions of measure functions, densities and discrete atoms in bivariate spectral measures. The sum-norm $\|\boldsymbol{x}\| = x_1 + x_2$ shall apply for the angle and the radius in the pseudo-polar transformation (2.7). We further assume that Pickands' dependence function $A(w) = w(1-w)V(w, 1-w)$ is twice differentiable on $(0,1)$. Then $A'(w) = \rho([0,w]) - 1$. Discrete mass at 0 is equal to $1 + A'(0^+)$; at 1 it is equal to $1 - A'(1^-)$. More generally, discrete mass at $w \in (0,1)$ is equal to $A'(w^+) - A'(w^-)$ when A is at least once differentiable.

For the standard normal distribution, the symbols Φ and φ shall denote the distribution function and the density respectively.

A.1. The extremal t process

Denote by Cor the correlation function of the underlying Gaussian process $\mathbf{Q}_{\text{gauss}}$ in the construction (2.2) and by η the power in $\mathbf{Q} = c_\eta (\mathbf{Q}_{\text{gauss}}{}^+)^\eta$. We use the notations $c = \sqrt{\eta + 1}/\sqrt{1 - \text{Cor}(\Delta s)^2}$ and $u(z_1, z_2) = c((z_1/z_2)^{1/\eta} - \text{Cor}(\Delta s))$. For student's t distribution, we denote the cumulative distribution function for df degrees of freedom by t_{df} and its density by f_{df}. The standard-scale exponent function is given as

$$V(z_1, z_2; \eta, \text{Cor}(\Delta s)) = z_1^{-1} t_{\eta+1}(u(z_2, z_1)) + z_2^{-1} t_{\eta+1}(u(z_1, z_2)).$$

For $\rho(w) = \rho([0, w]; \eta, \text{Cor}(\Delta s))$, we obtain

$$\rho(w) = 1 + t_{\eta+1}(u(w, 1-w)) + \frac{c}{\eta(1-w)} \left(\frac{w}{1-w}\right)^{1/\eta} f_{\eta+1}(u(w, 1-w))$$

$$- t_{\eta+1}(u(1-w, w)) - \frac{c}{\eta w} \left(\frac{1-w}{w}\right)^{1/\eta} f_{\eta+1}(u(1-w, w))$$

for $w \in (0,1)$, and we notice a discrete mass for the angles $w = 0$ and $w = 1$ on the axes: $\rho(\{0\}; \eta, \text{Cor}(\Delta s)) = \rho(\{1\}; \eta, \text{Cor}(\Delta s)) = t_{\eta+1}(-c\,\text{Cor}(\Delta s))$. The spectral density $f(w) = f(w; \eta, \text{Cor}(\Delta s))$ on $(0,1)$ can be written

$$f(w) = c_\eta \left[1 - \text{Cor}(\Delta s)^2\right]^{0.5(\eta+1)} [w(1-w)]^{1/\eta - 1}$$

$$\times \left[w^{2/\eta} + (1-w)^{2/\eta} - 2\text{Cor}(\Delta s)w^{1/\eta}(1-w)^{1/\eta}\right]^{-0.5(2+\eta)},$$

with $c_\eta = \frac{\Gamma(0.5(\eta+2))}{\eta^2 \pi^{1.5} \Gamma(0.5(\eta+1))}$ [33, page 114].

A.2. The Gaussian extreme value process

With $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{smallmatrix}\right)$ and $c^2 = (\Delta s)^T \boldsymbol{\Sigma}^{-1}(\Delta s)$, we have

$$V(z_1, z_2; c) = z_1^{-1}\Phi\left(0.5c + c^{-1}\log\frac{z_2}{z_1}\right) + z_2^{-1}\Phi\left(0.5c + c^{-1}\log\frac{z_1}{z_2}\right).$$

Setting $v_1(w) = 0.5c + c^{-1}\log((1-w)/w)$ and $v_2(w) = 0.5c + c^{-1}\log(w/(1-w))$ yields

$$\rho([0,w];c) = \Phi(-v_1(w)) + \Phi(v_2(w)) + c^{-1}\big\{-w^{-1}\varphi(v_1(w)) + (1-w)^{-1}\varphi(v_2(w))\big\}$$

and, using the the notation $\tilde{w} = 1 - w$, the spectral measure density

$$f(w; c) = \frac{1}{cw\tilde{w}}\left(\frac{1}{w}\varphi(v_1(w)) + \frac{1}{\tilde{w}}\varphi(v_2(w)) - \frac{v_1(w)}{cw}\varphi(v_1(w)) - \frac{v_2(w)}{c\tilde{w}}\varphi(v_2(w))\right).$$

A.3. The Brown–Resnick process

A Brown–Resnick type process [29] is characterized by a semi-variogram γ. Often, the semi-variogram of fractional Brownian motion $\gamma(\Delta s) = \|\psi^{-1}\Delta s\|^\alpha$ with the shape $\alpha \in (0,2]$ and the scale $\psi > 0$ (*power semi-variogram*) is used, leading to the classical Brown–Resnick process [6]. Then

$$\begin{aligned} V(z_1, z_2; \Delta s) = & z_1^{-1}\Phi\left(0.5\sqrt{\gamma(\Delta s)} + \sqrt{\gamma(\Delta s)}^{-1}\log\frac{z_2}{z_1}\right) \\ & + z_2^{-1}\Phi\left(0.5\sqrt{\gamma(\Delta s)} + \sqrt{\gamma(\Delta s)}^{-1}\log\frac{z_1}{z_2}\right). \end{aligned}$$

The Gaussian extreme value process arises as special case for $\alpha = 2$. Consequently, the respective expressions carry over after replacing c by $\sqrt{\gamma(\Delta s)}$.

A.4. Correlation models

- stable: $\mathrm{Cor}(\Delta s) = \exp(-\|\mathrm{scale}^{-1}\Delta s\|^{\mathrm{shape}})$ with scale > 0, $0 < \mathrm{shape} \leq 2$,
- Cauchy: $\mathrm{Cor}(\Delta s) = [1 + (\|\mathrm{scale}^{-1}\Delta s\|)^2]^{-\mathrm{shape}}$ with scale > 0, shape > 0.

For more details, see [40].

References

[1] AMEMIYA, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge, MA.

[2] AULBACH, S. and FALK, M. (2012). Testing for a generalized Pareto process. *Electron. J. Stat.* **6** 1779–1802. MR2988464

[3] BACRO, J. N. and GAETAN, C. (2014). Estimation of spatial max-stable models using threshold exceedances. *Stat. Comput.* **24** 651–662. MR3223547

[4] BALKEMA, A. A. and RESNICK, S. I. (1977). Max-infinite divisibility. *J. Appl. Probab.* **14** 309–319. MR0438425

[5] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications.* Wiley. MR2108013

[6] Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probab.* 732–739. MR0517438

[7] Buishand, T. A., de Haan, L. and Zhou, C. (2008). On spatial extremes: With application to a rainfall problem. *Ann. Appl. Stat.* **2** 624–642. MR2524349

[8] Capéraà, P. and Fougères, A. L. (2000). Estimation of a bivariate extreme value distribution. *Extremes* **3** 311–329.

[9] Carlstein, E., Do, K. A., Hall, P., Hesterberg, T. and Künsch, H. R. (1998). Matched-block bootstrap for dependent data. *Bernoulli* 305–328. MR1653268

[10] Cho, Y., Davis, R. A. and Ghosh, S. (2014). Asymptotic properties of the empirical spatial extremogram. *arXiv:1408.0412*.

[11] Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer. MR1932132

[12] Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *J. R. Stat. Soc. B* **53** 377–392. MR1108334

[13] Davis, R. A., Klüppelberg, C. and Steinkohl, C. (2013). Max-stable processes for modeling extremes observed in space and time. *J. Korean Stat. Society* **42** 399–414. MR3255398

[14] Davis, R. A. and Mikosch, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli* **15** 977–1009. MR2597580

[15] Davison, A. C. and Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proc. R. Stat. Soc. A* **468** 581–608. MR2874052

[16] Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application.* Cambridge Univ. Press, Cambridge, UK. MR1478673

[17] Davison, A. C., Padoan, S. and Ribatet, M. (2012). Statistical modelling of spatial extremes. *Stat. Sci.* **27** 161–186. MR2963980

[18] de Carvalho, M., Oumow, B., Segers, J. and Warchol, M. (2013). A Euclidean likelihood estimator for bivariate tail dependence. *Commun. Stat. (Theory and Methods)* **42** 1176–1192. MR3031275

[19] de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.* **12** 1194–1204. MR0757776

[20] Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface* **8** 9–17.

[21] Einmahl, J. H. J., de Haan, L. and Sinha, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stoch. Processes Appl.* **70** 143–171. MR1475660

[22] Einmahl, J. H. J., de Haan, L. and Piterbarg, V. I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Stat.* **29** 1401–1423. MR1873336

[23] Einmahl, J. H. J. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Stat.* **37** 2953–2989. MR2541452

[24] ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Huesler-Reiss distributions and Brown-Resnick processes. *J. R. Statist. Soc. B* **77** 239–265. MR3299407

[25] FALK, M. and MICHEL, R. (2009). Testing for a multivariate generalized Pareto distribution. *Extremes* **12** 33–51. MR2480722

[26] FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process. *Bernoulli* **20** 1717–1737. MR3263087

[27] FIEBIG, U., STROKORB, K. and SCHLATHER, M. (2014). The realization problem for tail correlation functions. *arXiv:1405.6876*.

[28] HUSER, R. and DAVISON, A. C. (2014). Space-time modelling of extreme events. *J. R. Stat. Soc. B* **76** 439–461. MR3164873

[29] KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440

[30] LEDFORD, A. W. and TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169. MR1399163

[31] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *Soc. Ind. Appl. Math. Rev.* **10** 422–437. MR0242239

[32] OPITZ, T. (2013a). Extremal t processes: Elliptical domain of attraction and a spectral representation. *J. Multivar. Anal.* **122** 409–413. MR3189331

[33] OPITZ, T. (2013b). Extrêmes multivariés et spatiaux: Approches spectrales et modèles elliptiques. PhD Thesis, 146 pages.

[34] PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Am. Stat. Assoc.* **105** 263–277. MR2757202

[35] PICKANDS, J. (1981). Multivariate extreme value theory. *Bull. Int. Stat. Inst.* **1** 859–878. MR0820979

[36] RESNICK, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer, Berlin. MR0900810

[37] RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, Berlin. MR2271424

[38] ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. Springer, Berlin. MR2080278

[39] SANG, H. and GENTON, M. (2014). Tapered composite likelihood for spatial max-stable models. *Spatial Statistics* **8** 86–103.

[40] SCHABENBERGER, O. and GOTWAY, C. A. (2004). *Statistical Methods for Spatial Data Analysis.* Chapman & Hall/CRC, Boca Raton. MR2134116

[41] SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44. MR1947786

[42] SCHLATHER, M. and TAWN, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90** 139–156. MR1966556

[43] SHAO, J. and TU, D. (1995). *The Jackknife and Bootstrap.* Springer, Berlin. MR1351010

[44] SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.

[45] STROKORB, K., BALLANI, F. and SCHLATHER, M. (2014). Systematic co-occurrence of tail correlation functions among max-stable processes. *arXiv:1402.4632*.

[46] STROKORB, K., BALLANI, F. and SCHLATHER, M. (2015). Tail correlation functions of max-stable processes. To appear in Extremes.

[47] R CORE TEAM (2013). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

[48] THIBAUD, E. and OPITZ, T. (2014). Efficient inference and simulation for elliptical Pareto processes. Submitted.

[49] VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21** 5–42. MR2796852