Modeling spatial tail dependence with Cauchy convolution processes

Pavel Krupskii and Raphaël Huser

University of Melbourne, Parkville, Victoria, 3010, Australia
Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering
(CEMSE) Division, King Abdullah University of Science and Technology (KAUST),
Thuwal 23955-6900, Saudi Arabia
e-mail: pavel.krupskiy@unimelb.edu.au*;* raphael.huser@kaust.edu.sa

Abstract: We study the class of dependence models for spatial data obtained from Cauchy convolution processes based on different types of kernel functions. We show that the resulting spatial processes have appealing tail dependence properties, such as tail dependence at short distances and independence at long distances with suitable kernel functions. We derive the extreme-value limits of these processes, study their smoothness properties, and detail some interesting special cases. To get higher flexibility at sub-asymptotic levels and separately control the bulk and the tail dependence properties, we further propose spatial models constructed by mixing a Cauchy convolution process with a Gaussian process. We demonstrate that this framework indeed provides a rich class of models for the joint modeling of the bulk and the tail behaviors. Our proposed inference approach relies on matching model-based and empirical summary statistics, and an extensive simulation study shows that it yields accurate estimates. We demonstrate our new methodology by application to a temperature dataset measured at 97 monitoring stations in the state of Oklahoma, US. Our results indicate that our proposed model provides a very good fit to the data, and that it captures both the bulk and the tail dependence structures accurately.

MSC2020 subject classifications: Primary 62H05; secondary 60G70, 62H11.
Keywords and phrases: Copula, extreme-value model, kernel convolution process, short-range spatial dependence, spatial process, tail dependence.

Received September 2021.

Contents

1. Introduction

Assessment of environmental risk associated with unprecedented air or sea temperatures [12, 28, 34], extreme flooding [10, 32, 66], strong wind gusts [17, 35, 53], or high air pollution levels [68, 69] requires the computation of joint tail probabilities. Because the process of interest is always observed at a finite set of monitoring sites, spatial modeling is needed whenever the required probabilities involve one or more unobserved locations, and the assumed tail dependence structure plays a crucial role in estimating these risks.

　　Gaussian processes have been widely used in the literature to model spatio-temporal dependence, because they are computationally convenient and they are parameterized using various types of covariance functions that can capture features such as the dependence range and the smoothness of realizations (see, e.g., Gneiting [23] and Gneiting et al. [24]). However, Gaussian models have restrictive symmetries and cannot capture strong tail dependence that is often found in spatial data, which makes them unsuitable when the main interest lies in the tails. More flexible yet computationally feasible spatial models are required. To circumvent the limitations of multivariate normality, copula models have become increasingly popular and have found a wide range of environmental applications in geology [25], hydrology [3, 4] and climatology [18], among others. A copula is simply defined as a multivariate cumulative distribution function (CDF) with standard uniform $\mathrm{Unif}(0,1)$ marginals. Sklar [62] showed that for any continuous d-dimensional CDF F with marginals $F_1, \ldots, F_d$ there exists a unique copula C such that $F(z_1, \ldots, z_d) = C\{F_1(z_1), \ldots, F_d(z_d)\}$. A random vector $(Z_1, Z_2)^\top$ with margins F_1, F_2 and copula C is said to be upper tail-dependent if the limit

$$\lambda_U = \lim_{u \to 1} \Pr\{Z_1 > F_1^{-1}(u) \mid Z_2 > F_2^{-1}(u)\} = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u}, \quad (1)$$

exists and is positive, i.e., $\lambda_U > 0$, and is upper tail-independent if $\lambda_U = 0$. An analogous definition holds for the lower tail based on the coefficient

$\lambda_L = \lim_{u \to 0} \Pr\{Z_1 \leq F_1^{-1}(u) \mid Z_2 \leq F_2^{-1}(u)\} = \lim_{u \to 0} C(u, u)/u$. A two-dimensional copula C is said to be tail-symmetric if $\lambda_L = \lambda_U$ and is reflection-symmetric if $C = C^R$, where $C^R(u_1, u_2) := -1 + u_1 + u_2 + C(1 - u_1, 1 - u_2)$ is the reflected copula. While multivariate normal vectors are always reflection-symmetric, as well as tail-independent when the correlation is less than one [47], other copula models can be used to construct flexible multivariate distributions with arbitrary marginals and various tail dependence structures. Certain copula families, such as vine models [1, 46], are very flexible but lack interpretability with spatial data. By contrast, factor copula models [45] have been proposed as flexible models capturing non-Gaussian features like reflection and/or tail asymmetry and strong tail dependence, and they can be naturally extended to the spatial context; see Krupskii et al. [44] and Krupskii and Genton [43]. However, since these models are built from common underlying random factors affecting all spatial sites simultaneously, they are unable to capture full independence at large distances. Hence, these models are only suitable for spatial data observed over small and homogeneous spatial regions.

To accurately model the data's tail behavior, an alternative approach might be to rely on models justified by Extreme-Value Theory; see Davison and Huser [13] for a general review on statistics of extremes, and Davison et al. [14], Davison et al. [15] and Huser and Wadsworth [38] for reviews on spatial extremes. Classical extreme-value models, such as max-stable processes—characterized by extreme-value copulas—and generalized Pareto processes, stem from asymptotic theory for block maxima and high threshold exceedances, respectively. However, despite their popularity, these extremal models suffer from several drawbacks: first, these limit models cannot capture weakening of dependence for increasing quantile levels. In particular, with Pareto processes, the conditional exceedance probability $\{1 - 2u + C(u, u)\}/(1 - u)$ that appears in (1) is constant in u above a certain uniform quantile [58], while with extreme-value copulas, one has $C^k(u_1^{1/k}, u_2^{1/k}) = C(u_1, u_2)$ for all $(u_1, u_2)^\top \in [0, 1]^2$, $k = 1, 2, \ldots$. While these strong restrictions on the form of the copula C are indeed justified asymptotically, they may not be satisfied at finite levels (always considered in finite samples), and this has major implications in practice for assessing the risk of simultaneous extremes over a spatial region. Several recent studies have indeed shown that environmental extreme events are often found to be more spatially localized when they are more extreme [37], a property that these asymptotic extreme-value models cannot capture. Second, a consequence of these stability properties is that these extreme-value models are always tail-dependent unless they are exactly independent. As a result, non-trivial extreme-value models cannot capture independence at large distances, which makes them unsuitable over large spatial domains similarly to factor copula models. Third, these models have complicated likelihood functions that are costly to evaluate for inference [11, 21, 33, 56], though recent progress on graphical models for Pareto processes opens the door to higher-dimensional likelihood inference [16]. Finally, because these models describe the limiting behavior of extreme events, they are typically fitted using only a small fraction of observations, thus wasting a lot of informa-

tion that might potentially be useful for accurate estimation of unknown model parameters.

To circumvent limitations of asymptotic extreme-value models, recent work has focused on the development of "sub-asymptotic" models for extremes that provide additional tail flexibility at finite levels, and that can smoothly bridge both tail dependence classes under the same parametrization; see Bopp et al. [6], Hua [30], Huser et al. [35, 33, 36], Su and Hua [65], Wadsworth and Tawn [70], Wadsworth et al. [72], and the recent review paper Huser and Wadsworth [38]. More recently, an alternative approach based on single-site conditioning has been proposed by Wadsworth and Tawn [71] to flexibly capture various forms of tail dependence structures, although the proposed model does not possess a convenient unconditional formulation. However, except for the rather artificial max-mixture model of Wadsworth and Tawn [70] and the conditional extremes model of Wadsworth and Tawn [71], sub-asymptotic models proposed in the literature cannot capture changes in the tail dependence class as a function of distance between sites. In other words, while it is reasonable to expect that strong tail dependence prevails at short distances and tail independence at larger distances, most proposed models in the literature assume either tail independence or tail dependence between all pairs of sites at any distance, and do not capture full independence as the distance between sites increases arbitrarily.

In this paper, we address these shortcomings by considering process convolutions of the form

$$Z(\boldsymbol{s}) = \int_{\mathbb{R}^q} k(\boldsymbol{s}, \boldsymbol{s}^*) W(\mathrm{d}\boldsymbol{s}^*), \qquad \boldsymbol{s} \in \mathbb{R}^q, \tag{2}$$

and variants thereof, where $k(\boldsymbol{s}, \boldsymbol{s}^*) \geq 0$ is a nonnegative integrable kernel function (i.e., such that $\int_{\mathbb{R}^q} k(\boldsymbol{s}, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^* < \infty$ for all $\boldsymbol{s} \in \mathbb{R}^q$) and W is a particular Lévy process [59] with independent increments. Note that the integral in (2) is a deterministic integral of Riemann-Stieltjes type, here calculated for each sample path. For simplicity, we hereafter assume that $q = 2$ (unless specified otherwise), although most of our results and models can easily be extended to the case $q = 1$ or $q > 2$. Process convolutions have been used extensively to model spatial data; see for example Calder and Cressie [9], Higdon [29], Paciorek and Schervish [55] and Zhu and Wu [74]. However, the marginal CDF of $W(\boldsymbol{s})$, F_W, is usually assumed to be Gaussian, thus leading to a Gaussian process $Z(\boldsymbol{s})$ in (2), which does not have tail dependence. Trans-Gaussian processes $Z^*(\boldsymbol{s}) = t\{Z(\boldsymbol{s})\}$, where $t(\cdot)$ is a monotone increasing transformation, could be used to model spatial data with non-Gaussian marginals [8], but the process $Z^*(\boldsymbol{s})$ still possesses the Gaussian copula and thus has the same restrictive dependence structure. More flexible tail structures can be obtained by considering non-Gaussian distributions for F_W in (2) [40, 52], and the unpublished manuscript of Opitz [54] investigates the dependence properties of the resulting process for an indicator kernel $k(\boldsymbol{s}, \boldsymbol{s}^*) = \mathbf{1}\{\boldsymbol{s}^* \in \mathcal{A}(\boldsymbol{s})\}$ defined in terms of a hypograph indicator set $\mathcal{A}(\boldsymbol{s})$. In this paper, we consider instead general classes of kernels in (2) but assume that F_W is the Cauchy CDF. Because the Cauchy distribution is stable, the process $Z(\boldsymbol{s})$ in (2) remains Cauchy, which facilitates

inference and theoretical calculations, and thanks to the heavy-tailedness of F_W, we will see that the resulting copula can have interesting tail dependence structures depending on the choice of the kernel k. While it would also be interesting to consider other types of (potentially skewed) stable distributions for F_W, we here focus on the Cauchy family, which yields tractable inference and already provides a fairly rich class of models. In this paper, we study the dependence properties of these Cauchy convolution process models under general forms of kernel functions k, and we derive their limiting extreme-value copulas, which turn out to be characterized in terms of a moving maximum representation. We show that a wide class of (existing or new) extremal dependence structures can be obtained, but unlike these limit extreme-value models, Cauchy convolution processes have a more flexible sub-asymptotic behavior. Moreover, when the kernel function is compactly supported, the resulting process $Z(\boldsymbol{s})$ has the appealing property of local tail dependence, in the sense that it possesses strong tail dependence at short distances only and full independence at larger distances. We also propose a new spatial sum-mixture model, where the proposed Cauchy process (2) is mixed with a lighter-tailed Gaussian process. This allows to get even higher flexibility at sub-asymptotic levels and separately control bulk and tail properties, while retaining the same extremal dependence structure. Some new theoretical results on the "smoothness" properties of the resulting extreme-value copulas are also derived.

To make inference for the Cauchy process convolution model (2) or its more flexible spatial mixture extension efficiently, we develop a fast estimation approach that consists in matching suitable empirical and model-based summary statistics. Compared to likelihood-based inference, this approach allows us to easily fit our models in higher dimensions. Unlike most extreme-value inference methods, which typically rely on extreme data from one tail only and discard all the other observations, we opt here for fitting the proposed model to the whole dataset from low to high quantiles (i.e., without applying any kind of censoring) for several reasons: first and foremost, we are not interested in modeling extremes only, but the whole distribution, as joint moderately large events from the "bulk" may in practice be as critical for risk assessment as individual severe extreme events; second, our approach based on the complete dataset makes full use of the available information, thus getting more accurate parameter estimates; and lastly, our proposed model is highly flexible in the bulk and the tails, so it can generally provide a good fit overall, without compromising any part of the distribution.

The rest of the paper is organized as follows. In Section 2, we detail our proposed Cauchy convolution model, study its dependence properties and tail behavior, derive some interesting special cases, and discuss approximate simulation algorithms for the Cauchy convolution process itself and its extreme-value limits. We also study our proposed spatial mixture process, and explore its improved flexibility. In Section 3, we describe our proposed inference approach, while in Section 4, we report the results of a simulation study, and we illustrate the proposed methodology by application to a temperature dataset from the state of Oklahoma, US. Section 5 concludes with a discussion and some

perspective on future research. All proofs are deferred to Appendix A.

2. Modeling

2.1. Cauchy convolution processes and their extreme-value limits

We consider the process convolution (2), where W is a Lévy process with independent Cauchy increments, i.e., such that $W(\mathrm{d}\boldsymbol{s}^*) \sim \mathrm{Cauchy}(\mathrm{d}\boldsymbol{s}^*)$ are independent increments, where $\mathrm{Cauchy}(c)$ is the Cauchy distribution with scale parameter c and probability density function (pdf) $f_{\mathcal{C}}(z;c) = \pi^{-1}c/(z^2 + c^2)$, $z \in \mathbb{R}$. By some slight abuse of notation, we here write $\mathrm{Cauchy}(\mathrm{d}\boldsymbol{s}^*)$ to denote the Cauchy distribution with infinitesimal scale $c = |\mathrm{d}\boldsymbol{s}^*| > 0$, where $\mathrm{d}\boldsymbol{s}^*$ is some infinitesimal spatial unit with area $|\mathrm{d}\boldsymbol{s}^*|$. Note that the Cauchy distribution is a sum-stable and infinitely divisible distribution, such that the process $Z(\boldsymbol{s})$ in (2) is well-defined and may be represented as

$$Z(\boldsymbol{s}) = \int_{\mathbb{R}^q} k(\boldsymbol{s}, \boldsymbol{s}^*)\tilde{W}(\boldsymbol{s}^*)\mathrm{d}\boldsymbol{s}^*, \quad \boldsymbol{s} \in \mathbb{R}^q,$$

where $\tilde{W}(\boldsymbol{s})$ is a Cauchy white noise process with $\tilde{W}(\boldsymbol{s}) \sim \mathrm{Cauchy}(1)$.

The finite-dimensional distributions of the Cauchy process convolution $Z(\boldsymbol{s})$ in (2) are not tractable in the general case. However, it is possible to derive the extreme-value (EV) limit of this process as Proposition 1 below shows. Before stating this result, we first recall some fundamentals about extreme-value theory. Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ be a d-dimensional random vector with margins $F_1, \ldots, F_d$ and copula C as defined in the Introduction (Section 1). The copula C^n describes the copula of the vector of componentwise maxima from i.i.d. copies $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})^\top$ of $\boldsymbol{X}$, $i = 1, \ldots, n$, i.e., $\boldsymbol{M}_n = (M_{n1}, \ldots, M_{nd})^\top$ with $M_{nj} = \max(X_{1j}, \ldots, X_{nj})$. Extreme-value copulas, denoted C_{EV}, describe the class of dependence structures that arise as limits of $\boldsymbol{M}_n$ (when properly renormalized), i.e.,

$$C_{\mathrm{EV}}(u_1, \ldots, u_d) = \lim_{n \to \infty} C^n(u_1^{1/n}, \ldots, u_d^{1/n}), \qquad (u_1, \ldots, u_d)^\top \in [0,1]^d. \quad (3)$$

It can be shown that extreme-value copulas are such that for any $k = 1, 2, \ldots$, one has $C_{\mathrm{EV}}(u_1, \ldots, u_d) = C_{\mathrm{EV}}^k(u_1^{1/k}, \ldots, u_d^{1/k})$, $(u_1, \ldots, u_d)^\top \in [0,1]^d$, and they can be characterized as

$$C_{\mathrm{EV}}(u_1, \ldots, u_d) = \exp\{-\ell(-\log u_1, \ldots, -\log u_d)\}, \qquad (u_1, \ldots, u_d)^\top \in [0,1]^d, \quad (4)$$

where ℓ is called the *stable (upper) tail dependence function* and completely determines the limiting extremal dependence structure in the upper tail. From (3) and (4), the stable tail dependence function can be expressed as the limit $\ell(w_1, \ldots, w_d) = \lim_{n \to \infty} n\{1 - C(1 - w_1/n, \ldots, 1 - w_d/n)\}$, and it lies between the bounds $\max(w_1, \ldots, w_d) \leq \ell(w_1, \ldots, w_d) \leq \sum_{j=1}^d w_j$, $w_1, \ldots, w_d \geq 0$, corresponding to perfect dependence and independence, respectively. Therefore,

extreme-value copulas cannot be negatively dependent. As Cauchy processes are reflection-symmetric, their extremal dependence structures are identical in both tails, so hereafter we shall simply refer to ℓ as the *stable tail dependence function.* More details on extreme-value theory, copula models, and their properties can be found, e.g., in Davison and Huser [13], Gudendorf and Segers [26], and Segers [61].

Proposition 1 (Stable tail dependence function of the Cauchy process)**.** Consider the Cauchy process convolution defined as in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{\text{i.i.d.}}$ Cauchy$(\mathrm{d}\boldsymbol{s}^*)$. For any collection of sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d \in \mathbb{R}^2$, we write concisely $Z_1 = Z(\boldsymbol{s}_1), \ldots, Z_d = Z(\boldsymbol{s}_d)$. Assume that $k(\boldsymbol{s}, \boldsymbol{s}^*) \leq k_{\max} < \infty$ is a nonnegative bounded integrable kernel function. Let $\ell(w_1, \ldots, w_d) : [0, \infty)^d \mapsto [0, \infty)$ be the stable tail dependence function of the random vector $(Z_1, \ldots, Z_d)^\top$, then

$$\ell(w_1, \ldots, w_d) = \int_{\mathbb{R}^2} \max_{j=1,\ldots,d} w_j \zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^*, \qquad \zeta(\boldsymbol{s}_j, \boldsymbol{s}) = \frac{k(\boldsymbol{s}_j, \boldsymbol{s})}{\int_{\mathbb{R}^2} k(\boldsymbol{s}_j, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^*}.$$

The result of Proposition 1 implies that max-stable processes resulting from Cauchy convolution processes are from the class of moving maximum processes (see, e.g., de Haan [27], Schlather [60], Strokorb et al. [64] and the references therein), whose stable tail dependence function is of form given above. Such extreme-value processes, arising as limits of properly renormalized pointwise block maxima of Cauchy convolution processes (2) with block size tending to infinity, admit the stochastic representation

$$Z_{\text{EV}}(\boldsymbol{s}) = \sup_{i=1,2,\ldots} \xi_i \zeta(\boldsymbol{s}, \boldsymbol{s}_i^*), \qquad \zeta(\boldsymbol{s}, \boldsymbol{s}^*) = \frac{k(\boldsymbol{s}, \boldsymbol{s}^*)}{\int_{\mathbb{R}^2} k(\boldsymbol{s}, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^*}, \qquad (5)$$

where $\{(\xi_i, \boldsymbol{s}_i^*)\}$ are points from a Poisson process on $(0, \infty) \times \mathbb{R}^2$ with intensity $\xi^{-2} \mathrm{d}\xi \times \mathrm{d}\boldsymbol{s}^*$. The process in (5) is max-stable and has unit Fréchet margins, i.e., $\Pr\{Z_{\text{EV}}(\boldsymbol{s}) \leq z\} = \exp(-1/z)$, $z > 0$. A prominent example is the model introduced by Smith [63], where the kernel has the shape of a Gaussian density, but the class is much wider than this specific example. We also note that Fasen [19] obtained a result similar to Proposition 1 for continuous-time mixed moving average processes. Similarly, Rootzén [57] studied the tail properties of stable moving average processes and established continuity of sample paths of these processes.

The Smith model [63] is known to have very smooth sample paths; see, e.g., Davison et al. [14]. Although this is already quite clear from the stochastic representation (5) and from spatial realizations, we now show more formally that the extreme-value limits of Cauchy processes of the form (2) are indeed "smooth" in a certain mathematical sense, and therefore that these asymptotic models may be too rigid for modeling block maxima with rough spatial dependence. "Smoothness" of realizations is determined by the form of dependence at short distances, and the next proposition precisely details the behavior of the stable tail dependence function for two variables from the limiting extreme-value process that are located close to each other.

Proposition 2 (Behavior of the stable tail dependence function at short distances)**.** Suppose that the assumptions of Proposition 1 hold. Moreover, assume that the kernel function in (2) may be written as $k(\boldsymbol{s}, \boldsymbol{s}^*) = g(\|\boldsymbol{s} - \boldsymbol{s}^*\|)$, where g is an integrable nonnegative monotonically decreasing function. Then, for any sites $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^2$, the stable tail dependence function $\ell(w_1, w_2)$ of $\{Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)\}^\top$ satisfies $\ell(w_1, w_2) \leq \{1 + K\delta + o(\delta)\} \max(w_1, w_2)$, as $\delta := \|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \to 0$, where K is some constant that does not depend on w_1 and w_2. Furthermore, we can select K such that $\ell(1, 1) = 1 + K\delta + o(\delta)$.

Let $Z_{\mathrm{EV}}(\boldsymbol{s})$ be the extreme-value limit (5) of the Cauchy convolution process $Z(\boldsymbol{s})$ defined as in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{\mathrm{i.i.d.}} \mathrm{Cauchy}(\mathrm{d}\boldsymbol{s}^*)$, which is characterized by the stable tail dependence function ℓ given in Proposition 1. To summarize the dependence structure in a spatial process, it is common to consider scale-free measures of association such as Spearman's rank correlation coefficient, $S_\rho(\delta)$, or the upper tail dependence coefficient defined in (1), $\lambda_U(\delta)$, here expressed as a function of the spatial distance δ between two sites. As Cauchy processes are reflection-symmetric, we have $\lambda_U(\delta) = \lambda_L(\delta)$, so we shall simply write $\lambda(\delta)$ to denote the coefficient of (lower or upper) tail dependence. Recall that while $\lambda(\delta)$ is informative about the strength of tail dependence (and is thus identical for $Z(\boldsymbol{s})$ and $Z_{\mathrm{EV}}(\boldsymbol{s})$), $S_\rho(\delta)$ mostly controls dependence in the bulk (and generally differs for $Z(\boldsymbol{s})$ and $Z_{\mathrm{EV}}(\boldsymbol{s})$). If $(Y_1, Y_2)^\top$ is a random vector distributed according to a joint distribution with continuous margins F_1, F_2 and underlying copula C, then Spearman's rank correlation is defined as $\mathrm{corr}\{F_1(Y_1), F_2(Y_2)\}$ and may be equivalently expressed in terms of the copula C as $12 \iint_{[0,1]^2} C(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2 - 3$. The next corollary exploits Proposition 2 to show that the coefficients $\lambda(\delta)$ and $S_\rho(\delta)$ corresponding to the extreme-value copula C_{EV} stemming from the limiting extreme-value process $Z_{\mathrm{EV}}(\boldsymbol{s})$ have indeed a quite restrictive behavior at the origin, i.e., for small distances $\delta \approx 0$.

Corollary 1 (Behavior of the tail dependence and Spearman's correlation coefficients at short distances)**.** Under the assumptions of Proposition 2, the limiting extreme-value process $Z_{\mathrm{EV}}(\boldsymbol{s})$ of the Cauchy convolution process $Z(\boldsymbol{s})$ (defined as in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{\mathrm{i.i.d.}} \mathrm{Cauchy}(\mathrm{d}\boldsymbol{s}^*)$) has dependence coefficients satisfying $\lambda(\delta) = 1 - K\delta + o(\delta)$ and $S_\rho(\delta) > 1 - 4K\delta$, as $\delta \to 0$.

Corollary 1 implies that moving maximum extreme-value processes resulting from Cauchy convolution processes are not suitable for modeling spatial extremes data such that $S_\rho(\delta) = 1 + O(\delta^\alpha)$, or $\lambda(\delta) = 1 + O(\delta^\alpha)$, with $\alpha < 1$. In Section 2.2, we describe some specific examples to illustrate this property, and in Section 2.3, we show that Cauchy convolution processes capture the subasymptotic dependence structure more flexibly than their extreme-value limits, and we also introduce spatial mixtures that can have different bulk and tail behaviors.

As already seen, the shape of the kernel k in (2) is crucial as it determines the extremal dependence structure of Cauchy convolution processes. In the next corollary, we further show that the support of the bivariate extreme-value copula C_{EV} may not be the whole unit square $[0, 1]^2$ depending on the kernel. This

result may be used to guide the selection of a suitable kernel at a preliminary modeling stage, in order to avoid unreasonable joint behaviors.

Corollary 2 (The support of C_{EV}). Let $\delta = ||\boldsymbol{s}_1 - \boldsymbol{s}_2||$ as before. Assume that the assumptions of Proposition 2 hold, and that

$$G(\delta) = \max_{(s_1^*, s_2^*)^\top \in \mathcal{S}_{\cup+}(\delta)} g(\|(s_1^* + \delta, s_2^*)^\top\|)/g(\|(s_1^*, s_2^*)^\top\|) < \infty,$$

where $\mathcal{S}_{\cup+}(\delta) := \{(s_1^*, s_2^*) \in \mathbb{R}^2 : g(\|(s_1^*, s_2^*)^\top\|) > 0 \text{ and/or } g(\|(s_1^* + \delta, s_2^*)^\top\|) > 0\}$, with the convention that $x/0 = \infty$ for all $x > 0$. Then, $\ell(w_1, w_2) = w_1$ for all $(w_1, w_2)^\top \in [0, \infty)^2$ with $w_1/w_2 > G(\delta) \geq 1$, and similarly, $\ell(w_1, w_2) = w_2$ for all $(w_1, w_2)^\top \in [0, \infty)^2$ with $w_2/w_1 > G(\delta) \geq 1$, which implies that the extreme-value copula $C_{\mathrm{EV}}(u_1, u_2)$ has density zero on the region defined by $\{(u_1, u_2)^\top \in [0, 1]^2 : u_1 < u_2^{G(\delta)} \text{ or } u_2 < u_1^{G(\delta)}\}$.

In particular, for $g(t) = \exp(-t^\alpha)$, $t \geq 0$, with $0 < \alpha \leq 1$, we obtain by the triangle inequality that

$$G(\delta) \leq \max_{(s_1^*, s_2^*) \in \mathbb{R}^2} \exp\{\|(s_1^*, s_2^*)^\top\|^\alpha - \|(s_1^* + \delta, s_2^*)^\top\|^\alpha\}$$

$$\leq \exp\{\|(s_1^*, s_2^*)^\top\|^\alpha - \|(s_1^*, s_2^*)^\top\|^\alpha + \delta^\alpha\} = \exp(\delta^\alpha),$$

which is finite for all $\delta \geq 0$. Thus, Corollary 2 implies that for all kernels of the form $k(\boldsymbol{s}, \boldsymbol{s}^*) = \exp\{-(\|\boldsymbol{s} - \boldsymbol{s}^*\|/\lambda)^\alpha\}$, with $\lambda > 0$ and $0 < \alpha \leq 1$, the extreme-value copula C_{EV} resulting from Cauchy convolution processes does not have full support, thus preventing "very low extreme values" at one location to occur with "very high extreme values" at another location. By contrast, it is easy to verify that $G(\delta) = \infty$ for all $\alpha > 1$, and for all kernels $k(\boldsymbol{s}, \boldsymbol{s}^*) = g(\|\boldsymbol{s} - \boldsymbol{s}^*\|)$ that are compactly supported, i.e., such that $g(t) = 0$ for all $t > r$ for some range $r > 0$. This odd behavior is illustrated in Section 2.4. In practice, it may be sensible to restrict ourselves to $\alpha > 1$ when using $k(\boldsymbol{s}, \boldsymbol{s}^*) = \exp\{-(\|\boldsymbol{s} - \boldsymbol{s}^*\|/r)^\alpha\}$, or to use a different (potentially compactly supported) kernel, to avoid pathological behaviors.

2.2. Special cases

We now give some specific examples that have a tractable bivariate extreme-value copula C_{EV}.

Example 1 (Marshall–Olkin copula). *Consider the indicator kernel $k(\boldsymbol{s}, \boldsymbol{s}^*) = \mathbf{1}\{\boldsymbol{s}^* \in \mathcal{A}(\boldsymbol{s})\}$, where $\mathcal{A}(\boldsymbol{s}) \subset \mathbb{R}^q$ is a compact subset of $\mathbb{R}^q$ with area $0 < |\mathcal{A}(\boldsymbol{s})| < \infty$ for each $\boldsymbol{s} \in \mathbb{R}^q$. From Proposition 1, it is easy to check that the stable tail dependence function of $\{Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)\}^\top$ is*

$$\ell(w_1, w_2) = \frac{|\mathcal{A}_1 \backslash \mathcal{A}_2|}{|\mathcal{A}_1|} \cdot w_1 + \frac{|\mathcal{A}_2 \backslash \mathcal{A}_1|}{|\mathcal{A}_2|} \cdot w_2 + |\mathcal{A}_1 \cap \mathcal{A}_2| \cdot \max\left\{\frac{w_1}{|\mathcal{A}_1|}, \frac{w_2}{|\mathcal{A}_2|}\right\},$$

where we have written $\mathcal{A}_i = \mathcal{A}(\boldsymbol{s}_i)$, $i = 1, 2$, for simplicity. The corresponding limiting extreme-value process is thus driven by the Marshall–Olkin copula with a singular component [50].

It is also possible to obtain the extreme-value limit of the Cauchy convolution process in closed form for non-trivial kernels in some special cases. In particular, one can consider the stationary kernel $k(\boldsymbol{s}, \boldsymbol{s}^*) = g(\|\boldsymbol{s} - \boldsymbol{s}^*\|)$ where g is a nonnegative continuous function.

Example 2 (Smith model [63]). *Assume that $q = 1$ and that $k(s, s^*) = \phi(s - s^*; \sigma^2)$, where $\phi(\cdot; \sigma^2)$ is the Gaussian density function with mean zero and variance σ^2. For simplicity, let also assume that $s_1 < s_2$. Adopting the notation of Proposition 2, it follows that $\{s^* \in \mathbb{R} : w_1\zeta(s_1, s^*) > w_2\zeta(s_2, s^*)\} = \{s^* \in \mathbb{R} : s^* < \frac{\sigma^2}{s_2 - s_1} \log(w_1/w_2) + \frac{s_1 + s_2}{2}\}$, and thus*

$$\ell(w_1, w_2) = w_1 \Phi\left(\frac{\lambda_1^*}{2} + \frac{1}{\lambda_1^*} \log \frac{w_1}{w_2}\right) + w_2 \Phi\left(\frac{\lambda_1^*}{2} + \frac{1}{\lambda_1^*} \log \frac{w_2}{w_1}\right), \quad \lambda_1^* = \frac{|s_1 - s_2|}{\sigma},$$

where $\Phi(\cdot)$ denotes the standard normal CDF. The limiting extreme-value copula is thus the Hüsler–Reiss copula [39]. Here, we can easily verify that $\lambda(\delta) = 2 - \ell(1, 1) = 1 + O(\delta)$, as $\delta = |s_1 - s_2| \to 0$. Note that $\lambda_1^* = \{\gamma(s_1 - s_2)\}^{1/2}$, where $\gamma(h) = (|h|/\sigma)^2$ is a valid variogram. This model corresponds to the Smith max-stable model [63], which is a smooth limiting case of the Brown–Resnick model [41] with variogram $\gamma(h) = (|h|/\sigma)^\alpha$, $\sigma > 0, \alpha \in (0, 2)$. More details for the case of $q = 2$ are provided in Appendix B.

Example 3 (Laplace kernel in $\mathbb{R}$). *Assume that $q = 1$ and let $k(s, s^*) = \frac{1}{2\lambda} \exp(-|s - s^*|/\lambda)$. For simplicity assume $s_1 < s_2$. It follows that for $G_\lambda(\delta) = \exp(\delta/\lambda)$ with $\delta = |s_1 - s_2|$ and $M(s_1, s_2) = \{s^* \in \mathbb{R} : w_1\zeta(s_1, s^*) \geq w_2\zeta(s_2, s^*)\}$,*

$$M(s_1, s_2) = \begin{cases} \left\{s^* \in \mathbb{R} : s^* \leq \frac{\lambda}{2} \log(w_1/w_2) + \frac{s_1 + s_2}{2}\right\}, & \frac{1}{G_\lambda(\delta)} \leq \frac{w_1}{w_2} \leq G_\lambda(\delta), \\ \mathbb{R}, & \frac{w_1}{w_2} > G_\lambda(\delta), \\ \emptyset, & \frac{w_1}{w_2} < \frac{1}{G_\lambda(\delta)}, \end{cases}$$

and, for any sites $s_1, s_2 \in \mathbb{R}$,

$$\ell(w_1, w_2) = \begin{cases} w_1 + w_2 - \sqrt{w_1 w_2 / G_\lambda(\delta)}, & \frac{1}{G_\lambda(\delta)} \leq \frac{w_1}{w_2} \leq G_\lambda(\delta), \\ w_1, & \frac{w_1}{w_2} > G_\lambda(\delta), \\ w_2, & \frac{w_1}{w_2} < \frac{1}{G_\lambda(\delta)}. \end{cases}$$

Thus, the corresponding extreme-value copula density is positive only when $u_2^{G_\lambda(\delta)} < u_1 < u_2^{1/G_\lambda(\delta)}$. For this model, we can again easily verify that $\lambda(\delta) = 1 + O(\delta)$ as $\delta = |s_1 - s_2| \to 0$.

Example 4 (Kernels with compact support). *Notice that if the set $\mathcal{S}_{\cap+}(\boldsymbol{s}_1, \boldsymbol{s}_2) := \{\boldsymbol{s}^* \in \mathbb{R}^q : k(\boldsymbol{s}_1, \boldsymbol{s}^*) > 0, k(\boldsymbol{s}_2, \boldsymbol{s}^*) > 0\} = \emptyset$ for some sites $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^q$, then*

$\ell(w_1, w_2) = w_1 + w_2$. *In particular, if the kernel k is compactly supported such that $k(\boldsymbol{s}, \boldsymbol{s}^*) = 0$ whenever $\|\boldsymbol{s} - \boldsymbol{s}^*\| > r$ for some radius $r > 0$, then $\mathcal{S}_{\cap+}(\boldsymbol{s}_1, \boldsymbol{s}_2)$ is empty if and only if $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \geq 2r$. This implies that the variables $Z_{\mathrm{EV}}(\boldsymbol{s}_1)$ and $Z_{\mathrm{EV}}(\boldsymbol{s}_2)$ from the limiting extreme-value process are independent whenever $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \geq 2r$. By construction, two realizations $Z(\boldsymbol{s}_1)$ and $Z(\boldsymbol{s}_2)$ with $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \geq 2r$ from the process convolution in* (2) *are not only tail-independent but fully independent in this case. A flexible family of compactly supported kernels includes $k(\boldsymbol{s}, \boldsymbol{s}^*) = \{1 - (\|\boldsymbol{s} - \boldsymbol{s}^*\|/r)^\alpha\}_+^\eta$, where $a_+ = \max(0, a)$, for some parameters $r, \alpha, \eta > 0$, though for identifiability concerns one may fix either α and/or η in practice. Here, the parameter r defines the range of spatial dependence for this process and should not be fixed. Another example with compactly supported kernel is detailed in Appendix* B.

2.3. *Mixture of Cauchy and Gaussian processes*

The Cauchy convolution process $Z(\boldsymbol{s})$ defined as in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{\mathrm{i.i.d.}}$ Cauchy$(\mathrm{d}\boldsymbol{s}^*)$ has the appealing property of being tail-dependent (unless exactly independent) and the strength of dependence as a function of distance between two spatial locations is controlled by the kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*)$. In particular, if the kernel k has a compact support, the process is only dependent locally (i.e., at small distances), and is independent at large enough distances. Moreover, importantly, given that the Cauchy convolution process $Z(\boldsymbol{s})$ can be seen as a sum-mixture of heavy-tailed noise, rather than a max-mixture like its extreme-value limit $Z_{\mathrm{EV}}(\boldsymbol{s})$ (recall Proposition 1 and (5)) or the max-mixture models proposed by Wadsworth and Tawn [70], it can generate more flexible patterns and realistic realizations, as demonstrated below. However, Cauchy convolution processes also have certain drawbacks when modeling spatial data, namely:

1. Like Gaussian processes, the resulting copula is reflection-symmetric (and in particular, tail-symmetric), which might not be realistic in some applications;
2. Depending on the kernel k and the distance between sites $\boldsymbol{s}_1, \boldsymbol{s}_2$, two variables $Z(\boldsymbol{s}_1)$ and $Z(\boldsymbol{s}_2)$ can either be tail-dependent, or exactly independent, but the intermediate case of (non-trivial) tail independence is not possible. In other words, the process cannot capture tail independence (unless exactly independent) and thus, it still lacks flexibility at sub-asymptotic levels;
3. The strength of dependence in the bulk of the joint distribution of $\{Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_d)\}^\top$ and in its tails cannot be controlled separately with this process.

While the issue highlighted in the first point is application-specific and should be addressed in future research, we have not found it to be a major limitation in our temperature data application described in Section 4.2. The second and third points highlight, however, issues that are more critical from a risk assessment perspective where the tail dependence structure needs to be estimated
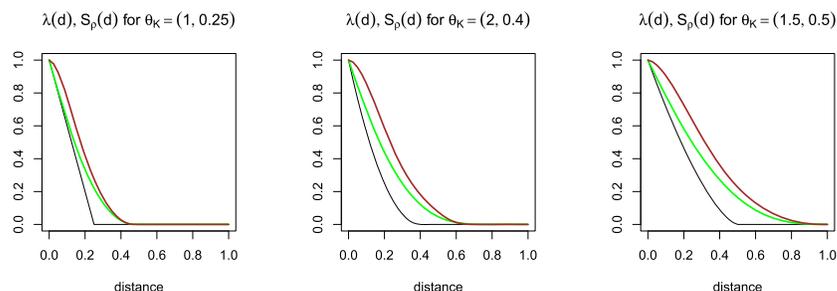
FIG 1. *Coefficient of upper tail dependence, $\lambda(\delta)$ (green line), and Spearman's rank correlation coefficient, $S_\rho(\delta)$ (brown line), for the Cauchy convolution process $Z(s)$ defined as in (2) with kernel function $k(s, s^*; \theta_K) = (1 - \|s - s^*\|/r)_+^\eta$ (black line) and kernel parameters $\theta_K = (\eta, r)^\top = (1, 0.25)^\top$ (left), $\theta_K = (2, 0.4)^\top$ (middle) and $\theta_K = (1.5, 0.5)^\top$ (right).*

with accuracy. In applications, it is common to observe very weak or zero tail dependence at large distances, while fairly strong overall dependence prevails in the bulk. In other words, it can happen in practice that the data suggest both $\lambda(\delta) \approx 0$ and $S_\rho(\delta) \gg 0$ for reasonably large distances δ, but the Cauchy convolution process (2) cannot capture this situation, as the strength and range of dependence in the bulk and the tails are necessarily similar to each other, and cannot be controlled separately, as illustrated in Figure 1.

To increase flexibility of the Cauchy process convolution model and circumvent the issues highlighted in the second and third points above, we propose to modify the original process by mixing it with a tail-independent process possessing lighter tails. Specifically, we define

$$\tilde{Z}(s) = Z(s) + \beta Z_G(s), \tag{6}$$

where $Z(s)$ is the Cauchy process (2) with $W(\mathrm{d}s^*) \sim_{\mathrm{i.i.d.}} \mathrm{Cauchy}(\mathrm{d}s^*)$, $Z_G(s)$ is a stationary Gaussian process with standard normal $N(0, 1)$ marginals and some correlation function $\rho_G(\delta)$, $\delta \geq 0$, and $\beta \geq 0$. For simplicity, we assume that $\rho_G(\delta)$ is an isotropic correlation function, though all the results in this section can be readily extended to the anisotropic or non-stationary context. As the next proposition shows, the new process possesses the same asymptotic behavior as the process (2), such that the spatial mixture model (6) still captures local tail dependence if the kernel is compactly supported.

Proposition 3 (Tail behavior of the spatial mixture process $\tilde{Z}(s)$)**.** Under the assumptions of Proposition 1, the stable tail dependence function corresponding to the joint distribution of $\{\tilde{Z}(s_1), \ldots, \tilde{Z}(s_d)\}^\top$, with the process $\tilde{Z}(s)$ defined as in (6), is $\ell(w_1, \ldots, w_d)$ given in Proposition 1.

While non-trivial tail independence can be captured by the spatial mixture process, it is important to remark that the tail dependence structure still remains fairly restricted in this case. To see this, assume that the random variables $Z(s_1), \ldots, Z(s_d)$ from the Cauchy process appearing in (6) are independent.

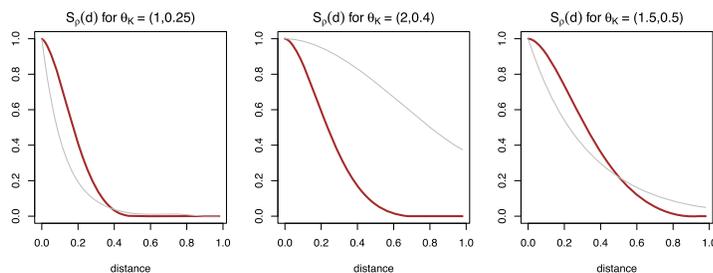FIG 2. *Spearman's correlation coefficient, $S_\rho(\delta)$, for the Cauchy convolution process $Z(\boldsymbol{s})$ defined in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{i.i.d.}$ Cauchy($\mathrm{d}\boldsymbol{s}^*$) (brown), and for the process $\tilde{Z}(\boldsymbol{s})$ defined in (6) (grey), based on the kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) = (1 - \|\boldsymbol{s} - \boldsymbol{s}^*\|/r)_+^\eta$ with $\boldsymbol{\theta}_K = (\eta, r)^\top = (1, 0.25)^\top$ and $\rho_G(\delta) = \exp(-8d)$ (left), $\boldsymbol{\theta}_K = (2, 0.4)^\top$ and $\rho_G(\delta) = \exp(-\delta^2)$ (middle) and $\boldsymbol{\theta}_K = (1.5, 0.5)^\top$ and $\rho_G(\delta) = \exp(-3\delta)$ (right). For the model $\tilde{Z}(\boldsymbol{s})$, we set $\beta = 2$.*

Then, from the proof of Proposition 3, we get that

$$p_{L,n} + o(e^{-n/(2\beta^2)}) \leq \Pr(\tilde{Z}_1 \leq n, \ldots, \tilde{Z}_d \leq n) \leq p_{U,n} + o(e^{-n/(2\beta^2)}),$$

where $p_{L,n} = \Pr(Z_1 \leq n - \sqrt{n}, \ldots, Z_d \leq n - \sqrt{n}) = \{\Pr(Z_1 \leq n)\}^d + o(\{\Pr(Z_1 \leq n)\}^d) = \{\Pr(\tilde{Z}_1 \leq n)\}^d + o(\{\Pr(\tilde{Z}_1 \leq n)\}^d)$, and $p_{U,n} = \Pr(Z_1 \leq n + \sqrt{n}, \ldots, Z_d \leq n + \sqrt{n}) = \{\Pr(Z_1 \leq n)\}^d + o(\{\Pr(Z_1 \leq n)\}^d) = \{\Pr(\tilde{Z}_1 \leq n)\}^d + o(\{\Pr(\tilde{Z}_1 \leq n)\}^d)$, and this implies that the tail order of the copula linking $\tilde{Z}(\boldsymbol{s}_1), \ldots, \tilde{Z}(\boldsymbol{s}_d)$ is equal to $\kappa = d$. In other words, this copula has tail quadrant independence, which is a weak form of tail independence. Nevertheless, the spatial mixture process (6) enjoys great flexibility at sub-asymptotic levels, as demonstrated below.

To illustrate the improved flexibility of model (6), Figure 2 shows $S_\rho(\delta)$ for the process $\tilde{Z}(\boldsymbol{s})$ in (6) with various kernel functions, $\beta = 2$, and underlying correlation function for the Gaussian process equal to $\rho_G(\delta) = \exp(-8\delta), \exp(-\delta^2)$, or $\exp(-3\delta)$, respectively. We can see that $\rho_G(\delta)$ mainly controls the Spearman's correlation of the process $\tilde{Z}(\boldsymbol{s})$, whereas the kernel parameters $\boldsymbol{\theta}_K$ control its tail dependence structure, as expected. The new spatial mixture process therefore allows for a greater flexibility both in the tails and in the middle of the joint distribution, capturing a wide range of sub-asymptotic dependence behaviors.

2.4. Simulation and realizations

Fast approximate simulation from the Cauchy convolution process $Z(\boldsymbol{s})$ defined in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{\text{i.i.d.}}$ Cauchy($\mathrm{d}\boldsymbol{s}^*$) can be easily obtained at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d \in \mathbb{R}^2$ as $Z(\boldsymbol{s}_j) \approx \delta_m^2 \sum_{k,l=1}^m k(\boldsymbol{s}_j, \boldsymbol{s}_{k,l}) W_{k,l}$, $W_{k,l} \sim_{\text{i.i.d}}$ Cauchy(1), $j = 1, \ldots, d$, where $\{\boldsymbol{s}_{k,l}\}_{k,l=1}^m$ is a fine rectangular grid in $S_m \subset \mathbb{R}^2$, with the cell size $\delta_m \times \delta_m$, covering the simulation sites, with large m (see the proof of Proposition 1). Similarly, the process $\tilde{Z}(\boldsymbol{s})$ in (6) can be simulated based on a finite approximation as $\tilde{Z}(\boldsymbol{s}_j) \approx \delta_m^2 \sum_{k,l=1}^m k(\boldsymbol{s}_j, \boldsymbol{s}_{k,l}) W_{k,l} + \beta Z_G(\boldsymbol{s}_j)$, $W_{k,l} \sim_{\text{i.i.d}}$

Cauchy(1), $j = 1, \ldots, d$, where $\{Z_G(\boldsymbol{s}_1), \ldots, Z_G(\boldsymbol{s}_d)\}^\top$ is a realization from the Gaussian random field $Z_G(\boldsymbol{s})$ at the sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$, independent of the Cauchy variables $\{W_{k,l}\}$.

Simulation from the associated limiting extreme-value process $Z_{\mathrm{EV}}(\boldsymbol{s})$ is also quite straightforward. Given that the kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*)$ is bounded, simulating moving maximum max-stable random fields can be performed exactly by exploiting the stochastic representation (5); see Schlather [60]. If the kernel is compactly supported with range $r > 0$, then the study region should be expanded on all sides by r units, and if it has support over $\mathbb{R}^2$, then the study region should be expanded sufficiently to ensure that the contributions of "distant" points in (5) become negligible. Alternatively, fast approximate simulations are also possible; see Appendix C for more details.

Figure 3 shows realizations of the processes $Z(\boldsymbol{s})$, $\tilde{Z}(\boldsymbol{s})$ and $Z_{\mathrm{EV}}(\boldsymbol{s})$ in $[0, 1]^2$ obtained for different kernel functions. For the indicator kernel with compact support we used a rectangular grid on $[-r, 1 + r]^2$ to prevent edge effects. For kernels with infinite support, we used a regular grid on $S_m = [-1, 2]^2$ because $k(\boldsymbol{s}, \boldsymbol{s}^*) < 5 \cdot 10^{-5}$ for $\|\boldsymbol{s} - \boldsymbol{s}^*\| > 1$ for the considered kernels. We use $m = 200$ in all cases. We can see that the realizations from the Cauchy process $Z(\boldsymbol{s})$ (second column) can indeed generate quite complex patterns, as opposed to its fairly rigid extreme-value limit $Z_{\mathrm{EV}}(\boldsymbol{s})$ (fourth column). Furthermore, the spatial mixture process $\tilde{Z}(\boldsymbol{s})$ (third column) has much rougher spatial realizations than the Cauchy process $Z(\boldsymbol{s})$, confirming the higher degree of flexibility of $\tilde{Z}(\boldsymbol{s})$ to capture the sub-asymptotic dependence structure. Moreover, in all simulations, the effect of the kernel on the spatial dependence structure is obvious, especially with the extreme-value process $Z_{\mathrm{EV}}(\boldsymbol{s})$ (fourth column). Clearly, smoother kernels result in smoother random fields.

Figure 4 shows bivariate scatter plots for datasets of size 1000 generated from the extreme-value copulas linking $Z_{\mathrm{EV}}(\boldsymbol{s}_1)$ and $Z_{\mathrm{EV}}(\boldsymbol{s}_2)$ at distance $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| = 0.125$ for the four extremal processes displayed in Figure 3. With the indicator kernel, the singular component on the diagonal $u_1 = u_2$ is clearly seen. Green lines $u_1 = u_2^{\pm \exp(20\delta)}$ and $u_1 = u_2^{\pm \exp(10\delta^{0.5})}$, with $\delta = 0.125$, delimit the area with positive density for the data generated using the exponential and powered exponential kernels; recall Corollary 2. Note that for the powered exponential kernel, the copula density is positive but very small near the boundaries. By contrast, the Gaussian kernel yields a positive density on the whole unit square $[0, 1]^2$, and the indicator kernel has a positive density on $[0, 1]^2 \setminus \{(u_1, u_2)^\top : u_1 = u_2\}$.

In the next section, we detail our proposed inference approach for the Cauchy convolution process $Z(\boldsymbol{s})$ observed at d locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$, and for the modified spatial mixture process $\tilde{Z}(\boldsymbol{s})$.
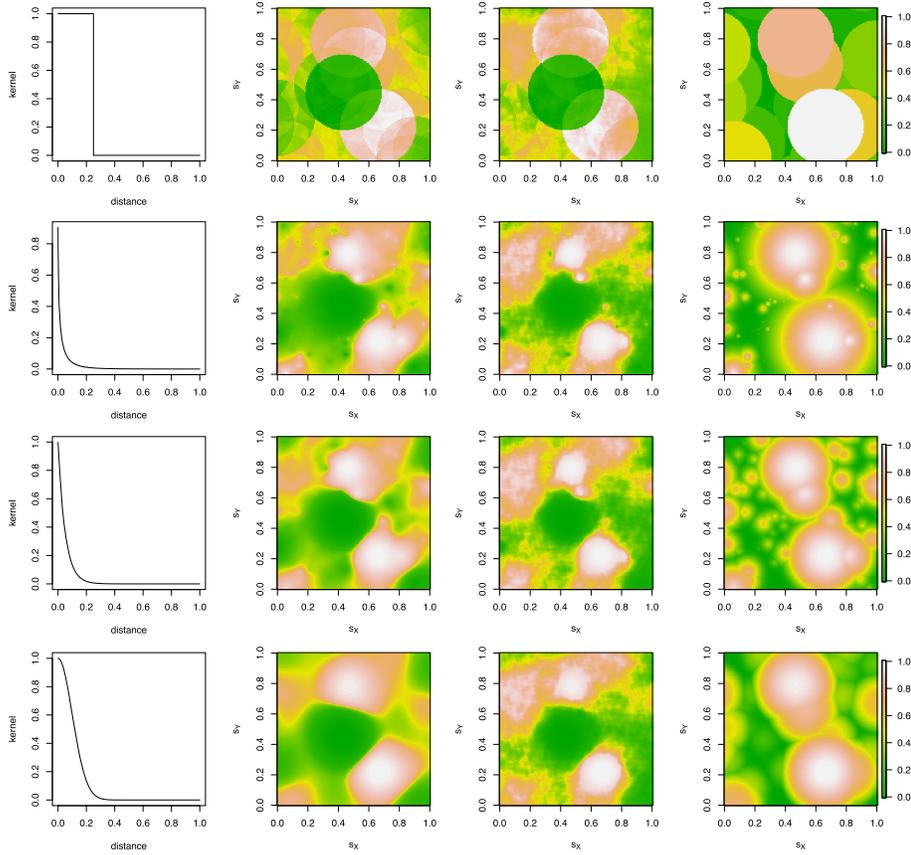
FIG 3. *Realizations on $[0,1]^2$ from the Cauchy convolution process $Z(\boldsymbol{s})$ defined in* (2) *with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{i.i.d.}$ Cauchy$(\mathrm{d}\boldsymbol{s}^*)$ (2nd column), $\tilde{Z}(\boldsymbol{s})$ defined in* (6) *(3rd column), and their extreme-value limit $Z_{\mathrm{EV}}(\boldsymbol{s})$ (4th column) for different isotropic kernels $k(\boldsymbol{s}, \boldsymbol{s}^*) = g(\|\boldsymbol{s} - \boldsymbol{s}^*\|)$ (the 1st column shows the function g). Each simulated process is marginally transformed to have $\mathrm{Unif}(0,1)$ marginals. The same random seed was used for all realizations. The kernel is $k(\boldsymbol{s}, \boldsymbol{s}^*) = \mathbf{1}(\|\boldsymbol{s} - \boldsymbol{s}^*\| < 0.25)$ (1st row), $k(\boldsymbol{s}, \boldsymbol{s}^*) = \exp\left(-10\|\boldsymbol{s} - \boldsymbol{s}^*\|^{1/2}\right)$ (2nd row), $k(\boldsymbol{s}, \boldsymbol{s}^*) = \exp\left(-20\|\boldsymbol{s} - \boldsymbol{s}^*\|\right)$ (3rd row), and $k(\boldsymbol{s}, \boldsymbol{s}^*) = \exp\left(-50\|\boldsymbol{s} - \boldsymbol{s}^*\|^2\right)$ (4th row). We use $\beta = 2$ and $\rho_G(\delta) = \exp(-\delta)$ for $\tilde{Z}(\boldsymbol{s})$.*

3. Inference

3.1. General setting and marginal estimation

Throughout this section, we assume that $\{\mathbf{y}_i = (y_{i1}, \ldots, y_{id})^\top\}_{i=1}^n$ are n i.i.d. realizations of a stationary process $Y(\boldsymbol{s})$ measured at d spatial locations $\boldsymbol{s}_1, \ldots,$ $\boldsymbol{s}_d \in \mathbb{R}^2$, and whose dependence structure (i.e., copula) is the same as either (i) the Cauchy convolution process $Z(\boldsymbol{s})$ defined as in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{\mathrm{i.i.d.}}$ Cauchy$(\mathrm{d}\boldsymbol{s}^*)$ using some parametric kernel function k (Section 3.2); or (ii) the
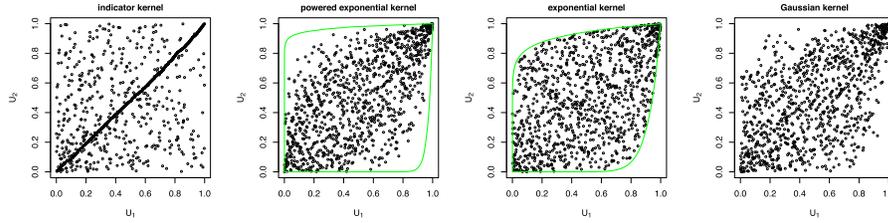
FIG 4. *Bivariate scatter plots of datasets of size* 1000 *generated from the extreme-value copulas linking* $Z_{\mathrm{EV}}(s_1)$ *and* $Z_{\mathrm{EV}}(s_2)$ *at distance* $\delta = \|s_1 - s_2\| = 0.125$ *based on the kernel* $k(s, s^*) = \mathbf{1}(\|s - s^*\| < 0.25)$, $k(s, s^*) = \exp\left(-10\|s - s^*\|^{1/2}\right)$, $k(s, s^*) = \exp\left(-20\|s - s^*\|\right)$, *and* $k(s, s^*) = \exp\left(-50\|s - s^*\|^2\right)$ *(left to right). Green lines* $u_1 = u_2^{\pm \exp(10\delta^{1/2})}$ *and* $u_1 = u_2^{\pm \exp(20\delta)}$ *delimit the area with positive density in the* 2nd *and* 3rd *panels, respectively.*

spatial mixture model extension $\tilde{Z}(s)$ defined as in (6) (Section 3.3); or (iii) the extreme-value limit process $Z_{\mathrm{EV}}(s)$ defined as in (5) (Section 3.4), but with potentially different marginal distributions. We also assume that $F(\cdot; \boldsymbol{\theta}_F)$ and $f(\cdot; \boldsymbol{\theta}_F)$ are the marginal distribution and density functions, respectively, of the observed process $Y(s)$ and that $\boldsymbol{\theta}_F$ is the vector of marginal parameters. In other words, the process $Y(s)$ has the same dependence properties (both in the bulk and the tails) as $Z(s)$, $\tilde{Z}(s)$ or $Z_{\mathrm{EV}}(s)$, but the marginal distributions are essentially arbitrary, thus allowing for a greater flexibility by modeling margins and dependence separately.

We may estimate marginal parameters non-parametrically in a first step using the empirical distribution (computed for each site separately, or for the entire pooled dataset). Alternatively, we can also estimate the vector of marginal parameters $\boldsymbol{\theta}_F$ by maximizing the marginal (composite) log-likelihood function $\ell_F(\mathbf{y}; \boldsymbol{\theta}_F) = \sum_{i=1}^{n} \sum_{j=1}^{d} \log f(y_{ij}; \boldsymbol{\theta}_F)$, and we denote the respective estimate by $\widehat{\boldsymbol{\theta}}_F$. Such an approach, which neglects spatial dependence to estimate marginal parameters, is known to be valid (i.e., yielding consistency and asymptotic normality of estimators) under mild regularity conditions [67]. The data can then be transformed to the uniform $\mathrm{Unif}(0,1)$ scale using the probability integral transform. More precisely, pseudo-uniform scores $\{\mathbf{u}_i = (u_{i1}, \ldots, u_{id})^\top\}_{i=1}^{n}$ can be obtained by setting $u_{ij} = F(y_{ij}; \widehat{\boldsymbol{\theta}}_F)$.

In the next sections, we show how dependence parameters may be estimated from the pseudo-uniform scores $\{\mathbf{u}_i\}_{i=1}^{n}$ by matching some suitable empirical and model-based summary statistics (for $Z(s)$ in (2), and $\tilde{Z}(s)$ in (6)), or by exploiting composite likelihood inference (for $Z_{\mathrm{EV}}(s)$ in (5)).

3.2. Parameter estimation for the Cauchy convolution process model defined in (2)

Here, we assume that the pseudo-uniform scores $\{\mathbf{u}_i\}_{i=1}^{n}$ obtained in Section 3.1 are driven by the copula stemming from the Cauchy convolution process $Z(s)$

in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{\mathrm{i.i.d.}} \mathrm{Cauchy}(\mathrm{d}\boldsymbol{s}^*)$ using some parametric kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)$, where $\boldsymbol{\theta}_K$ is the vector of dependence parameters controlling the kernel function. While the joint likelihood function for the process $Z(\boldsymbol{s})$ is not tractable in general, we can exploit its stochastic representation in (2) to estimate $\boldsymbol{\theta}_K$.

We first back-transform the pseudo-uniform scores to the standard Cauchy scale. More precisely, we compute $\{\boldsymbol{z}_i^* = (z_{i1}^*, \ldots, z_{id}^*)^\top\}_{i=1}^n$ with $z_{ij}^* = F_{\mathcal{C}}^{-1}(u_{ij})$, where $F_{\mathcal{C}}^{-1}(q) = \tan\{\pi(q-0.5)\}$ is the quantile function of the standard Cauchy distribution. By stability of the Cauchy distribution, the marginal distribution of the process convolution $Z(\boldsymbol{s})$ in (2) are Cauchy with scale parameter $c(\boldsymbol{s}; \boldsymbol{\theta}_K) = \int_{\mathbb{R}^2} k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)\mathrm{d}\boldsymbol{s}^*$. This implies that $Z^*(\boldsymbol{s}) := Z(\boldsymbol{s})/c(\boldsymbol{s}; \boldsymbol{\theta}_K) \sim \mathrm{Cauchy}(1)$ and that $\{\boldsymbol{z}_i^*\}_{i=1}^n$ can be treated as pseudo-observations from $Z^*(\boldsymbol{s})$. Moreover, we have

$$
\begin{aligned}
Z_{jk}^* = Z^*(\boldsymbol{s}_j) - Z^*(\boldsymbol{s}_k) &= \frac{Z(\boldsymbol{s}_j)}{c(\boldsymbol{s}_j; \boldsymbol{\theta}_K)} - \frac{Z(\boldsymbol{s}_k)}{c(\boldsymbol{s}_k; \boldsymbol{\theta}_K)} \\
&= \int_{\mathbb{R}^2} \{\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) - \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)\} W(\mathrm{d}\boldsymbol{s}^*) \sim \mathrm{Cauchy}\{c_{jk}(\boldsymbol{\theta}_K)\},
\end{aligned}
$$

where

$$
c_{jk}(\boldsymbol{\theta}_K) = \int_{\mathbb{R}^2} |\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) - \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)|\,\mathrm{d}\boldsymbol{s}^*, \tag{7}
$$

$$
\zeta(\boldsymbol{s}, \boldsymbol{s}^*) \equiv \zeta(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) = \frac{k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)}{\int_{\mathbb{R}^2} k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)\mathrm{d}\boldsymbol{s}^*}.
$$

For each pair of sites $\{\boldsymbol{s}_j, \boldsymbol{s}_k\}$, the scale parameter c_{jk} may be estimated non-parametrically from the pseudo-observations $\{z_{ijk}^* = z_{ij}^* - z_{ik}^*\}_{i=1}^n$ by maximizing the corresponding Cauchy likelihood function. The maximum likelihood estimator $\widehat{c}_{jk}$ satisfies the equation $\sum_{i=1}^n \widehat{c}_{jk}^2/(\widehat{c}_{jk}^2 + z_{ijk}^{*\,2}) = n/2$, whose positive root can be easily found using numerical routines. The estimator $\widehat{c}_{jk}$ is a consistent and asymptotically normal estimator of c_{jk}. Alternatively, the median of absolute values, $\mathrm{median}\{|z_{1jk}|, \ldots, |z_{njk}|\}$, may also be used as a more robust and faster-to-compute non-parametric estimator of c_{jk}, but which is about 20% less efficient than the maximum likelihood estimator (in terms of the ratio of their variances); see Zhang [73]. To estimate the vector of parameters $\boldsymbol{\theta}_K$, we can then use a least squares approach and compute

$$
\widehat{\boldsymbol{\theta}}_K = \arg\min_{\boldsymbol{\theta}_K} \sum_{j<k} \omega_{jk}\{c_{jk}(\boldsymbol{\theta}_K) - \widehat{c}_{jk}\}^2, \tag{8}
$$

where $\omega_{kj} \geq 0$ are some non-negative weights. While equal weights $\omega_{jk} = 1$ are commonly chosen, binary weights specified according to the distance between sites, e.g., $\omega_{jk} = \mathbf{1}(\|\boldsymbol{s}_j - \boldsymbol{s}_k\| \leq \delta_{\max})$ for some cut-off distance $\delta_{\max} > 0$, may be helpful to reduce the computational burden and/or prioritize goodness-of-fit at small distances. The estimator in (8) is a special case of minimum distance estimators and therefore it is a consistent and asymptotically normal estimator of $\boldsymbol{\theta}_K$ [51].

3.3. *Parameter estimation for the spatial mixture model extension defined in* (6)

We now assume that the pseudo-uniform scores $\{\mathbf{u}_i\}_{i=1}^n$ obtained in Section 3.1 are driven by the copula stemming from the extended model $\tilde{Z}(\boldsymbol{s})$ in (6). In addition to the parametric kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)$ described by the vector of parameters $\boldsymbol{\theta}_K$, we now need to estimate the correlation function $\rho_G(\delta; \boldsymbol{\theta}_G)$ parametrized by a vector $\boldsymbol{\theta}_G$, and the mixture parameter $\beta \geq 0$.

Parameter estimation is now more tricky, but parameters can nevertheless be estimated in two steps by noticing that the copula of the limiting extreme-value process $Z_{\mathrm{EV}}(\boldsymbol{s})$ only depends on the kernel parameters $\boldsymbol{\theta}_K$. In the first step, $\boldsymbol{\theta}_K$ can thus be estimated by matching empirical and model-based estimates of the tail dependence coefficient for different pairs of sites. More precisely, let $\lambda_{jk}(\boldsymbol{\theta}_K)$ be the tail dependence coefficient defined in (1) corresponding to the pair of variables $\{\tilde{Z}(\boldsymbol{s}_j), \tilde{Z}(\boldsymbol{s}_k)\}$. On the one hand, λ_{jk} can be estimated non-parametrically from the pseudo-uniform scores $\{(u_{ij}, u_{ik})^\top\}_{i=1}^n$ for each pair of sites $\{\boldsymbol{s}_j, \boldsymbol{s}_k\}$, e.g., as

$$\widehat{\lambda}_{jk} = \frac{1}{n(1-u)} \sum_{i=1}^n \mathbf{1}(u_{ij} > u, u_{ik} > u), \tag{9}$$

where $u \approx 1$ is a high threshold on the uniform scale. The empirical estimator $\widehat{\lambda}_{jk}$ is consistent as $n \to \infty$ and $u \equiv u_n \to 1$ such that $n(1-u_n) \to \infty$. Many other valid non-parametric estimators of the tail dependence coefficient exist. In particular, as the copula stemming from $\tilde{Z}(\boldsymbol{s})$ is reflection-symmetric, it would be possible to combine information from the lower and upper tails to estimate $\lambda_{jk}(\boldsymbol{\theta}_K)$. Hereafter, we rely on an improved estimator proposed by Lee et al. [48], which works well for small sample sizes. On the other hand, from Propositions 1 and 3, we have

$$\lambda_{jk}(\boldsymbol{\theta}_K) = 2 - \ell(1,1) = 2 - \int_{\mathbb{R}^2} \max\{\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*), \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)\} \mathrm{d}\boldsymbol{s}^*$$

$$= 1 - \int_{\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) > \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)} \zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^* + 1 - \int_{\zeta(\boldsymbol{s}_k, \boldsymbol{s}^*) > \zeta(\boldsymbol{s}_j, \boldsymbol{s}^*)} \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^*$$

$$= \int_{\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) \leq \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)} \zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^* + \int_{\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) \geq \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)} \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^*, \tag{10}$$

where $\ell(w_1, w_2)$ is the stable tail dependence function of $\{\tilde{Z}(\boldsymbol{s}_j), \tilde{Z}(\boldsymbol{s}_k)\}^\top$, and $\zeta(\boldsymbol{s}, \boldsymbol{s}^*) \equiv \zeta(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) = k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)/\int_{\mathbb{R}^2} k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) \mathrm{d}\boldsymbol{s}^*$. The coefficient $\lambda_{jk}(\boldsymbol{\theta}_K)$ can be expressed in simple form as a function of $\boldsymbol{\theta}_K$ for many parametric families of kernels. We consider one example in the simulation study reported in Section 4.1. The parameter $\boldsymbol{\theta}_K$ can then be estimated by least squares as follows:

$$\widehat{\boldsymbol{\theta}}_K = \arg\min_{\boldsymbol{\theta}_K} \sum_{j<k} \omega_{jk} \{\lambda_{jk}(\boldsymbol{\theta}_K) - \widehat{\lambda}_{jk}\}^2, \tag{11}$$

where $\widehat{\lambda}_{jk}$ are the empirical estimates from (9) and $\omega_{jk} \geq 0$ are some non-negative weights as in (8). Notice that this approach based on the tail dependence coefficient could also be applied to the Cauchy process $Z(\boldsymbol{s})$, as it shares the same extremal dependence structure, but the least squares estimator (8) is more efficient than (11) as it uses information from both the bulk and the tails.

After having estimated the kernel parameters $\boldsymbol{\theta}_K$, we now need to estimate the parameters $\boldsymbol{\theta}_G$ of the correlation function $\rho_G(\delta) = \rho_G(\delta; \boldsymbol{\theta}_G)$ and the parameter $\beta \geq 0$ in (6). Let $F(\cdot; \gamma, \beta)$ be the CDF of the variable $\gamma W + \beta Z$ where W and Z are independent random variables following the standard Cauchy and Gaussian distributions, respectively. The distribution function of this sum of variables may be expressed in integral form as

$$F(w; \gamma, \beta) = \frac{1}{2} + \frac{1}{\pi} \int_{\mathbb{R}} \arctan\{(w - \beta z)/\gamma\}\phi(z)\mathrm{d}z, \tag{12}$$

where $\phi(\cdot)$ is the standard normal density function. Numerical integration can be used to quickly and accurately evaluate (12). The corresponding density function can be computed by differentiating (12) under the integral sign. Fixing the value of $\beta = \beta_0 \geq 0$, we then back-transform the pseudo-uniform scores to the scale of $F(\cdot; 1, \beta_0)$ as $\{\tilde{\boldsymbol{z}}_i = (\tilde{z}_{i1}, \ldots, \tilde{z}_{id})^\top\}_{i=1}^n$, where $\tilde{z}_{ij} = F^{-1}(u_{ij}; 1, \beta_0)$. If β_0 is the true value of β, then the vectors $\{\tilde{\boldsymbol{z}}_i\}_{i=1}^n$ can be considered as pseudo-observations from the process $\tilde{Z}(\boldsymbol{s})$ in (6). Now, notice that for any two sites $\{\boldsymbol{s}_j, \boldsymbol{s}_k\}$, we have

$$\tilde{Z}_{jk}^+ = \tilde{Z}(\boldsymbol{s}_j) + \tilde{Z}(\boldsymbol{s}_k) \sim F\{\cdot; 2, \beta_{jk}^+(\boldsymbol{\theta}_G)\},$$
$$\tilde{Z}_{jk}^- = \tilde{Z}(\boldsymbol{s}_j) - \tilde{Z}(\boldsymbol{s}_k) \sim F\{\cdot; c_{jk}(\boldsymbol{\theta}_K), \beta_{jk}^-(\boldsymbol{\theta}_G)\},$$

where $c_{jk}(\boldsymbol{\theta}_K)$ is given in (7) and $\beta_{jk}^+(\boldsymbol{\theta}_G) = \beta_0\{2 + 2\rho_G(\delta_{jk}; \boldsymbol{\theta}_G)\}^{1/2}$, $\beta_{jk}^-(\boldsymbol{\theta}_G) = \beta_0\{2 - 2\rho_G(\delta_{jk}; \boldsymbol{\theta}_G)\}^{1/2}$, $\delta_{jk} = \|\boldsymbol{s}_j - \boldsymbol{s}_k\|$. To obtain empirical estimates of $\boldsymbol{\theta}_G$, we first obtain non-parametric estimates $\widehat{\beta}_{jk}^+$ of $\beta_{jk}^+(\boldsymbol{\theta}_G)$ by assuming that $\tilde{Z}_{jk}^+ \sim F\{\cdot; 2, \beta_{jk}^+(\boldsymbol{\theta}_G)\}$ and maximizing the corresponding likelihood function using the pseudo-observations $\{\tilde{z}_{jk}^+ = \tilde{z}_{ij} + \tilde{z}_{ik}\}_{i=1}^n$. We use the same maximum likelihood approach to get non-parametric estimates $\widehat{\beta}_{jk}^-$ of $\beta_{jk}^-(\boldsymbol{\theta}_G)$, but now assuming that $\tilde{Z}_{jk}^- \sim F\{\cdot; c_{jk}(\widehat{\boldsymbol{\theta}}_K), \beta_{jk}^-(\boldsymbol{\theta}_G)\}$ with $\widehat{\boldsymbol{\theta}}_K$ obtained in (11) and using the pseudo-observations $\{\tilde{z}_{jk}^- = \tilde{z}_{ij} - \tilde{z}_{ik}\}_{i=1}^n$ instead. The vector of parameters $\boldsymbol{\theta}_G$ and the parameter $\beta \geq 0$ can then be jointly estimated by least squares as

$$(\widehat{\boldsymbol{\theta}_G^\top}, \widehat{\beta})^\top = \arg\min_{(\boldsymbol{\theta}_G^\top, \beta_0)^\top} \sum_{j<k} \omega_{jk}^+ \{\beta_{jk}^+(\boldsymbol{\theta}_G) - \widehat{\beta}_{jk}^+\}^2 + \sum_{j<k} \omega_{jk}^- \{\beta_{jk}^-(\boldsymbol{\theta}_G) - \widehat{\beta}_{jk}^-\}^2, \tag{13}$$

where $\omega_{jk}^+ \geq 0$ and $\omega_{jk}^- \geq 0$ are some non-negative weights associated to each pair of sites $\{\boldsymbol{s}_j, \boldsymbol{s}_k\}$ as in (8) and (11). In practice, the minimization in (13) can be performed over $\boldsymbol{\theta}_G$ for a grid of fixed values $\beta = \beta_0 \geq 0$, and then the value β can be selected in a second step as the one that provides the lowest objective function overall. The estimators in (11) and (13) are again special

cases of minimum distance estimators and they are therefore consistent and asymptotically normal if $\widehat{\lambda}_{j,k}$ and $\widehat{\beta}_{jk}^{+}, \widehat{\beta}_{jk}^{-}$ are consistent and asymptotically normal estimators of $\lambda_{j,k}$ and $\beta_{jk}^{+}(\boldsymbol{\theta}_G), \beta_{jk}^{-}(\boldsymbol{\theta}_G)$, respectively [51].

3.4. Parameter estimation for the extreme-value model $Z_{\mathrm{EV}}(s)$ in (5)

We now assume that the pseudo-uniform scores $\{\mathbf{u}_i\}_{i=1}^n$ obtained in Section 3.1 are driven by the copula stemming from the extreme-value process $Z_{\mathrm{EV}}(\boldsymbol{s})$ in (5) with parametric kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)$, obtained as limit of the Cauchy convolution process (2) or the spatial mixture model extension (6). Such an extreme-value copula has the form (4). When the stable tail dependence function ℓ given in Proposition 1 has a simple explicit form, a pairwise (composite) likelihood approach may be used to estimate the dependence parameters $\boldsymbol{\theta}_K$ [49, 56, 67]. Let $C_{\mathrm{EV},jk}$ be the bivariate extreme-value copula restricted to the pair of sites $\{\boldsymbol{s}_j, \boldsymbol{s}_k\}$, and $c_{\mathrm{EV},jk}$ be the corresponding density function. More precisely, we have $C_{\mathrm{EV},jk}(u_j, u_k) = \exp\{-\ell_{jk}(-\log u_j, -\log u_k; \boldsymbol{\theta}_K)\}$, where $\ell_{jk}(w_j, w_k) = \ell(w_j \mathbf{e}_j + w_k \mathbf{e}_k)$ (with $\mathbf{e}_j$ the j-th canonical basis vector of $\mathbb{R}^d$) is the (j,k)-th margin of $\ell(w_1, \ldots, w_d)$, and

$$
\begin{aligned}
c_{\mathrm{EV},jk}(u_j, u_k) = {}& (u_j u_k)^{-1} \exp\{-\ell_{jk}(-\log u_j, -\log u_k; \boldsymbol{\theta}_K)\} \\
& \times \{\partial_1 \ell_{jk}(-\log u_j, -\log u_k; \boldsymbol{\theta}_K) \partial_2 \ell_{jk}(-\log u_j, -\log u_k; \boldsymbol{\theta}_K) \\
& \quad - \partial_{12} \ell_{jk}(-\log u_j, -\log u_k; \boldsymbol{\theta}_K)\},
\end{aligned}
$$

where ∂_1 means differentiation with respect to the first argument, and so forth. The parameter vector $\boldsymbol{\theta}_K$ may then be estimated by maximizing a pairwise log-likelihood function as follows:

$$
\widehat{\boldsymbol{\theta}}_K = \arg\max_{\boldsymbol{\theta}_K} \sum_{i=1}^n \sum_{j<k} \omega_{jk} \log\{c_{\mathrm{EV},jk}(u_{ij}, u_{ik}; \boldsymbol{\theta}_K)\}, \tag{14}
$$

where $\omega_{jk} \geq 0$ are some non-negative weights. Under mild regularity conditions, the maximum pairwise likelihood estimator (14) is consistent and asymptotically normal, but with some loss of efficiency compared to the usual maximum likelihood estimator since it only uses information contained in pairs of variables; see, e.g., Huser [31], Chapter 3, and Padoan et al. [56]. Notice that when the stable tail dependence function ℓ is not tractable or is difficult to compute, it is also possible to estimate $\boldsymbol{\theta}_K$ by matching empirical and model-based coefficient of tail dependence given by $\lambda_{jk} = 2 - \ell_{jk}(1,1)$. This approach is similar to the least squares estimator (11), except that the coefficients λ_{jk} may be estimated from the full dataset of maxima instead of just from the tail. Various estimators have been proposed to estimate λ_{jk} non-parametrically, and in our simulation study we relied on an estimator of the so-called Pickands dependence function; see for example [22].

4. Numerical experiments

4.1. Simulation study

We now perform a simulation study to verify the performance of the inference schemes proposed in Sections 3.2, 3.3, and 3.4, before illustrating our proposed methodology by application to real temperature data in Section 4.2. For estimating kernel parameters in the mixture process $\tilde{Z}(\boldsymbol{s})$, we rely on the least squares approach (11) and we use a nonparametric estimator of the tail dependence coefficient proposed by Lee et al. [48] that was found to be efficient and accurate in small samples. For the extreme-value process $Z_{\mathrm{EV}}(\boldsymbol{s})$, we similarly estimate kernel parameters by matching empirical and model-based coefficients of tail dependence given now by $\lambda_{jk} = 2 - \ell_{jk}(1,1)$. This approach is computationally tractable in high dimensions even if the stable tail dependence function ℓ has no simple form, and it is akin to the least squares estimator (11) except that the coefficients λ_{jk} can be estimated from the full dataset instead of just from the tail. Various nonparametric estimators have been proposed for λ_{jk}, and we here rely on an estimator of the so-called Pickands dependence function; see Genest and Segers [22].

We simulate datasets comprised of $n = 100, 200, 500$ replicates of the Cauchy convolution process $Z(\boldsymbol{s})$ in (2) for some kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)$, its spatial mixture process extension $\tilde{Z}(\boldsymbol{s})$ in (6), and the extreme-value limit $Z_{\mathrm{EV}}(\boldsymbol{s})$ in (5), at $d = 9, 25, 100$ sites on the regular grid $\left\{ (j/(m+1), k/(m+1))^\top \right\}_{j,k=1}^m \subset [0,1]^2$, for $m = 3, 5, 10$. To illustrate the methods, we consider here the compactly supported kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) = (1 - \|\boldsymbol{s} - \boldsymbol{s}^*\|/r)_+^\eta$ with true kernel parameters chosen as $\boldsymbol{\theta}_K = (\eta, r)^\top = (1, 0.25)^\top$. For the spatial mixture process $\tilde{Z}(\boldsymbol{s})$ in (6), we additionally specify the correlation function of the Gaussian process to be $\rho_G(\delta; \theta_G) = \exp(-\theta_G \delta)$ with rate $\theta_G = 1$, and the mixture parameter is fixed to $\beta = 2$.

Following the notation from Section 3.2, we need to find the theoretical expression of the scale parameter $c_{jk}(\boldsymbol{\theta}_K)$ in (7). By symmetry, straightforward calculations yield

$$
\begin{aligned}
c_{jk}(\boldsymbol{\theta}_K) &= \int_{\mathbb{R}^2} |\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) - \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)| \, \mathrm{d}\boldsymbol{s}^* \\
&= 2 \int_{\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) > \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)} \zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^* + 2 \int_{\zeta(\boldsymbol{s}_j, \boldsymbol{s}^*) < \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*)} \zeta(\boldsymbol{s}_k, \boldsymbol{s}^*) \mathrm{d}\boldsymbol{s}^* - 2 \\
&= \frac{4}{c^*(\boldsymbol{\theta}_K)} \int_{\|\boldsymbol{s}^* - \boldsymbol{s}_k\| \leq \min\{r, \|\boldsymbol{s}_j - \boldsymbol{s}^*\|\}} \left(1 - \frac{\|\boldsymbol{s}_k - \boldsymbol{s}^*\|}{r} \right)^\eta \mathrm{d}\boldsymbol{s}^* - 2,
\end{aligned}
$$

where $\zeta(\boldsymbol{s}, \boldsymbol{s}^*) = k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K)/ \int_{\mathbb{R}^2} k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) \mathrm{d}\boldsymbol{s}^*$ and the normalizing factor is equal to $c^*(\boldsymbol{\theta}_K) = \int_{\mathbb{R}^2} k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) \mathrm{d}\boldsymbol{s}^* = 2\pi r^2 / \{(\eta + 1)(\eta + 2)\}$. After a change of variables, we find that

$$
c_{jk}(\boldsymbol{\theta}_K) = \frac{4r^2}{c^*(\boldsymbol{\theta}_K)} \int_\pi^{2\pi} \left\{ \frac{g_{jk}(\phi; r)^{\eta+1}}{\eta + 1} - \frac{g_{jk}(\phi; r)^{\eta+2}}{\eta + 2} \right\} \mathrm{d}\phi
$$

$$= \frac{2(\eta+1)(\eta+2)}{\pi} \int_\pi^{2\pi} \left\{ \frac{g_{jk}(\phi;r)^{\eta+1}}{\eta+1} - \frac{g_{jk}(\phi;r)^{\eta+2}}{\eta+2} \right\} \mathrm{d}\phi,$$

where $g_{jk}(\phi;r) = 1 - \max\{0, 1 + \|\boldsymbol{s}_j - \boldsymbol{s}_k\|/(2r\sin\phi)\}$. This integral can be easily computed numerically. Furthermore, from (10), we can also deduce that $\lambda_{jk}(\boldsymbol{\theta}_K) = 1 - c_{jk}(\boldsymbol{\theta}_K)/2$.

For each simulated dataset we use the Student-t marginal distribution with four degrees of freedom for the processes $Z(\boldsymbol{s})$ and $\tilde{Z}(\boldsymbol{s})$, and the Fréchet marginal distribution with the shape parameter 4 for the process $Z_{\mathrm{EV}}(\boldsymbol{s})$. Degrees of freedom and shape parameters are estimated in a first step using the marginal likelihood approach described in Section 3.1. Dependence (i.e., copula) parameters are then estimated in a second step based on a pairwise least squares inference approach (recall Sections 3.2, 3.3, and 3.4). Here, we set the weights to $\omega_{jk}^+ = \omega_{jk}^- = 1$ for all pairs of sites $\{\boldsymbol{s}_j, \boldsymbol{s}_k\}$ with $\|\boldsymbol{s}_j - \boldsymbol{s}_k\| < \delta_{\max}$ in (8), (11), and (13), where we use $\delta_{\max} = 0.4$ for $d = 9, 25$ and $\delta_{\max} = 0.25$ for $d = 100$. Other weights are set to zero, so we only include pairs of locations at small distances to improve the accuracy of estimates and make computations faster, especially for $d = 100$. More specifically, the values of $\delta_{\max}$ are chosen to keep 20, 150, and 790 close-by pairs for inference, i.e., about 56%, 50%, and 16% of all pairs for $d = 9, 25, 100$, respectively, in order to achieve a reasonable trade-off between computational and statistical efficiency. Parameters to be estimated are the kernel parameters $\boldsymbol{\theta}_K = (\eta, r)^\top$ for $Z(\boldsymbol{s})$ and $Z_{\mathrm{EV}}(\boldsymbol{s})$, and $(\eta, r, \theta_G, \beta)^\top$ for the spatial mixture process $\tilde{Z}(\boldsymbol{s})$. To assess the estimators' performance, we repeat the simulations $N = 500$ times and compute the root mean squared errors (RMSE) for all estimated parameters. We also compute $\Delta_{\max} = \frac{1}{N}\sum_{i=1}^N \max_{x\in(0,1)} |k(\boldsymbol{s}_0, \boldsymbol{s}(x); \boldsymbol{\theta}_K) - k(\boldsymbol{s}_0, \boldsymbol{s}(x); \widehat{\boldsymbol{\theta}}_{K,i})|$, and $\Delta_{\mathrm{avg}} = \frac{1}{N}\sum_{i=1}^N \int_0^1 |k(\boldsymbol{s}_0, \boldsymbol{s}(x); \boldsymbol{\theta}_K) - k(\boldsymbol{s}_0, \boldsymbol{s}(x); \widehat{\boldsymbol{\theta}}_{K,i})|\mathrm{d}x$, where $\boldsymbol{s}_0 = (0,0)^\top$ and $\boldsymbol{s}(x) = (0,x)^\top$, while $\boldsymbol{\theta}_K = (1, 0.25)^\top$ denotes the true kernel parameters and $\widehat{\boldsymbol{\theta}}_{K,i}$ is its estimate for the i-th simulation $(i = 1, \ldots, N)$. Hence, $\Delta_{\max}$ and Δ_{avg} represent the maximum and mean integrated absolute differences between the true kernel and its estimate along a horizontal segment passing through the origin, averaged across the $N = 500$ simulations.

Table 1 reports the results. As expected, estimates are more accurate with larger sample sizes, as shown by significantly smaller RMSE and $(\Delta_{\max}, \Delta_{\mathrm{avg}})$ values as n increases. Moreover, using data from more locations (i.e., increasing d) can further improve the accuracy of parameter estimates.

For small n and d, RMSE values are quite large due to kernel parameters being more tricky to identify. Similar kernels (hence, dependence structures) may be obtained for different combinations of parameters $\boldsymbol{\theta}_K = (\eta, r)^\top$, and thus the effects of η (shape) and ρ (dependence range) are difficult to distinguish, especially in low sample sizes. Figure 5 shows the estimated kernel profiles along the x-axis, i.e., $k(\boldsymbol{s}_0, \boldsymbol{s}(x), \widehat{\boldsymbol{\theta}}_{K,i})$ with $0 < x < 1$, for each simulated dataset of $Z(\boldsymbol{s})$ and $Z_{\mathrm{EV}}(\boldsymbol{s})$ with $d = 100$. The true kernel appears to be nevertheless very well estimated, even when the sample size is not very large. Similar results are obtained with other sets of parameters and with different kernel functions.

TABLE 1

Performance metrics (i.e., RMSE of estimated parameters and $(\Delta_{\max}, \Delta_{\mathrm{avg}})$) calculated for the least squares inference schemes detailed in Sections 3.2, 3.3, and 3.4 for data simulated from the processes $Z(s)$ (top panel), $\tilde{Z}(s)$ (middle panel), and $Z_{\mathrm{EV}}(s)$ (bottom panel), respectively. The true kernel function is chosen as $k(s, s^*; \theta_K) = (1 - \|s - s^*\|/r)_+^\eta$, with parameters $\theta_K = (\eta, r)^\top = (1, 0.25)^\top$. For the spatial mixture process $\tilde{Z}(s)$, the correlation function of the underlying Gaussian process is $\rho_G(\delta; \theta_G) = \exp(-\theta_G \delta)$ with $\theta_G = 1$, and the mixture parameter is $\beta = 2$. Datasets are simulated on a regular grid at $d = 9, 25, 100$ locations, with $n = 100, 200, 500$ independent replicates.

Data simulated from the Cauchy process $Z(s)$ in (2), with inference based on Section 3.2.

| Sample size | RMSE for $\theta_K^\top = (\eta, r)$ (top) and $(\Delta_{\max}, \Delta_{\mathrm{avg}})$ (bottom) | | |
	$d = 9$	$d = 25$	$d = 100$
$n = 100$	(2.49, 0.16)	(1.62,0.11)	(0.30,0.02)
	(0.21,0.03)	(0.14,0.02)	(0.07,0.01)
$n = 200$	(1.86,0.12)	(0.69,0.05)	(0.22,0.02)
	(0.16,0.03)	(0.08,0.01)	(0.05,0.01)
$n = 500$	(0.85,0.05)	(0.26,0.02)	(0.13,0.01)
	(0.11,0.02)	(0.04,0.01)	(0.03,0.00)

Data simulated from the spatial mixture process $\tilde{Z}(s)$ in (6), with inference based on Section 3.3.

| Sample size | RMSE for $(\eta, r, \theta_G, \beta)$ (top) and $(\Delta_{\max}, \Delta_{\mathrm{avg}})$ (bottom) | | |
	$d = 9$	$d = 25$	$d = 100$
$n = 100$	(1.77,0.14,0.76,0.86)	(1.02,0.08,0.67,0.75)	(1.00,0.08,0.52,0.74)
	(0.36,0.05)	(0.15,0.02)	(0.11,0.02)
$n = 200$	(1.80,0.14,0.73,0.79)	(0.78,0.06,0.65,0.73)	(0.63,0.04,0.46,0.64)
	(0.33,0.05)	(0.13,0.02)	(0.09,0.01)
$n = 500$	(1.52,0.11,0.56,0.61)	(0.45,0.03,0.51,0.55)	(0.52,0.03,0.30,0.40)
	(0.21,0.03)	(0.09,0.01)	(0.08,0.01)

Data simulated from the extreme-value process $Z_{\mathrm{EV}}(s)$ in (5), with inference based on a least squares approach based on pairwise tail dependence coefficients (see Appendix).

| Sample size | RMSE for $\theta_K^\top = (\eta, r)$ (top) and $(\Delta_{\max}, \Delta_{\mathrm{avg}})$ (bottom) | | |
	$d = 9$	$d = 25$	$d = 100$
$n = 100$	(1.63,0.12)	(0.65,0.05)	(0.31,0.03)
	(0.26,0.04)	(0.10,0.02)	(0.06,0.01)
$n = 200$	(1.30,0.09)	(0.43,0.03)	(0.22,0.02)
	(0.17,0.03)	(0.08,0.01)	(0.04,0.01)
$n = 500$	(0.82,0.06)	(0.28,0.02)	(0.13,0.01)
	(0.12,0.02)	(0.05,0.01)	(0.03,0.00)

4.2. Temperature data application

We now illustrate the proposed methodology to analyze temperature data from the state of Oklahoma, United States. Here, we use daily temperature averages measured at 97 monitoring stations at maximum distance 531 km. The time period considered here is the year 2018, which contains $n = 366$ days in total. The dataset can be freely downloaded from the website `mesonet.org`. After removing the obvious seasonal component, we then fit an `AR(2)` model to account for temporal dependence, and we fit the skew-t distribution jointly to the residuals using the marginal likelihood approach described in Section 3.1. Finally,
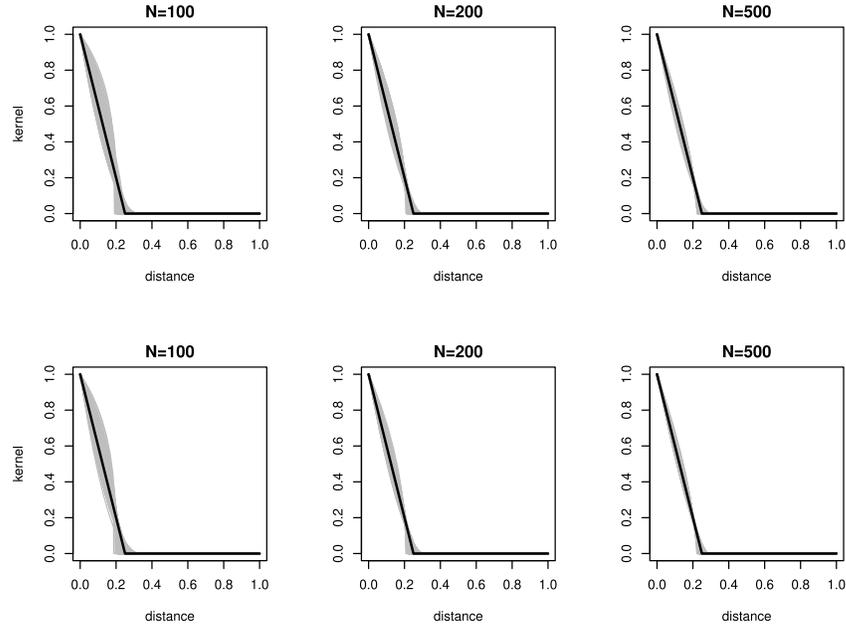
FIG 5. *Estimated kernel profiles* $k(s_0, s(x); \widehat{\theta}_{K,i})$, $0 < x < 1$, $i = 1, \ldots, N$ *(light gray lines), based on* $N = 500$ *simulated processes* $Z(s)$ *(top row) and extreme-value limit* $Z_{EV}(s)$ *(bottom row), with* $n = 100$ *(left column),* 200 *(middle column) and* 500 *(right column) independent replicates. Black thick solid lines show the true kernels* $k(s_0, s(x; \theta_K))$ *with* $\theta_K = (1, 0.25)^{\top}$.

we transform the residuals to the $\text{Unif}(0, 1)$ scale using the estimated marginal distribution functions. To explore the dependence features of the data, the left panel of Figure 6 displays bivariate scatter plots of normal scores (i.e., residuals further transformed to the standard normal distribution) for some selected pairs of stations. The sharp and non-elliptical tails that can be seen in all bivariate scatter plots indicate clear evidence of non-Gaussianity and strong tail dependence. Furthermore, these diagnostics do not reveal any significant tail asymmetry. Therefore, while a (transformed) Gaussian process would clearly not be suitable to model the data due to its very weak form of dependence in the tails, our proposed Cauchy convolution process provides a more adequate alternative. Furthermore, the right panel of Figure 6 shows Spearman's correlation estimates for all pairs of stations, plotted as a function of distance between stations. Although these empirical estimates appear quite dispersed, this plot reveals that the behavior of Spearman's correlation function is approximately linear near the origin, which suggests that the observed process is quite smooth and that our Cauchy convolution model should fit well.

All the stations are located in a relatively small geographical region in Oklahoma State, and we excluded stations in the mountainous area in the northwestern part of the state. The largest distance between any two stations is about 531 km, so we can reasonably assume that the data are stationary over space.
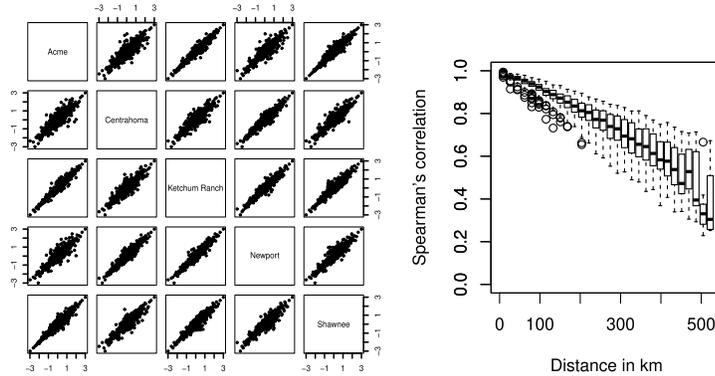
FIG 6. Left: *Bivariate scatter plots of normal scores based on the temperature data residuals (obtained from preliminary marginal fits) for some selected pairs of stations.* Right: *Spearmans's correlation estimates for all pairs of stations, plotted as a function of distance between stations using boxplots for different distance classes.*

We select the kernel function $k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) = (1 - \|\boldsymbol{s} - \boldsymbol{s}^*\|/r)_+^\eta$, $\boldsymbol{\theta}_K = (\eta, r)^\top \in (0, \infty)^2$, and estimate its parameters using the inference approach described in Section 3.3 using close-by pairs at maximum distance $\delta_{\max} = 300$ km. Notice that although this kernel is compactly supported on $[0, r]$ with $r > 0$ representing the dependence range, it reduces to the exponential kernel $k(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_K) = \exp(-\xi\|\boldsymbol{s} - \boldsymbol{s}^*\|)$ when $\eta := \eta(r)$ is such that $\eta(r) \to \infty$ and $\eta(r)/r \to \xi > 0$ as $r \to \infty$. The estimated kernel parameters are $\widehat{\boldsymbol{\theta}}_K = (\widehat{\eta}, \widehat{r})^\top = (8985, 1270)^\top$ with the estimated range $\widehat{r} > 0$ expressed in units of 1000 km. Thus, as expected, our results here imply fairly long range dependence for this temperature dataset, with $k(\boldsymbol{s}, \boldsymbol{s}^*) \approx e^{-7.08\|\boldsymbol{s} - \boldsymbol{s}^*\|}$; the fitted kernel function is plotted on the left panel of Figure 7, from which the exponential decay is evident. Nevertheless, notice that our modeling approach has the great benefit of estimating whether the dependence range is finite or infinite (obtained as a boundary limiting case), rather than fixing a priori. Moreover, even if the range parameter r is here much larger than the study region, the effective tail dependence range at which the estimated kernel function drops below 0.05 is only about 423 km.

We then estimate the remaining parameters of the spatial mixture process (6), where the underlying Gaussian process is specified with a powered exponential correlation function defined as $\rho_G(\delta) = \{1 - \tau\mathbb{I}(\delta \neq 0)\}\exp(-\theta_G\delta^\alpha)$, $\delta \geq 0$, where $\mathbb{I}(\cdot)$ is the indicator function, $\theta_G > 0$, $\alpha \in (0, 2]$, and $\tau \in [0, 1]$ is the nugget effect capturing small scale variability. More precisely, we here fix the kernel parameters to $\boldsymbol{\theta}_K = (8985, 1270)^\top$ as estimated above, and then estimate the parameters $(\theta_G, \alpha, \tau, \beta)^\top$ using the least squares approach (13) described in Section 3.3. We obtain $\widehat{\theta}_G = 0.90$ (0.29), $\widehat{\alpha} = 0.82$ (0.20), $\widehat{\tau} = 0.013$ (0.018), and $\widehat{\beta} = 1.50$ (0.31), with standard errors calculated using the bootstrap shown in parentheses. Since $\widehat{\beta}$ is positive and quite far from zero, the fitted spatial mixture process $\tilde{Z}(\boldsymbol{s})$ turns out to be quite different from the Cauchy convo-
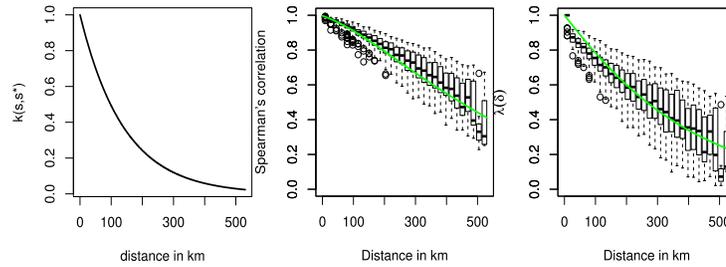
FIG 7. Left: *Fitted kernel function* $k(\boldsymbol{s}, \boldsymbol{s}^*; \widehat{\boldsymbol{\theta}}_K) = (1 - \|\boldsymbol{s} - \boldsymbol{s}^*\|/\widehat{r})_+^{\widehat{\eta}}$ *with* $\widehat{\boldsymbol{\theta}}_K = (\widehat{\eta}, \widehat{r})^\top = (8985, 1270)^\top$. Middle: *Empirical estimates of Spearman's correlation for all pairs of stations (boxplots) and fitted Spearman's correlation function* $S_\rho(\delta)$ *(solid green light), plotted as a function of distance* δ *between stations.* Right: *Empirical upper tail dependence coefficient estimates for all pairs of stations (boxplots) and fitted tail dependence function* $\lambda(\delta)$ *(solid green light), plotted as a function of distance* δ *between stations.*

lution model $Z(\boldsymbol{s})$, providing increased flexibility to capture the behavior in the bulk of the distribution. Furthermore, the fitted process corresponds to a smooth Cauchy convolution process mixed with a quite rough Gaussian random field (with smoothness parameter $\widehat{\alpha} = 0.82$), which yields realizations that are relatively—but not overly—smooth and thus more realistic. However, notice that the estimated nugget effect $\widehat{\tau}$ is here very small, indicating negligible micro-scale variability.

To assess the goodness-of-fit for the spatial mixture process (6), we then compute empirical and model-based Spearman's correlation estimates $S_\rho(\delta)$ for all pairs of sites $\{\boldsymbol{s}_j, \boldsymbol{s}_k\}$ and plot them in the middle panel of Figure 7 as a function of distance $\delta = \|\boldsymbol{s}_j - \boldsymbol{s}_k\|$. Model-based estimates are obtained by Monte Carlo using a large number of simulations from the fitted model. Similarly, the right panel of Figure 7 shows empirical and model-based estimates of the upper tail dependence coefficient $\lambda(\delta)$ plotted as function of distance δ. We obtain very similar results for the lower tail dependence coefficient (not shown) as the data show no significant tail asymmetry. These plots show that the model (6) fits the data very well, both the in the bulk of the distribution (as measured by $S_\rho(\delta)$) and in the tails (as measured by $\lambda(\delta)$).

For comparison, we also fit some alternative copula models to the same temperature dataset. We consider the following models: the copula stemming from the Cauchy convolution process in (2) (Model 1); the Gaussian copula (Model 2) with the same powered-exponential correlation function as above for the process Z_G; and the Student-t copula (Model 3) with $\nu > 0$ degrees of freedom and same underlying correlation function as Z_G and Model 2. Notice that Model 1 is a special case of our proposed model (6) when $\beta = 0$, and Model 2 can also be obtained as a limiting case of (6) as $\beta \to \infty$. We then compute empirical and model-based Spearman's correlation estimates $S_\rho(\delta)$ and the upper tail dependence coefficient $\lambda(\delta)$ for all pairs of sites $\{\boldsymbol{s}_j, \boldsymbol{s}_k\}$ and plot them in the first and second row of Figure 8 as a function of distance $\delta = \|\boldsymbol{s}_j - \boldsymbol{s}_k\|$. Model-based
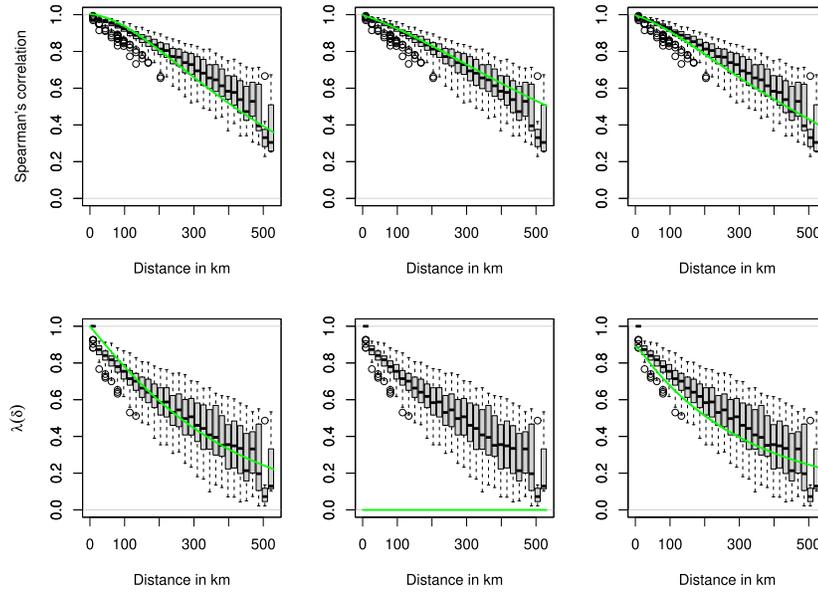
FIG 8. Top: *Empirical estimates of Spearman's correlation for all pairs of stations (boxplots) and fitted Spearman's correlation function $S_\rho(\delta)$ (solid green light), plotted as a function of distance δ between stations.* Bottom: *Empirical upper tail dependence coefficient estimates for all pairs of stations (boxplots) and fitted tail dependence function $\lambda(\delta)$ (solid green light), plotted as a function of distance δ between stations. Estimates are obtained using Model 1 (left), Model 2 (middle) and Model 3 (right).*

estimates are obtained by Monte Carlo using a large number of simulations from the fitted model.

Model 1 has a fairly good fit to the data, though the more flexible mixture process (6) has a slightly better fit in the bulk of the distribution, especially at relatively large distances. Model 2 has a good fit in the bulk of the distribution but cannot capture limiting tail dependence, and Model 3 tends to slightly underestimate the upper tail dependence structure, and is to rigid to separately control bulk and tail dependence characteristics. Overall, the spatial mixture process thus provides a better fit to the temperature data. We also recall here that our proposed mixture model (6) is the only one among the different models fitted that can capture short-range tail dependence and long-range tail independence, such that the proposed model should provide good fits on larger domains, as well.

5. Concluding remarks

In this paper, we have proposed a Cauchy kernel convolution copula model for non-Gaussian spatial data and studied its dependence properties, with a particular eye on its tail behavior. In particular, with compactly supported kernels,

our proposed model can capture complex dependence structures that possess short-range tail dependence, and long-range independence. Moreover, to further increase its flexibility in the bulk of the distribution and better capture the sub-asymptotic dependence behavior, we have also proposed a parsimonious copula model constructed by mixing a Cauchy convolution process with a Gaussian process. With this spatial mixture model extension, bulk and tail properties can be separately controlled with a few parameters, and a smooth transition between tail dependence classes is achieved as a function of distance between stations.

Our proposed inference scheme relies on a least-square minimum distance approach, which matches suitably chosen empirical and model-based summary statistics. It yields consistent estimates and it is guaranteed to be very fast even in high dimensions, thus circumventing the computational difficulties of likelihood-based inference. We have shown that our inference scheme works well using an extensive simulation study, and we have successfully applied it to a real temperature data example. Model parameters are generally easy to identify from each other, and the underlying kernel function can be accurately estimated, even in low sample sizes.

A limitation of our approach is that, by construction, the proposed Cauchy model and its spatial mixture extension are tail symmetric, and can only capture smooth extreme-value dependence structures characterized by moving maximum processes similar to the well-known Smith max-stable process [63]. However, as opposed to fitting the (rigid) Smith model directly to spatially-indexed block maxima, we here propose to fit the Cauchy convolution process or the flexible spatial mixture process to the whole dataset instead. Therefore, even if the limiting extreme-value dependence structure is very smooth, our proposed models, which have much rougher realizations, can still adequately capture dependence characteristics at finite levels.

Nevertheless, the sub-asymptotic tail structure of the spatial mixture extension of the Cauchy model is still quite restrictive in the sense that there is a discontinuity in the coefficient of tail dependence η (reciprocal of the tail order) as a function of distance, being equal to $\eta = 1$ for short distances and jumping to $\eta = 1/2$ at sufficiently large distances. The reason is that in the sum-mixture process $Z(\boldsymbol{s}) + \beta Z_G(\boldsymbol{s})$, the heavy-tailed Cauchy convolution process $Z(\boldsymbol{s})$ dominates the Gaussian process $Z_G(\boldsymbol{s})$ in the tail. Unreported proofs suggest that this would remain the case if the margins of the $Z_G(\boldsymbol{s})$ process were replaced with a distribution with Weibull-like tails (i.e., potentially heavier-tailed than the Gaussian distribution but still lighter-tailed than the power-law Cauchy tails). However, despite this relatively restrictive limiting structure in the tail-independent case, we have shown through various theoretical diagnostics and goodness-of-fit plots in the data application that our model still possesses high flexibility at finite levels, all the way from low to high quantiles, separately capturing a wide range of different bulk and tail dependencies as a function of distance. Other constructions (e.g., considering product-mixtures rather than sum-mixtures) might be helpful to capture a smoother sub-asymptotic tail behavior, though this is still unclear and we leave this as a topic for future research.

Another important problem is conditional simulation, i.e., simulation of the Cauchy convolution process conditional on the observed values of this process at some locations or conditional on the value of a spatial aggregation functional. This is a difficult problem for non-Gaussian kernel convolution processes since the likelihood function is intractable. One way to tackle conditional simulation could be to use some version of rejection sampling given that approximate simulations from the proposed model can be performed very fast. While such a naive approach would be computationally inefficient when performing simulations conditional on fixed values at a moderate or large number of spatial locations, it could prove helpful when conditioning on the value of a single aggregation functional (e.g., the sum or maximum across sites). Approaches based on exponential tilting and importance sampling [5, 7] could potentially be adapted to further enhance conditional simulation, especially when the conditioning event is a low-probability rare event.

Further interesting research directions include generalizing our modeling approach to capture tail asymmetry, e.g., by considering kernel convolution processes constructed from asymmetric stable noise. It would also be interesting to study how to modify our model to capture rougher tail dependence structures of Brown–Resnick type.

Appendix A: Proofs

Lemma 1 (Lower tail probability). *For $j = 1, \ldots, d$, let $Y_j^* = \sum_{k=1}^{K} a_{k,j}(m) Y_k$, where $a_{k,j}(m) \geq 0$ and $Y_1, \ldots, Y_K \sim_{i.i.d.}$ Cauchy(1). Let $c_j(m,n) > 0$ and $\gamma_k = \gamma_k(m,n) := \max_{j=1,\ldots,d} \{a_{k,j}(m)/c_j(m,n)\}$ and assume that $\gamma_k \leq c_0/(nK)$ for any m, k and some constant $c_0 > 0$. Then for any m and large enough n,*

$$
\left| \Pr\{Y_1^* \leq c_1(m,n), \ldots, Y_d^* \leq c_d(m,n)\} - \left(1 - \frac{1}{\pi} \sum_{k=1}^{K} \gamma_k\right) \right| \leq \frac{(d+1)c_0^2}{\pi^2 n^2}.
$$

Proof. For the standard Cauchy CDF $F_{\mathcal{C}}$ and $z > 1$, we have

$$
1 - \frac{1}{\pi z} \leq F_{\mathcal{C}}(z) \leq 1 - \frac{1}{\pi z} + \frac{1}{3\pi z^3}.
$$

By stability of the Cauchy distribution, we find thus that, for each $j = 1, \ldots, d$ and large enough n,

$$
\begin{aligned}
\delta_j(m,n) &= \Pr\left\{ \sum_{k=1}^{K} a_{k,j}(m) Y_k \leq c_j(m,n) \right\} \\
&\quad - \Pr\left\{ Y_1 \leq \frac{c_j(m,n)}{a_{1,j}(m)}, \ldots, Y_K \leq \frac{c_j(m,n)}{a_{K,j}(m)} \right\} \\
&\leq 1 - \frac{\sum_{k=1}^{K} a_{k,j}(m)}{\pi c_j(m,n)} \left(1 - \frac{c_0^2}{3n^2} \right) - \prod_{k=1}^{K} \left\{ 1 - \frac{a_{k,j}(m)}{\pi c_j(m,n)} \right\}
\end{aligned}
$$

$$\leq \quad 1 - \frac{\sum_{k=1}^{K} a_{k,j}(m)}{\pi c_j(m,n)} \left(1 - \frac{c_0^2}{3n^2} \right) - 1 + \frac{\sum_{k=1}^{K} a_{k,j}(m)}{\pi c_j(m,n)}$$

$$\leq \quad \frac{c_0^3}{3\pi n^3} \leq \frac{c_0^2}{\pi^2 n^2},$$

$$\delta_j(m,n) \quad \geq \quad 1 - \frac{\sum_{k=1}^{K} a_{k,j}(m)}{\pi c_j(m,n)} - \prod_{k=1}^{K} \left\{ 1 - \frac{a_{k,j}(m)}{\pi c_j(m,n)} \left(1 - \frac{c_0^2}{3\pi n^2 K^2} \right) \right\}$$

$$\geq \quad -\frac{c_0^3}{3\pi^2 n^3 K} - \left(1 - \frac{c_0^2}{3\pi n^2 K^2} \right)^2 \sum_{k<l} \frac{a_{k,j}(m) a_{l,j}(m)}{\pi^2 c_j(m,n)^2}$$

$$\geq \quad -\frac{c_0^3}{3\pi^2 n^3 K} - \left(1 - \frac{c_0^2}{3\pi n^2 K^2} \right)^2 \frac{c_0^2}{2\pi^2 n^2} \geq -\frac{c_0^2}{\pi^2 n^2}.$$

Therefore, with $p_{m,n} = \Pr\{Y_1^* \leq c_1(m,n), \ldots, Y_d^* \leq c_d(m,n)\}$,

$$\delta(m,n) = \left| p_{m,n} - \Pr\left(Y_1 \leq \frac{1}{\gamma_1}, \ldots, Y_d \leq \frac{1}{\gamma_d} \right) \right| \leq \sum_{j=1}^{d} |\delta_j(m,n)| \leq \frac{d c_0^2}{\pi^2 n^2}.$$

Let $p_K^* = \Pr\left(Y_1 \leq \frac{1}{\gamma_1}, \cdots, Y_K \leq \frac{1}{\gamma_K} \right)$. Now,

$$p_K^* \quad = \quad \prod_{k=1}^{K} F_{\mathcal{C}} \left(\frac{1}{\gamma_k} \right) \leq \prod_{k=1}^{K} \left\{ 1 - \frac{\gamma_k}{\pi} \left(1 - \frac{c_0^2}{3n^2} \right) \right\}$$

$$\leq \quad 1 - \frac{1}{\pi} \left(1 - \frac{c_0^2}{3n^2} \right) \sum_{k=1}^{K} \gamma_k + \frac{1}{\pi^2} \left(1 - \frac{c_0^2}{3n^2} \right)^2 \sum_{k<l}^{K} \gamma_k \gamma_l$$

$$\leq \quad 1 - \frac{1}{\pi} \sum_{k=1}^{K} \gamma_k + \frac{c_0^3}{3\pi n^3} + \frac{c_0^2}{2\pi^2 n^2} \leq 1 - \frac{1}{\pi} \sum_{k=1}^{K} \gamma_k + \frac{c_0^2}{\pi^2 n^2},$$

$$p_K^* \quad \geq \quad \prod_{k=1}^{K} \left(1 - \frac{\gamma_k}{\pi} \right) \geq 1 - \frac{1}{\pi} \sum_{k=1}^{K} \gamma_k.$$

Hence, $\left| p_{m,n} - \left(1 - \frac{1}{\pi} \sum_{k=1}^{K} \gamma_k \right) \right| \leq (d+1) c_0^2 / (\pi^2 n^2)$ as required. $\qquad \square$

Proof of Proposition 1. Define the expanding, but increasingly dense rectangular grid $S_m \subset \mathbb{R}^2$ as

$$S_m = \{\boldsymbol{s}_{k,l}^*\}_{k,l=1}^{m} = \left\{ -\frac{(m-1)\delta_m}{2}, -\frac{(m-3)\delta_m}{2}, \cdots, \frac{(m-3)\delta_m}{2}, \frac{(m-1)\delta_m}{2} \right\}^2,$$

with $\delta_m \to 0$ such that $m\delta_m \to \infty$, as $m \to \infty$. This implies that $\int_{S_m} h(x) \mathrm{d}x \to \int_{\mathbb{R}^2} h(x) \mathrm{d}x$, $m \to \infty$, for any integrable function $h(x)$. Define

$$Z_{j,m} = \delta_m^2 \sum_{k,l=1}^{m} k(\boldsymbol{s}_j, \boldsymbol{s}_{k,l}^*) W_{k,l}, \quad W_{k,l} = W(\boldsymbol{s}_{k,l}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(1). \qquad (15)$$

Intuitively, for large m, $Z_{j,m}$ is a finite approximation of $Z_j = Z(\boldsymbol{s}_j)$. Formally, we indeed have $(Z_{1,m}, \ldots, Z_{d,m})^\top \to_d (Z_1, \ldots, Z_d)^\top$ as $m \to \infty$, where "$\to_d$" denotes convergence in distribution. It can be seen that (15) is a linear factor model with m^2 independent and identically distributed common factors $W_{k,l}$ as considered in Krupskii and Genton [42].

Let $F_{j,m}$ be the CDF of $Z_{j,m}$, $j = 1, \ldots, d$. Define

$$\ell_{m,n}(w_1, \ldots, w_d) = n \left[1 - \Pr \left\{ Z_{1,m} \le F_{1,m}^{-1}(1 - \frac{w_1}{n}), \ldots, Z_{p,m} \le F_{d,m}^{-1}(1 - \frac{w_d}{n}) \right\} \right].$$

The stable tail dependence function associated with the random vector $(Z_{1,m}, \ldots, Z_{d,m})^\top$ is

$$\ell_m(w_1, \ldots, w_d) = \lim_{n \to \infty} \ell_{m,n}(w_1, \ldots, w_d).$$

By Moore-Osgood Theorem (see Angus [2], p. 140), we need to prove that $\ell_{m,n}(w_1, \ldots, w_d)$ converges uniformly to $\ell_m(w_1, \ldots, w_d)$, as $n \to \infty$.

By stability of the Cauchy distribution, we have $Z_{j,m} \sim \text{Cauchy}(c_{j,m})$ with scale parameter $c_{j,m} = \delta_m^2 \sum_{k,l=1}^m k(\boldsymbol{s}_j, \boldsymbol{s}_{k,l}^*)$, and therefore

$$F_{j,m}^{-1}\left(1 - \frac{w_j}{n} \right) = c_{j,m} \cot \left(\frac{\pi w_j}{n} \right), \qquad j = 1, \ldots, d.$$

Using the Laurent series of the cotangent function around the origin with $\cot(z) = 1/z - z/3 + \epsilon(z)$, $|\epsilon(z)| < z^3$ for $z < 1$, we deduce that

$$\frac{nc_{j,m}}{\pi w_j}\left(1 - \frac{c_L}{n^2} \right) \le \frac{nc_{j,m}}{\pi w_j} - \frac{\pi w_j c_{j,m}}{2n} \le F_{j,m}^{-1}\left(1 - \frac{w_j}{n} \right) \le \frac{nc_{j,m}}{\pi w_j}$$

for any m if n is large enough, where $c_L = 2^{-1}\pi^2 \max_{j=1,\ldots,d} w_j^2$, is a constant that does not depend on m and n. We now use Lemma 1. Let $p_{m,n} = p_{m,n}(w_1, \ldots, w_d) = \Pr\{Z_{1,m} \le \eta_{1,m}(w_1, n), \ldots, Z_{d,m} \le \eta_{d,m}(w_d, n)\}$. Using Lemma 1 with $K = m^2$, $c_j(m,n) = F_{j,m}^{-1}\left(1 - \frac{w_j}{n} \right)$ and $c_0 = 2\pi \max_{\boldsymbol{s},j}\{w_j \zeta(\boldsymbol{s}_j, \boldsymbol{s})\}$, we find for $\ell_{m,n}(w_1, \ldots, w_d) = n\{1 - p_{m,n}(w_1, \ldots, w_d)\}$ that

$$\begin{aligned}
|\ell_{m,n}(w_1, \ldots, w_d) - \ell_m^*(w_1, \ldots, w_d)| &\le \frac{(d+1)c_0^2}{\pi^2 n}, \\
|\ell_m^*(w_1, \ldots, w_d) - \ell_m(w_1, \ldots, w_d)| &\le \frac{c_L}{(n^2 - c_L)} \ell_m(w_1, \ldots, w_d) \\
&\le \frac{c_L}{(n^2 - c_L)} M(w_1, \ldots, w_d),
\end{aligned}$$

where $M(w_1, \ldots, w_d) = \max_m \ell_m(w_1, \ldots, w_d)$ and

$$\begin{aligned}
\ell_m^*(w_1, \ldots, w_d) &= \frac{n\delta_m^2}{\pi} \sum_{k,l=1}^m \max_{j=1,\ldots,d} \left\{ \frac{k(\boldsymbol{s}_j, \boldsymbol{s}_{k,l}^*)}{F_{j,m}^{-1}\left(1 - \frac{w_j}{n} \right)} \right\}, \\
\ell_m(w_1, \ldots, w_d) &= \delta_m^2 \sum_{k,l=1}^m \max_{j=1,\ldots,d} \left\{ \frac{w_j k(\boldsymbol{s}_j, \boldsymbol{s}_{k,l}^*)}{c_{j,m}} \right\},
\end{aligned}$$

It follows that $\ell_{m,n}(w_1,\ldots,w_d) \to \ell_m(w_1,\ldots,w_d)$, as $n \to \infty$, uniformly over m.

This implies that the stable tail dependence function associated with the d-dimensional random vector $(Z_1,\ldots,Z_d)^\top$ can be calculated as $\ell(w_1,\ldots,w_d) = \lim_{n\to\infty}\lim_{m\to\infty}\ell_{m,n}(w_1,\ldots,w_d) = \lim_{m\to\infty}\lim_{n\to\infty}\ell_{m,n}(w_1,\ldots,w_d) = \lim_{m\to\infty}\ell_m(w_1,\ldots,w_d) = \int_{\mathbb{R}^2}\max_{j=1,\ldots,d}w_j\zeta(\boldsymbol{s}_j,\boldsymbol{s}^*)\mathrm{d}\boldsymbol{s}^*$, and the result of the proposition follows. $\qquad\square$

Proof of Proposition 2. We first prove the second part of the proposition. Without loss of generality, we assume that $\int_{\mathbb{R}^2}g(\|\boldsymbol{s}^*\|)\mathrm{d}\boldsymbol{s}^* = 1$. Using stationarity of the kernel and an orthonormal transform of the integration variables, we obtain from Proposition 1 that

$$
\begin{aligned}
\ell(1,1) &= \int_{\mathbb{R}^2}\max_{j=1,2}g(\|\boldsymbol{s}_j - \boldsymbol{s}^*\|)\mathrm{d}\boldsymbol{s}^*\\
&= \int_{\mathbb{R}^2}\max\{g(\|(s_1^*,s_2^*)^\top\|), g(\|(s_1^*+\delta,s_2^*)^\top\|)\}\mathrm{d}\boldsymbol{s}^*, \qquad \boldsymbol{s}^* = (s_1^*,s_2^*)^\top,\\
&= 1 + \int_{\{\boldsymbol{s}^*\in\mathbb{R}^2:s_1^*<-\delta/2\}}\{g(\|(s_1^*+\delta,s_2^*)^\top\|) - g(\|(s_1^*,s_2^*)^\top\|)\}\mathrm{d}\boldsymbol{s}^*\\
&= 1 + \int_{-\delta/2}^{\delta/2}\int_{\mathbb{R}}g(\|(s_1^*,s_2^*)^\top\|)\mathrm{d}s_2^*\mathrm{d}s_1^* = 1 + \delta\int_{\mathbb{R}}g(\|(0,s_2^*)^\top\|)\mathrm{d}s_2^* + o(\delta),
\end{aligned}
$$

so one can take $K = \int_{\mathbb{R}}g(\|(0,s_2^*)^\top\|)\mathrm{d}s_2^*$.

Now we prove the first part of the proposition. Without loss of generality, we assume that $w_1 > w_2$. We get $\ell(w_1,w_2) - w_1 = \int_{S(w_1,w_2)}\{w_2g(\|(s_1^*+\delta,s_2^*)^\top\|) - w_1g(\|(s_1^*,s_2^*)^\top\|)\}\mathrm{d}\boldsymbol{s}^*$, where $S(w_1,w_2) = \{\boldsymbol{s}^* \in \mathbb{R}^2 : w_2g(\|(s_1^*+\delta,s_2^*)^\top\|) > w_1g(\|(s_1^*,s_2^*)^\top\|)\} \subset \{\boldsymbol{s}^* \in \mathbb{R}^2 : s_1^* < -\delta/2\}$, so

$$
\begin{aligned}
\ell(w_1,w_2) - w_1 &\leq w_2\int_{S(w_1,w_2)}\{g(\|(s_1^*+\delta,s_2^*)^\top\|) - g(\|(s_1^*,s_2^*)^\top\|)\}\mathrm{d}\boldsymbol{s}^*\\
&\leq w_2\int_{\{\boldsymbol{s}^*\in\mathbb{R}^2:s_2^*<-\delta/2\}}\{g(\|(s_1^*+\delta,s_2^*)^\top\|) - g(\|(s_1^*,s_2^*)^\top\|)\}\mathrm{d}\boldsymbol{s}^*\\
&= w_2\{K\delta + o(\delta)\} \leq w_1\{K\delta + o(\delta)\},
\end{aligned}
$$

where the last equality follows the same line of proof as above. This implies that $\ell(w_1,w_2) \leq \{1+K\delta+o(\delta)\}w_1$, which concludes the proof of the proposition. $\quad\square$

Proof of Corollary 1. It is easy to verify that $\lambda(\delta) = 2 - \ell(1,1)$, where ℓ is the stable tail dependence function for two sites $\boldsymbol{s}_1,\boldsymbol{s}_2 \in \mathbb{R}^2$ at distance $\delta = \|\boldsymbol{s}_1-\boldsymbol{s}_2\|$ from each other, and by Proposition 2, we therefore obtain $\lambda(\delta) = 1-K\delta+o(\delta)$. Moreover, as the function $\ell(w_1,w_2)$ in non-decreasing in both arguments and homogeneous of order 1, we get $\ell(-\log u_1, -\log u_2) \leq -\log\{\min(u_1,u_2)\}\ell(1,1) = -\log\{\min(u_1,u_2)\}\{1 + K\delta + o(\delta)\}$ thanks to Proposition 2. This yields

$$
S_\rho(\delta) = 12\iint_{[0,1]^2}C_{\mathrm{EV}}(u_1,u_2)\mathrm{d}u_1\mathrm{d}u_2 - 3
$$

$$= 12 \iint_{[0,1]^2} \exp\{-\ell(-\log u_1, -\log u_2)\} \mathrm{d}u_1 \mathrm{d}u_2 - 3$$

$$\geq 12 \iint_{[0,1]^2} \min(u_1, u_2)^{1+K\delta} \mathrm{d}u_1 \mathrm{d}u_2 + o(\delta) - 3$$

$$= 12 \frac{2}{6 + 5K\delta + o(\delta)} - 3 + o(\delta)$$

$$= \frac{6 - 15K\delta + o(\delta)}{6 + 5K\delta + o(\delta)} + o(\delta) = 1 - \frac{10K\delta}{3} + o(\delta) > 1 - 4K\delta, \quad \delta \to 0,$$

which concludes the proof. $\qquad\square$

Proof of Corollary 2. If $w_1/w_2 > G(\delta)$, then $S(w_1, w_2) = \{(s_1^*, s_2^*)^\top \in \mathbb{R}^2 : w_2 g(\|(s_1^* + \delta, s_2^*)^\top\|) > w_1 g(\|(s_1^*, s_2^*)^\top\|)\}$ is the empty set, i.e., $S(w_1, w_2) = \emptyset$, which implies that $\ell(w_1, w_2) = w_1$ by following the proof of Proposition 2. The second part of the corollary follows by symmetry. $\qquad\square$

Proof of Proposition 3. Without loss of generality, we assume that $\int_{\mathbb{R}^2} k(s, s^*) \mathrm{d}s^* = 1$ for any $s \in \mathbb{R}^2$. Let $F_{\mathcal{C}}^{-1}(\cdot)$ be the quantile function of the standard Cauchy distribution. Denote $Z_j = Z(s_j)$, $Z_{G,j} = Z_G(s_j)$, $\tilde{Z}_j = \tilde{Z}(s_j)$, and $h_j = h_j(w_j, n) = F_{\mathcal{C}}^{-1}(1 - w_j/n)$, $j = 1, \ldots, d$. We have:

$$\Pr(\tilde{Z}_1 \leq h_1, \ldots, \tilde{Z}_d \leq h_d) = \Pr(\tilde{Z}_1 \leq h_1, \ldots, \tilde{Z}_d \leq h_d, \max_{j=1,\ldots,d} |\beta Z_{G,j}| \leq \sqrt{n}) + \epsilon_n,$$

where, thanks to Mills's ratio,

$$0 < \epsilon_n \leq \Pr(\max_{j=1,\ldots,d} |\beta Z_{G,j}| > \sqrt{n}) < \sum_{j=1}^{d} \Pr(|\beta Z_{G,j}| > \sqrt{n})$$

$$\leq d|\beta|\phi(\sqrt{n}/|\beta|)/\sqrt{n} = o(e^{-n/(2\beta^2)}),$$

and $\phi(\cdot)$ denotes the density function of the standard normal distribution. Let

$$p^* = p^*(w_1, \ldots, w_d, n) = \Pr(\tilde{Z}_1 \leq h_1, \ldots, \tilde{Z}_d \leq h_d, \max_{j=1,\ldots,d} |\beta Z_{G,j}| \leq \sqrt{n}).$$

We find that

$$p^* \leq \Pr(Z_1 \leq h_1 + \sqrt{n}, \ldots, Z_d \leq h_d + \sqrt{n}) = \Pr(Z_1 \leq h_1, \ldots, Z_d \leq h_d) + o(n^{-1}),$$

$$p^* \geq \Pr(Z_1 \leq h_1 - \sqrt{n}, \ldots, Z_d \leq h_d - \sqrt{n}, \max_{j=1,\ldots,d} |\beta Z_{G,j}| \leq \sqrt{n})$$

$$= \Pr(Z_1 \leq h_1 - \sqrt{n}, \ldots, Z_d \leq h_d - \sqrt{n}) + o(e^{-n/(2\beta^2)})$$

$$= \Pr(Z_1 \leq h_1, \ldots, Z_d \leq h_d) + o(n^{-1}).$$

Let $\tilde{h}_j(w_j, n)$ be the $(1 - w_j/n)$-quantile of $\tilde{Z}_j$, $j = 1, \ldots, d$. Note that $\Pr(\tilde{Z}_j < h_j) = \Pr(Z_j < h_j) + o(n^{-1})$, so $\tilde{h}_j(w_j, n) = h_j(w_j, n) + o(n)$, and therefore

$$\Pr(\tilde{Z}_1 \leq \tilde{h}_1, \ldots, \tilde{Z}_d \leq \tilde{h}_d) = \Pr(\tilde{Z}_1 \leq h_1, \ldots, \tilde{Z}_d \leq h_d) + o(n^{-1})$$

$$= \Pr(Z_1 \leq h_1, \ldots, Z_d \leq h_d) + o(n^{-1}),$$

and $(Z_1, \ldots, Z_d)^\top$ and $(\tilde{Z}_1, \ldots, \tilde{Z}_d)^\top$ thus have the same stable tail dependence function. $\qquad\square$

Appendix B: Further special cases

Example 5 (Smith model [63] in $\mathbb{R}^2$)**.** *Assume that $q = 2$ and that $k(\boldsymbol{s}, \boldsymbol{s}^*) = \phi_2(\boldsymbol{s} - \boldsymbol{s}^*; \sigma, \rho)$ where $\phi_2(\cdot; \sigma, \rho)$ denotes the density function of the bivariate Gaussian distribution with zero mean, variances both equal to σ^2, and correlation $\rho \in (-1, 1)$. It follows that*

$$
\begin{aligned}
\ell(w_1, w_2) &= w_1 \int_{\frac{\phi_2(\boldsymbol{s}*+\Delta \boldsymbol{s}; \sigma, \rho)}{\phi_2(\boldsymbol{s}^*; \sigma, \rho)} \leq \frac{w_1}{w_2}} \phi_2(\boldsymbol{s}^*; \sigma, \rho) \mathrm{d}\boldsymbol{s}^* \\
&\quad + w_2 \int_{\frac{\phi_2(\boldsymbol{s}*-\Delta \boldsymbol{s}; \sigma, \rho)}{\phi_2(\boldsymbol{s}^*; \sigma, \rho)} \leq \frac{w_2}{w_1}} \phi_2(\boldsymbol{s}^*; \sigma, \rho) \mathrm{d}\boldsymbol{s}^* \\
&= w_1 \Phi\left(\frac{\lambda_2^*}{2} + \frac{1}{\lambda_2^*} \log \frac{w_1}{w_2} \right) + w_2 \Phi\left(\frac{\lambda_2^*}{2} + \frac{1}{\lambda_2^*} \log \frac{w_1}{w_2} \right),
\end{aligned}
$$

where $\Delta \boldsymbol{s} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ and $\lambda_2^* = \{\gamma(\Delta \boldsymbol{s})\}^{1/2}$ with variogram $\gamma(\mathbf{h}) = \sigma^{-2} \mathbf{h}^{\top} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \mathbf{h}$.

When $\rho = 0$, we obtain the isotropic variogram $\gamma(\mathbf{h}) = (\|\mathbf{h}\|/\sigma)^2$, which is similar to the case in $\mathbb{R}$ above. Similar expressions can be obtained when variances are not equal (see Smith [63]) or when the kernel's variance-covariance matrix varies spatially (see Huser and Genton [34]).

Example 6 (Kernels with compact support)**.** *As another example with compactly supported kernel, consider the truncated Gaussian kernel $k(s, s^*) = \frac{1}{A} \phi(s - s^*; \sigma^2) \mathbf{1}\{|s - s^*| < r\}$ where $\phi(\cdot; \sigma^2)$ is Gaussian density with mean zero and variance σ^2, and $A = 2\Phi(r/\sigma) - 1$ is the normalizing factor such that $\int_{\mathbb{R}} k(s, s^*) \mathrm{d}s^* = 1$. Again, for simplicity we assume $s_1 < s_2$. If $\delta = s_2 - s_1 \geq 2r$, then $\ell(w_1, w_2) = w_1 + w_2$. Otherwise,*

$$
\begin{aligned}
M(s_1, s_2) &= \{s^* \in \mathbb{R} : w_1 \zeta(s_1, s^*) > w_2 \zeta(s_2, s^*)\} \\
&= \{s^* \in \mathbb{R} : s_1 - r < s^* < s_2 - r\},
\end{aligned}
$$

for $\log(w_1/w_2) \geq -\Delta$, where $\Delta = \delta(r - \delta/2)/\sigma^2$, and

$$
M(s_1, s_2) = \left\{ s^* \in \mathbb{R} : s_1 - r < s^* < s_1 + \min\left(r, \frac{\sigma^2}{\delta} \log \frac{w_1}{w_2} + \frac{\delta}{2} \right) \right\},
$$

for $\log(w_1/w_2) \leq -\Delta$. This implies that

$$
\ell(w_1, w_2) = \begin{cases}
\frac{w_1}{A} \left\{ \Phi(\frac{\sigma}{\delta} \log \frac{w_1}{w_2} + \frac{\delta}{2\sigma}) - \Phi(-\frac{r}{\sigma}) \right\} \\
\qquad + \frac{w_2}{A} \left\{ \Phi(\frac{\sigma}{\delta} \log \frac{w_2}{w_1} + \frac{\delta}{2\sigma}) - \Phi(-\frac{r}{\sigma}) \right\}, & \left| \log \frac{w_1}{w_2} \right| \leq \Delta, \\
\frac{w_1}{A} \left\{ \Phi(\frac{\delta - r}{\sigma}) - \Phi(-\frac{r}{\sigma}) \right\} + w_2, & \log \frac{w_1}{w_2} < -\Delta, \\
w_1 + \frac{w_2}{A} \left\{ \Phi(\frac{\delta - r}{\sigma}) - \Phi(-\frac{r}{\sigma}) \right\}, & \log \frac{w_1}{w_2} > \Delta.
\end{cases}
$$

Appendix C: Approximate simulation of the max-stable process (5)

The next proposition can be used to easily and quickly generate approximate simulations from the limiting moving maximum max-stable process corresponding to the process $Z(\boldsymbol{s})$ (or $\tilde{Z}(\boldsymbol{s})$) with arbitrary accuracy.

Proposition 4 (Approximate simulation of Z_{EV} on a dense grid)**.** Let $Z_{\mathrm{EV}}(\boldsymbol{s})$ be the moving maximum max-stable process (5) with unit Fréchet margins $F^*(z) = \exp(-1/z)$, $z > 0$, which corresponds to extreme-value limit of the Cauchy convolution process $Z(\boldsymbol{s})$ defined in (2) with $W(\mathrm{d}\boldsymbol{s}^*) \sim_{\mathrm{i.i.d.}} \mathrm{Cauchy}(\mathrm{d}\boldsymbol{s}^*)$. Moreover, let $\{\boldsymbol{s}_{k,l}\}_{k,l=1}^{m}$ be the rectangular grid $S_m \subset \mathbb{R}^2$ as defined in the proof of Proposition 1 with the cell size $\delta_m \times \delta_m$ with $\delta_m \to 0$ such that $m\delta_m \to \infty$, as $m \to \infty$. Following the notation of Proposition 1, define

$$Z_{\mathrm{EV};j,m} = \delta_m^2 \max_{k,l=1,\ldots,m} \left\{ \zeta(\boldsymbol{s}_j, \boldsymbol{s}_{k,l}) W_{k,l}^* \right\}, \quad W_{k,l}^* \sim_{\mathrm{i.i.d}} F^*, \quad j = 1,\ldots,d.$$

It follows that $(Z_{\mathrm{EV};1,m},\ldots,Z_{\mathrm{EV};d,m})^\top \to_d \{Z_{\mathrm{EV}}(\boldsymbol{s}_1),\ldots,Z_{\mathrm{EV}}(\boldsymbol{s}_d)\}^\top$, as $m \to \infty$.

Proof. Define $Z_{j,m} = \delta_m^2 \sum_{k,l=1}^{m} \zeta(\boldsymbol{s}_j, \boldsymbol{s}_{k,l}) W_{k,l}$, $W_{k,l} = W(\boldsymbol{s}_{k,l}) \sim_{\mathrm{i.i.d.}} \mathrm{Cauchy}(\pi)$. Without loss of generality we use i.i.d. Cauchy random variables with the scale parameter π here so that $\Pr(W_{k,l} \le z) = 1 - 1/z + o(1/z)$ and $\Pr(W_{k,l}^* \le z) = 1 - 1/z + o(1/z)$ as $z \to \infty$.

Let $\tilde{q}_{m,j,h}$ and $q_{m,j,h}$ be the $(1 - h/n)$-quantiles of $Z_{\mathrm{EV};j,m}$ and $Z_{j,m}$, respectively. We have $\tilde{q}_{m,j,h} - q_{m,j,h} = o(n)$ (see Chapter VIII.8 in Feller [20]), and therefore, with $\tilde{p}_{m,d} = \Pr\{Z_{\mathrm{EV};1,\mathrm{m}} \le \tilde{q}_{m,1,h_1},\ldots,Z_{\mathrm{EV};d,m} \le \tilde{q}_{m,d,h_d}\}$,

$$
\begin{aligned}
\tilde{p}_{m,d} &= \prod_{k,l=1}^{m} \Pr\left\{ W_{k,l}^* \le \min_{j=1,\ldots,d} \frac{m^2 \tilde{q}_{m,j,h_j}}{\zeta(\boldsymbol{s}_j, \boldsymbol{s}_{k,l})} \right\} \\
&= \prod_{k,l=1}^{m} \Pr\left\{ W_{k,l}^* \le \min_{j=1,\ldots,d} \frac{m^2 q_{m,j,h_j}}{\zeta(\boldsymbol{s}_j, \boldsymbol{s}_{k,l})} \right\} + o\left(n^{-1}\right).
\end{aligned}
$$

Let $p_{m,d} = \Pr\{Z_{1,m} \le q_{m,1,h_1},\ldots,Z_{d,m} \le q_{m,d,h_d}\}$. Using Lemma 1,

$$
\begin{aligned}
p_{m,d} &= \prod_{k,l=1}^{m} \Pr\left\{ W_{k,l} \le \min_{j=1,\ldots,d} \frac{m^2 q_{m,j,h_j}}{\zeta(\boldsymbol{s}_j, \boldsymbol{s}_{k,l})} \right\} + o\left(n^{-1}\right) \\
&= \prod_{k,l=1}^{m} \Pr\left\{ W_{k,l}^* \le \min_{j=1,\ldots,d} \frac{m^2 q_{m,j,h_j}}{\zeta(\boldsymbol{s}_j, \boldsymbol{s}_{k,l})} \right\} + o\left(n^{-1}\right).
\end{aligned}
$$

This implies that the limiting extreme-value distributions of $\mathbf{Z}_m = (Z_{1,m},\ldots,Z_{d,m})^\top$ and $\tilde{\mathbf{Z}}_m = (Z_{\mathrm{EV};1,m},\ldots,Z_{\mathrm{EV};d,m})^\top$ are the same. By Proposition 1, we obtain that $(Z_{\mathrm{EV};1,m},\ldots,Z_{\mathrm{EV};d,m})^\top \to_d \{Z_{\mathrm{EV}}(\boldsymbol{s}_1),\ldots,Z_{\mathrm{EV}}(\boldsymbol{s}_d)\}^\top$ as $m \to \infty$. □

References

[1] Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics 44, 182–198. MR2517884

[2] Angus, T. E., 2012. General Theory of Functions and Integration. Dover Publications, Incorporated. MR0184545

[3] Bárdossy, A., 2006. Copula-based geostatistical models for groundwater quality parameters. Water Resources Research 42.

[4] Bárdossy, A., Li, J., 2008. Geostatistical interpolation using copulas. Water Resources Research 44.

[5] Ben Rached, N., Kammoun, A., Alouini, M.S., Tempone, R., 2016. Unified importance sampling schemes for efficient simulation of outage capacity over generalized fading channels. IEEE Journal of Selected Topics in Signal Processing 10, 376–388.

[6] Bopp, G., Shaby, B.A., Huser, R., 2021. A hierarchical max-infinitely divisible spatial model for extreme precipitation. Journal of the American Statistical Association 116, 93–106. MR4227677

[7] Botev, Z., L'Ecuyer, P., 2017. Accurate computation of the right tail of the sum of dependent log-normal variates, in: 2017 Winter Simulation Conference (WSC), pp. 1880–1890.

[8] Bousset, L., Jumel, S., Garreta, V., Picault, H., Soubeyrand, S., 2015. Transmission of leptosphaeria maculans from a cropping season to the following one. Annals of Applied Biology 166(3), 530–543.

[9] Calder, C., Cressie, N. A., 2007. Some topics in convolution-based spatial modeling. Bulletin of the International Statistical Institute 62, 132–139.

[10] Castro-Camilo, D., Huser, R., 2020. Local likelihood estimation of complex tail dependence structures, applied to U.S. precipitation extremes. Journal of the American Statistical Association 115, 1037–1054. MR4143448

[11] Castruccio, S., Huser, R., Genton, M.G., 2016. High-order composite likelihood inference for max-stable distributions and processes. Journal of Computational and Graphical Statistics 25, 1212–1229. MR3572037

[12] Davison, A.C., Gholamrezaee, M.M., 2012. Geostatistics of extremes. Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences 468, 581–608. MR2874052

[13] Davison, A.C., Huser, R., 2015. Statistics of extremes. Annual Review of Statistics and its Application 2, 203–235.

[14] Davison, A.C., Huser, R., Thibaud, E., 2019. Spatial extremes, in: Gelfand, A.E., Fuentes, M., Hoeting, J.A., Smith, R.L. (Eds.), Handbook of Environmental and Ecological Statistics. CRC Press, pp. 711–744.

[15] Davison, A.C., Padoan, S., Ribatet, M., 2012. Statistical modelling of spatial extremes (with Discussion). Statistical Science 27, 161–186. MR2963980

[16] Engelke, S., Hitz, A.S., 2020. Graphical models for extremes (with Discussion). Journal of the Royal Statistical Society, Series B 82, 871–932. MR4136498

[17] Engelke, S., Malinowski, A., Kabluchko, Z., Schlather, M., 2015. Estimation

of Huesler–Reiss distributions and Brown–Resnick processes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77, 239–265. MR3299407

[18] Erhardt, T.M., Czado, C., Schepsmeier, U., 2015. R-vine models for spatial time series with an application to daily mean temperature. Biometrics 71(2), 323–332. MR3366237

[19] Fasen, V., 2005. Extremes of regularly varying Lévy-driven mixed moving average processes. Advances in Applied Probability 37(4), 993–1014. MR2193993

[20] Feller, W., 1970. An Introduction to Probability Theory and Its Applications. volume Volume 2. John Wiley & Sons, USA.

[21] de Fondeville, R., Davison, A.C., 2018. High-dimensional peaks-over-threshold inference. Biometrika 105, 575–592. MR3842886

[22] Genest, C., Segers, J., 2009. Rank-based inference for bivariate extreme-value copulas. Annals of Statistics 37, 2990–3022. MR2541453

[23] Gneiting, T., 2002. Nonseparable, stationary covariance functions for space-time data. Journal of the American Statistical Association 97, 590–600. MR1941475

[24] Gneiting, T., Genton, M. G., Guttorp, P., 2007. Geostatistical space-time models, stationarity, separability and full symmetry. In Finkenstaedt, B., Held, L. and Isham, V.(eds), *Statistics of Spatio-Temporal Systems*, Chapman & Hall / CRC Press, Monograph in Statistics and Applied Probability, Boca Raton.

[25] Gräler, B., Pebesma, E., 2011. The pair-copula construction for spatial data: a new approach to model spatial dependency. Procedia Environmental Sciences 7, 206–211.

[26] Gudendorf, G., Segers, J., 2010. Extreme-value copulas, in: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (Eds.), Copula Theory and Its Applications, Proceedings of the Workshop Held in Warsaw, 25–26 September 2009, pp. 127–145. Lecture Notes in Statistics — Proceedings. MR3075361

[27] de Haan, L., 1984. A spectral representation for max-stable processes. Annals of Probability 12, 1194–1204. MR0757776

[28] Hazra, A., Huser, R., 2021. Estimating high-resolution Red sea surface temperature hotspots, using a low-rank semiparametric spatial model. Annals of Applied Statistics 15, 572–596. MR4298957

[29] Higdon, D., 2002. Space and space-time modeling using process convolutions. In: Anderson, C. W., El-Shaarawi, A. H., Chatwin, P. C., Barnett, V. (eds) Quantitative Methods for Current Environmental Issues. Springer, London. MR2059819

[30] Hua, L., 2017. On a bivariate copula with both upper and lower full-range tail dependence. Insurance: Mathematics and Economics 73, 94–104. MR3625749

[31] Huser, R., 2013. Statistical Modeling and Inference for Spatio-Temporal Extremes. Ph.D. thesis. École Polytechnique Fédérale de Lausanne.

[32] Huser, R., Davison, A.C., 2014. Space-time modelling of extreme events. Journal of the Royal Statistical Society: Series B (Statistical Methodology)

76, 439–461. MR3164873

[33] Huser, R., Dombry, C., Ribatet, M., Genton, M.G., 2019. Full likelihood inference for max-stable data. Stat 8, e218. MR3938281

[34] Huser, R., Genton, M.G., 2016. Non-stationary dependence structures for spatial extremes. Journal of Agricultural, Biological and Environmental Statistics 21, 470–491. MR3542082

[35] Huser, R., Opitz, T., Thibaud, E., 2017. Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. Spatial Statistics. 21, 166–186. MR3692183

[36] Huser, R., Opitz, T., Thibaud, E., 2021. Max-infinitely divisible models and inference for spatial extremes. Scandinavian Journal of Statistics 48, 321–348. MR4233175

[37] Huser, R., Wadsworth, J.L., 2019. Modeling spatial processes with unknown extremal dependence class. Journal of the American Statistical Association 114, 434–444. MR3941266

[38] Huser, R., Wadsworth, J.L., 2022. Advances in statistical modeling of spatial extremes. Wiley Interdisciplinary Reviews (WIREs): Computational Statistics 14, e1537. MR4376783

[39] Hüsler, J., Reiss, R. D., 1989. Maxima of normal random vectors: between independence and complete dependence. Statistics and Probability Letters 7, 283–286. MR0980699

[40] Jónsdóttir, K.Y., Rønn-Nielsen, A., Mouridsen, K., Jensen, E.B.V., 2013. Lévy-based modelling in brain imaging. Scandinavian Journal of Statistics 40(3), 511–529. MR3091695

[41] Kabluchko, Z., Schlather, M., de Haan, L., 2009. Stationary max-stable fields associated to negative definite functions. Annals of Probability 37, 2042–2065. MR2561440

[42] Krupskii, P., Genton, M. G., 2018. Linear factor copula models and their properties. Scandinavian Journal of Statistics 45(4), 861–878. MR3884891

[43] Krupskii, P., Genton, M. G., 2019. A copula model for non-Gaussian multivariate spatial data. Journal of Multivariate Analysis 169, 264–277. MR3875599

[44] Krupskii, P., Huser, R., Genton, M. G., 2018. Factor copula models for replicated spatial data. Journal of the American Statistical Association 521, 467–479. MR3803479

[45] Krupskii, P., Joe, H., 2013. Factor copula models for multivariate data. Journal of Multivariate Analysis 120, 85–101. MR3072719

[46] Kurowicka, D., Joe, H., 2011. Dependence Modeling: Vine Copula Handbook. World Scientific, Singapore. MR2849701

[47] Ledford, A.W., Tawn, J.A., 1996. Statistics for near independence in multivariate extreme values. Biometrika 83, 169–187. MR1399163

[48] Lee, D., Joe, H., Krupskii, P., 2018. Tail-weighted dependence measures with limit being tail dependence coefficient. Journal of Nonparametric Statistics, 30(2), 262–290. MR3794394

[49] Lindsay, B., 1998. Composite likelihood methods. Contemporary Mathematics 80, 220–239. MR0999014

[50] Marshall, W. A., Olkin, I., 1967. A multivariate exponential distribution. Journal of the American Statistical Association 62, 30–44. MR0215400

[51] Millar, P. W., 1984. A general approach to the optimality of minimum distance estimators. Transactions of the American Mathematical Society 286, 377–418. MR0756045

[52] Noven, R.C., Veraart, A.E.D., Gandy, A., 2018. A latent trawl process model for extreme values. Journal of Energy Markets 11(3), 1–24.

[53] Oesting, M., Schlather, M., Friederichs, P., 2017. Statistical post-processing of forecasts for extremes using bivariate Brown-Resnick processes with an application to wind gusts. Extremes 20, 309–332. MR3638369

[54] Opitz, T., 2017. Spatial random field models based on Lévy indicator convolutions. arXiv preprint arXiv:1710.06826.

[55] Paciorek, C. J., Schervish, M. J., 2006. Spatial modelling using a new class of nonstationarycovariance functions. Environmetrics 17, 483–506. MR2240939

[56] Padoan, S.A., Ribatet, M., Sisson, S.A., 2010. Likelihood-based inference for max-stable processes. Journal of the American Statistical Association 105, 263–277. MR2757202

[57] Rootzén, H., 1978. Extremes of moving average of stable processes. The Annals of Probability 6(5), 847–869. MR0494450

[58] Rootzén, H., Segers, J., Wadsworth, J.L., 2018. Multivariate peaks over thresholds models. Extremes 21, 115–145. MR3764613

[59] Sato, K., 1999. Lévy processes and infinitely divisible distributions. Cambridge University Press, UK. MR1739520

[60] Schlather, M., 2002. Models for stationary max-stable random fields. Extremes 5, 33–44. MR1947786

[61] Segers, J., 2012. Max-stable models for multivariate extremes. REVSTAT 10, 61–82. MR2912371

[62] Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Institute of Statistics of the University of Paris 8, 229–231. MR0125600

[63] Smith, R., 1990. Max-stable processes and spatial extremes. Department of Mathematics, University of Surrey.

[64] Strokorb, K., Ballani, F., Schlather, M., 2015. Tail correlation functions of max-stable processes. Extremes 18, 241–271. MR3351816

[65] Su, J., Hua, L., 2017. A general approach to full-range tail dependence copulas. Insurance: Mathematics and Economics 77, 49–64. MR3724935

[66] Thibaud, E., Mutzner, R., Davison, A.C., 2013. Threshold modeling of extreme spatial rainfall. Water Resources Research 49, 4633–4644.

[67] Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Statistica Sinica 21, 5–42. MR2796852

[68] Vettori, S., Huser, R., Genton, M.G., 2019. Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. Biometrics 75, 831–841. MR4012089

[69] Vettori, S., Huser, R., Segers, J., Genton, M.G., 2020. Bayesian model averaging over tree-based dependence structures for multivariate extremes. Journal of Computational and Graphical Statistics 29, 174–190.

MR4085873

[70] Wadsworth, J.L., Tawn, J.A., 2012. Dependence modelling for spatial extremes. Biometrika 99, 253–272. MR2931252

[71] Wadsworth, J.L., Tawn, J.A., 2022. Higher-dimensional spatial extremes via single-site conditioning. Spatial Statistics 51, 100677.

[72] Wadsworth, J.L., Tawn, J.A., Davison, A.C., Elton, D.M., 2016. Modelling across extremal dependence classes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79, 149–175. MR3597968

[73] Zhang, J., 2010. A highly efficient L-estimator for the location parameter of the Cauchy distribution. Computational Statistics 25(1), 97–105. MR2586726

[74] Zhu, Z., Wu, Y., 2010. Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data. Journal of Computational and Graphical Statistics 19(1), 74–95. MR2654401