

Modeling Spatial Data with Cauchy Convolution Processes

Pavel Krupskii¹, Raphaël Huser²

February 16, 2021

Abstract

We study the class of models for spatial data obtained from Cauchy convolution processes based on different types of kernel functions. We show that the resulting spatial processes have some 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 consider some interesting special cases, including Marshall–Olkin and Hüsler–Reiss processes. We further consider mixtures between such Cauchy processes and Gaussian processes, in order to have a separate control over the bulk and the tail dependence behaviors. Our proposed approach for estimating model parameters relies on matching model-based and empirical summary statistics, while the corresponding extreme-value limit models may be fitted using a pairwise likelihood approach. We show with a simulation study that our proposed inference approach yields accurate estimates. Moreover, the proposed class of models allows for a wide range of flexible dependence structures, and we demonstrate our new methodology by application to a temperature dataset. 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.

Keywords: Copula; Extreme-value model; Kernel convolution process; Short-range spatial dependence; Spatial process; Tail dependence

¹University of Melbourne, Parkville, Victoria, 3010, Australia. E-mail: pavel.krupskiy@unimelb.edu.au.

²Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia.. E-mail: raphael.huser@kaust.edu.sa.

1 Introduction

Assessment of environmental risk associated with unprecedented air or sea temperatures (Davison and Gholamrezaee, 2012; Huser and Genton, 2016; Hazra and Huser, 2020), extreme flooding (Thibaud et al., 2013; Huser and Davison, 2014; Castro-Camilo and Huser, 2020), strong wind gusts (Engelke et al., 2015; Oesting et al., 2017; Huser et al., 2017), or high air pollution levels (Vettori et al., 2019, 2020) 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, 2002, and Gneiting et al., 2007). 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 (Gräler and Pebesma, 2011), hydrology (Bárdossy, 2006; Bárdossy and Li, 2008) and climatology (Erhardt et al., 2015), among others. A copula is simply defined as a multivariate cumulative distribution function (CDF) with standard uniform $\text{Unif}(0, 1)$ marginals. Sklar (1959) showed that for any continuous d -dimensional CDF F with marginals F_1, \dots, F_d there exists a unique copula C such that $F(z_1, \dots, z_d) = C\{F_1(z_1), \dots, F_d(z_d)\}$. 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

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

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 $\lambda_L = \lim_{u \rightarrow 0} \Pr\{Z_1 \leq F_1^{-1}(u) \mid Z_2 \leq F_2^{-1}(u)\} = \lim_{u \rightarrow 0} C(u, u)/u$. 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 (Ledford and Tawn, 1996), 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 (Aas et al., 2009; Kurowicka and Joe, 2011), are very flexible but lack interpretability with spatial data. By contrast, factor copula models (Krupskii and Joe, 2013) 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. (2018) and Krupskii and Genton (2019). 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 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 (2015) for a general review on statistics of extremes, and Davison et al. (2012), Davison et al. (2019) and Huser and Wadsworth (2020) 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 across all quantiles $u \in [0, 1]$ (Rootzén et al., 2018), while with extreme-value copulas, one has

$C^k(u_1^{1/k}, u_2^{1/k}) = C(u_1, u_2)$ for all $(u_1, u_2)^\top \in [0, 1]^2$, $k = 1, 2, \dots$. 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 (Huser and Wadsworth, 2019), 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. Thirdly, these models have complicated likelihood functions that are costly to evaluate for inference (Padoan et al., 2010; Castruccio et al., 2016; de Fondeville and Davison, 2018; Huser et al., 2019), though recent progress on graphical models for Pareto processes opens the door to higher-dimensional likelihood inference (Engelke and Hitz, 2020). Finally, because these models describe the limiting behavior of extreme events, they are typically fitted using only a fraction of observations, thus wasting a lot of information that is potentially useful for accurate estimation of unknown model parameters.

To circumvent some of the limitations of classical 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, e.g., Wadsworth and Tawn (2012), Wadsworth et al. (2016), Hua (2017), Su and Hua (2017), Huser et al. (2017), Huser et al. (2019), Bopp et al. (2020), Huser et al. (2020), and the recent review paper Huser and Wadsworth (2020). More recently, an alternative approach based on single-site conditioning has been proposed by Wadsworth and Tawn (2019) to flexibly capture various forms of tail dependence structures, although the proposed model does not possess a convenient unconditional formulation. However, except for the max-mixture model of Wadsworth and

Tawn (2012) and the conditional extremes model of Wadsworth and Tawn (2019), 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

$$Z(\mathbf{s}) = \int_{\mathbb{R}^q} k(\mathbf{s}, \mathbf{s}^*) W(d\mathbf{s}^*), \quad \mathbf{s} \in \mathbb{R}^q, \quad (2)$$

where $k(\mathbf{s}, \mathbf{s}^*) \geq 0$ is a nonnegative integrable kernel function (i.e., such that $\int_{\mathbb{R}^q} k(\mathbf{s}, \mathbf{s}^*) d\mathbf{s}^* < \infty$ for all $\mathbf{s} \in \mathbb{R}^q$) and W is a particular Lévy process (Sato, 1999) with independent increments. 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 Higdon (2002), Paciorek and Schervish (2006), Calder and Cressie (2007) and Zhu and Wu (2010). However, the marginal CDF of $W(\mathbf{s})$, F_W , is usually assumed to be Gaussian, thus leading to a Gaussian process $Z(\mathbf{s})$ in (2), which does not have tail dependence. Trans-Gaussian processes $Z^*(\mathbf{s}) = t\{Z(\mathbf{s})\}$, where $t(\cdot)$ is a monotone increasing transformation, could be used to model spatial data with non-Gaussian marginals (Bousset et al., 2015), but the process $Z^*(\mathbf{s})$ 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) (Jónsdóttir et al., 2013; Noven et al., 2018), and Opitz (2017) studied the dependence properties of the resulting process for an indicator kernel $k(\mathbf{s}, \mathbf{s}^*) = \mathbf{1}\{\mathbf{s}^* \in \mathcal{A}(\mathbf{s})\}$ defined in terms of a hypograph indicator set $\mathcal{A}(\mathbf{s})$. 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 $Z(\mathbf{s})$ 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 extremal dependence structures can be obtained, recovering some popular existing models and offering some new models, as well. In particular, when the kernel function is compactly supported, the resulting process $Z(\mathbf{s})$ 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. To make our Cauchy model even more flexible in the bulk of the distribution, we further extend it by mixing the Cauchy process with a lighter-tailed Gaussian process, so that we retain the same extremal dependence structure while capturing (non-trivial) tail independence at large distances. Furthermore, we provide some theoretical results on the spatial dependence strength of the resulting extreme-value copulas at short distances, providing new insights into their “smoothness” properties.

To make inference for the Cauchy process convolution model (2) or its more flexible 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 (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 be as crucial in practice for risk assessment as severe individual extremes; 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 overall fit, without compromising any part of the distribution. For completeness, we also discuss how the corresponding extreme-value copula may be fitted by pairwise likelihood.

The rest of the paper is organized as follows. In Section 2, we define our proposed Cauchy convolution model in more detail, study its dependence properties and tail behavior, we derive some interesting special cases, and we discuss approximate simulation algorithms for the Cauchy convolution process itself and its extreme-value limits. We also study our proposed process mixture, 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 the Supplementary Material.

2 Modeling

2.1 Cauchy convolution processes and their extreme-value limits

We consider the process convolution (2), where W is a Cauchy white noise process, i.e., such that $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$ are independent and identically distributed (i.i.d.) increments, where $\text{Cauchy}(c)$ is the Cauchy distribution with scale parameter c and probability density function (pdf) $f_C(z; c) = \pi^{-1}c/(z^2 + c^2)$, $z \in \mathbb{R}$. The finite-dimensional distributions of the Cauchy process convolution $Z(\mathbf{s})$ 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 $\mathbf{X} = (X_1, \dots, X_d)^\top$ be a d -dimensional random vector with margins F_1, \dots, F_d and copula C . The copula C^n describes

the copula of the vector of componentwise maxima from i.i.d. copies $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^\top$ of \mathbf{X} , $i = 1, \dots, n$, i.e., $\mathbf{M}_n = (M_{n1}, \dots, M_{nd})^\top$ with $M_{nj} = \max(X_{1j}, \dots, X_{nj})$. Extreme-value copulas, denoted C_{EV} , describe the class of dependence structures that arise as limits of \mathbf{M}_n (when properly renormalized), i.e.,

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

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

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

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 $\ell(w_1, \dots, w_d) = \lim_{n \rightarrow \infty} n\{1 - C(1 - w_1/n, \dots, 1 - w_d/n)\}$, and it lies between the bounds $\max(w_1, \dots, w_d) \leq \ell(w_1, \dots, w_d) \leq \sum_{j=1}^d w_j$, $w_1, \dots, w_d \geq 0$, 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 Gudendorf and Segers (2010), Segers (2012), and Davison and Huser (2015).

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

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

The result of Proposition 1 shows that max-stable processes resulting from Cauchy convolution processes are from the class of moving maximum processes (see, e.g., de Haan, 1984, Schlather, 2002, Strokorb et al., 2015 and the references therein), whose stable tail dependence function is of form given above. Such extreme-value processes admit the stochastic representation

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

where $\{(\xi_i, \mathbf{s}_i^*)\}$ are points from a Poisson process on $(0, \infty) \times \mathbb{R}^2$ with intensity $\xi^{-2} d\xi \times d\mathbf{s}^*$. The limit process in (5) is max-stable and has unit Fréchet margins, i.e., $\Pr\{Z_{\text{EV}}(\mathbf{s}) \leq z\} = \exp(-1/z)$, $z > 0$. A prominent example is the model introduced by Smith (1990), 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 (2005) obtained a result similar to Proposition 1 for continuous-time mixed moving average processes. Similarly, Rootzén (1978) studied the tail properties of stable moving average processes and established continuity of sample paths of these processes.

The Smith (1990) model is known to have very smooth sample paths; see, e.g., Davison et al. (2019). 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 sense, and therefore that this process is not suitable for modeling spatial data with rough extremal 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 Suppose that the assumptions of Proposition 1 hold. Moreover, assume that the kernel function in (2) may be written as $k(\mathbf{s}, \mathbf{s}^*) = g(\|\mathbf{s} - \mathbf{s}^*\|)$, where g is an integrable nonnegative monotonically decreasing function. Then, for any sites $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^2$, the stable tail dependence function $\ell(w_1, w_2)$ of $\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\}^\top$ satisfies $\ell(w_1, w_2) \leq \{1 + K\delta + o(\delta)\} \max(w_1, w_2)$, as $\delta :=$

$\|\mathbf{s}_1 - \mathbf{s}_2\| \rightarrow 0$, 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 $Z_{\text{EV}}(\mathbf{s})$ be the extreme-value limit (5) of the Cauchy convolution process $Z(\mathbf{s})$ defined as in (2) with $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$, 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 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 $Z(\mathbf{s})$ and $Z_{\text{EV}}(\mathbf{s})$), while $S_\rho(\delta)$ mostly controls dependence in the bulk (and generally differs for $Z(\mathbf{s})$ and $Z_{\text{EV}}(\mathbf{s})$). 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 $\text{corr}\{F_1(Y_1), F_2(Y_2)\}$ and may be equivalently expressed in terms of the copula C as $12 \iint_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3$. 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 $Z_{\text{EV}}(\mathbf{s})$ have indeed a quite restrictive behavior at the origin, i.e., for small distances $\delta \approx 0$.

Corollary 1 Under the assumptions of Proposition 2, the limiting extreme-value process $Z_{\text{EV}}(\mathbf{s})$ of the Cauchy convolution process $Z(\mathbf{s})$ (defined as in (2) with $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$) has dependence coefficients satisfying $\lambda(\delta) = 1 - K\delta + o(\delta)$ and $S_\rho(\delta) > 1 - 4K\delta$, as $\delta \rightarrow 0$.

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 how to extend Cauchy convolution processes in order to

capture the sub-asymptotic dependence structure more flexibly using mixtures.

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 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 $w_1/w_2 > G(\delta) \geq 1$. Similarly, $\ell(w_1, w_2) = w_2$ for all $(w_1, w_2)^\top \in [0, \infty)^2$ with $w_2/w_1 > G(\delta) \geq 1$.

This result implies that the extreme-value copula $C_{EV}(u_1, u_2)$ 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, for all kernels of the form $k(\mathbf{s}, \mathbf{s}^*) = \exp\{-\|\mathbf{s} - \mathbf{s}^*\|/\lambda\}^\alpha$, 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 values at one location to occur with very high values at another location. By contrast, it is easy to verify that $G(\delta) = \infty$ for all $\alpha > 1$, and for all kernels $k(\mathbf{s}, \mathbf{s}^*) = g(\|\mathbf{s} - \mathbf{s}^*\|)$ that are compactly supported, i.e., such that $g(t) = 0$ for all $t > r$ for some range $r > 0$. In practice, it may be sensible to restrict ourselves to $\alpha > 1$ when using $k(\mathbf{s}, \mathbf{s}^*) = \exp\{-\|\mathbf{s} - \mathbf{s}^*\|/r\}^\alpha$, or to use a different (potentially compactly supported) kernel.

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

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

The corresponding limiting extreme-value process is thus driven by the Marshall–Olkin copula with a singular component (Marshall and Olkin, 1967).

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 $k(\mathbf{s}, \mathbf{s}^*) = g(\|\mathbf{s} - \mathbf{s}^*\|)$ where g is a nonnegative continuous function.

Example 2 (Smith (1990) model) 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 (Hüsler and Reiss, 1989). 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 (1990) max-stable model, which is a smooth limiting case of the Brown-Resnick model (Kablichko et al., 2009) where the variogram is $\gamma(h) = (|h|/\sigma)^\alpha$, $\sigma > 0, \alpha \in (0, 2)$, instead. More details for the case of $q = 2$ are provided in the Supplementary Material.

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 $\delta = |s_2 - s_1|$,

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

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 $u_2^{G_\lambda(\delta)} < u_1 < u_2^{1/G_\lambda(\delta)}$.

Example 4 (Kernels with finite support) Notice that if the set $\mathcal{S}_{\cap+}(\mathbf{s}_1, \mathbf{s}_2) := \{\mathbf{s}^* \in \mathbb{R}^q : k(\mathbf{s}_1, \mathbf{s}^*) > 0, k(\mathbf{s}_2, \mathbf{s}^*) > 0\} = \emptyset$ for some sites $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^q$, then $\ell(w_1, w_2) = w_1 + w_2$. In particular, if the kernel k is compactly supported such that $k(\mathbf{s}, \mathbf{s}^*) = 0$ whenever $\|\mathbf{s} - \mathbf{s}^*\| > r$ for some radius $r > 0$, then $\mathcal{S}_{\cap+}(\mathbf{s}_1, \mathbf{s}_2)$ is empty if and only if $\|\mathbf{s}_1 - \mathbf{s}_2\| \geq 2r$. This implies that the variables $Z_{\text{EV}}(\mathbf{s}_1)$ and $Z_{\text{EV}}(\mathbf{s}_2)$ from the limiting extreme-value process are independent whenever $\|\mathbf{s}_1 - \mathbf{s}_2\| \geq 2r$. By construction, two realizations $Z(\mathbf{s}_1)$ and $Z(\mathbf{s}_2)$ 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 $k(\mathbf{s}, \mathbf{s}^*) = \{1 - (\|\mathbf{s} - \mathbf{s}^*\|/r)^\alpha\}_+^\eta$, 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 the Supplementary Material.

2.3 Mixture of Cauchy and Gaussian processes

The Cauchy convolution process $Z(\mathbf{s})$ defined as in (2) with $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$ 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

$k(\mathbf{s}, \mathbf{s}^*)$. Moreover, 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. Unfortunately, Cauchy convolution processes have three main drawbacks when modeling spatial data:

1. As 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 $\mathbf{s}_1, \mathbf{s}_2$, two variables $Z(\mathbf{s}_1)$ and $Z(\mathbf{s}_2)$ 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 lacks flexibility at sub-asymptotic levels;
3. The strength of dependence in the bulk of the joint distribution of $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_d)\}^\top$ 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 limiting factor 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 accurately. 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 simultaneously $\lambda(\delta) \approx 0$ 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.

To increase flexibility of the Cauchy process convolution model and circumvent the issues highlighted in the second and third points above, we can modify the original process by mixing it with

a tail-independent process possessing lighter tails. We define

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

where $Z(\mathbf{s})$ is the Cauchy process (2) with $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$, $Z_G(\mathbf{s})$ 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 tail properties of the new process remain intact, such that the mixture model (6) still captures local tail dependence if the kernel is compactly supported.

Proposition 3 Under the assumptions of Proposition 1, the stable tail dependence function corresponding to the joint distribution of $\{\tilde{Z}(\mathbf{s}_1), \dots, \tilde{Z}(\mathbf{s}_d)\}^\top$, with the process $\tilde{Z}(\mathbf{s})$ defined as in (6), is $\ell(w_1, \dots, w_d)$ given in Proposition 1.

In the next section we will see that the mixture process $\tilde{Z}(\mathbf{s})$ in (6) has much rougher realizations than the Cauchy process $Z(\mathbf{s})$ alone, and so it is more realistic for modeling real environmental data. The improved flexibility of model (6) is further illustrated in the Supplementary Material.

2.4 Simulation

Approximate simulations from the Cauchy convolution process $Z(\mathbf{s})$ defined in (2) with $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$ can be obtained at locations $\mathbf{s}_1, \dots, \mathbf{s}_d \in \mathbb{R}^2$ as $Z(\mathbf{s}_j) \approx m^{-2} \sum_{k,l=1}^m k(\mathbf{s}_j, \mathbf{s}_{k,l}) W_{k,l}$, $W_{k,l} \sim_{\text{i.i.d.}} \text{Cauchy}(1)$, $j = 1, \dots, d$, where $\{\mathbf{s}_{k,l}\}_{k,l=1}^m$ is a fine rectangular grid in $S_m \subset \mathbb{R}^2$ covering the simulation sites, with large m (see the proof of Proposition 1). Similarly, the process $\tilde{Z}(\mathbf{s})$ in (6) can be simulated based on a finite approximation as $\tilde{Z}(\mathbf{s}_j) \approx m^{-2} \sum_{k,l=1}^m k(\mathbf{s}_j, \mathbf{s}_{k,l}) W_{k,l} + \beta Z_G(\mathbf{s}_j)$, $W_{k,l} \sim_{\text{i.i.d.}} \text{Cauchy}(1)$, $j = 1, \dots, d$, where $\{Z_G(\mathbf{s}_1), \dots, Z_G(\mathbf{s}_d)\}^\top$ is a realization from the Gaussian random field $Z_G(\mathbf{s})$ at the sites $\mathbf{s}_1, \dots, \mathbf{s}_d$, independent of the Cauchy variables $\{W_{k,l}\}$.

Simulations from the associated limiting extreme-value process $Z_{\text{EV}}(\mathbf{s})$ are also quite straightforward. Given that the kernel function $k(\mathbf{s}, \mathbf{s}^*)$ is bounded, simulating moving maximum max-stable random fields can be performed exactly by exploiting the stochastic representation (5), simulating the Poisson points $\{\xi_i\}$ in decreasing order and setting an appropriate stopping rule; see Schlather (2002). 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; we provide more details in the Supplementary Material.

Figure 1 shows realizations of the processes $Z(\mathbf{s})$, $\tilde{Z}(\mathbf{s})$ and $Z_{\text{EV}}(\mathbf{s})$ in $[0, 1]^2$ obtained for different kernel functions. For the indicator kernel with finite 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 $k(\mathbf{s}, \mathbf{s}^*) < 5 \cdot 10^{-5}$ for $\|\mathbf{s} - \mathbf{s}^*\| > 1$ for the considered kernels. We use $m = 200$ in all cases.

We can see that the process $\tilde{Z}(\mathbf{s})$ (third column) has a rougher spatial dependence than the Cauchy process $Z(\mathbf{s})$ (second column), confirming the higher degree of flexibility of $\tilde{Z}(\mathbf{s})$ to capture the sub-asymptotic dependence structure. Moreover, in all simulations, the effect of the kernel on the spatial dependence structure is obvious, especially—but not only—with the extreme-value process $Z_{\text{EV}}(\mathbf{s})$ (fourth column). Smoother kernels result in smoother random fields. Figure 2 shows bivariate scatter plots for datasets of size 1000 generated from the extreme-value copulas linking $Z_{\text{EV}}(\mathbf{s}_1)$ and $Z_{\text{EV}}(\mathbf{s}_2)$ at distance $\|\mathbf{s}_1 - \mathbf{s}_2\| = 0.125$ for the four extremal processes displayed in Figure 1. 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

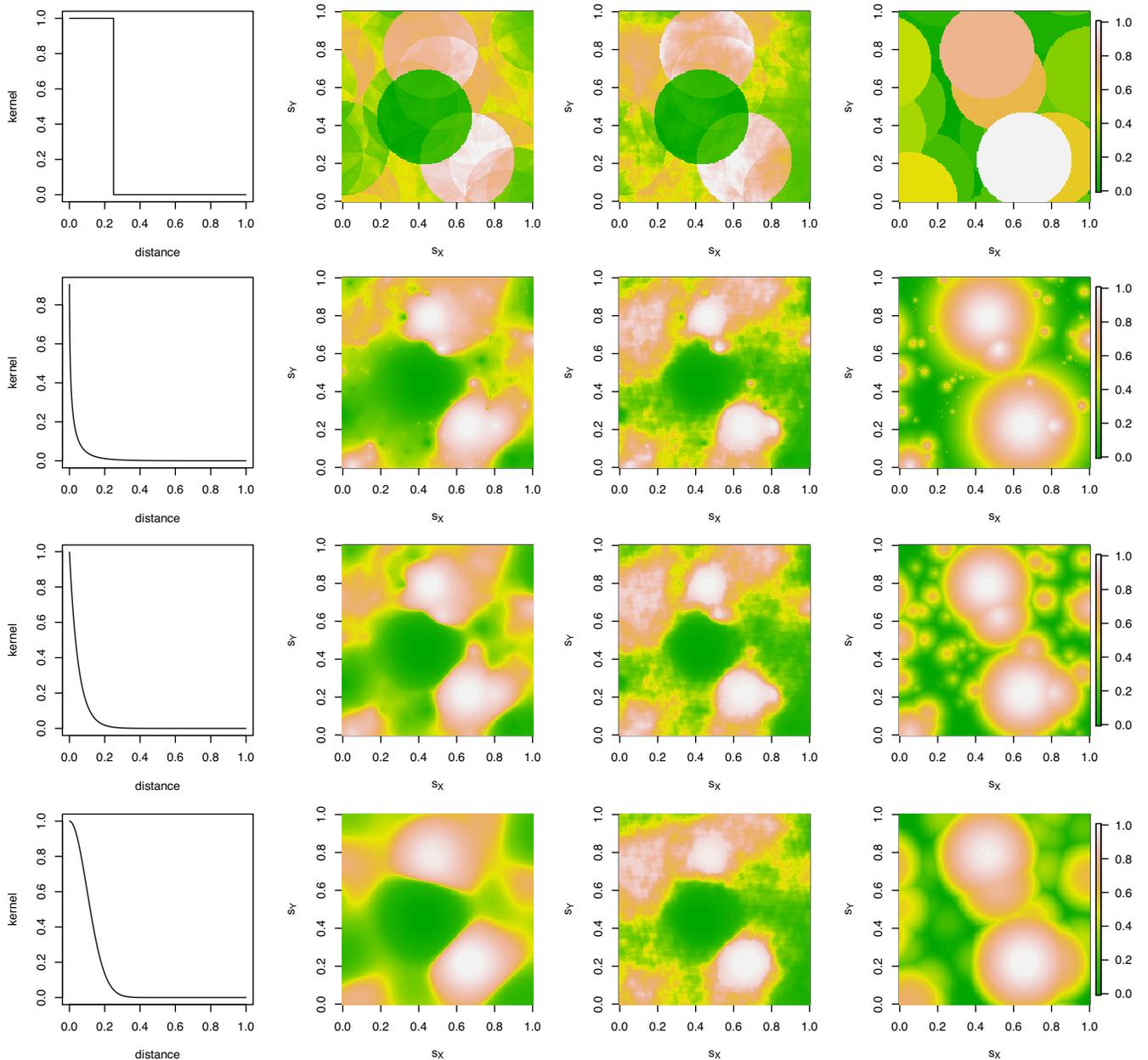


Figure 1: Realizations on $[0, 1]^2$ from the Cauchy convolution process $Z(\mathbf{s})$ defined in (2) with $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$ (2nd column), $\tilde{Z}(\mathbf{s})$ defined in (6) (3rd column), and their extreme-value limit $Z_{EV}(\mathbf{s})$ (4th column) for different isotropic kernels $k(\mathbf{s}, \mathbf{s}^*) = g(\|\mathbf{s} - \mathbf{s}^*\|)$ (the 1st column shows the function g). Each simulated process is marginally transformed to have $\text{Unif}(0, 1)$ marginals. The same random seed was used for all realizations. The kernel is $k(\mathbf{s}, \mathbf{s}^*) = \mathbf{1}(\|\mathbf{s} - \mathbf{s}^*\| < 0.25)$ (1st row), $k(\mathbf{s}, \mathbf{s}^*) = \exp(-10\|\mathbf{s} - \mathbf{s}^*\|^{1/2})$ (2nd row), $k(\mathbf{s}, \mathbf{s}^*) = \exp(-20\|\mathbf{s} - \mathbf{s}^*\|)$ (3rd row), and $k(\mathbf{s}, \mathbf{s}^*) = \exp(-50\|\mathbf{s} - \mathbf{s}^*\|^2)$ (4th row). We use $\beta = 2$ and $\rho_G(\delta) = \exp(-\delta)$ for $\tilde{Z}(\mathbf{s})$.

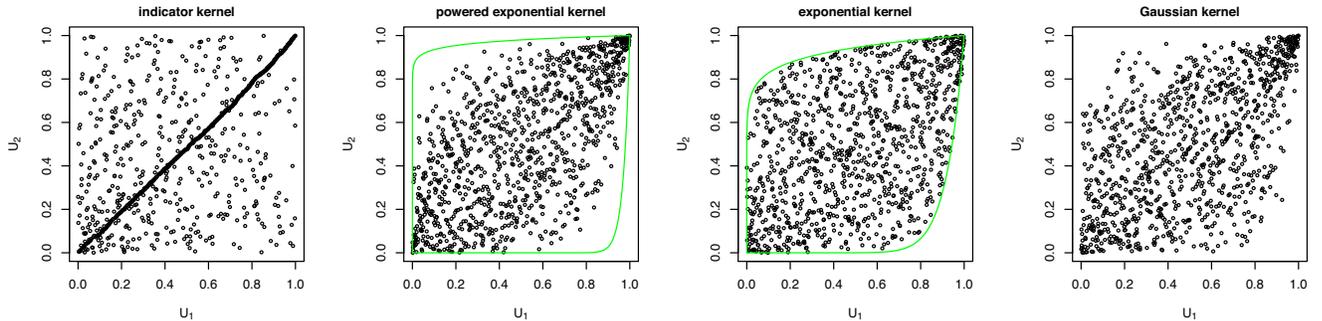


Figure 2: Bivariate scatter plots of datasets of size 1000 generated from the extreme-value copulas linking $Z_{\text{EV}}(\mathbf{s}_1)$ and $Z_{\text{EV}}(\mathbf{s}_2)$ at distance $\delta = \|\mathbf{s}_1 - \mathbf{s}_2\| = 0.125$ for the extreme-value process $Z_{\text{EV}}(\mathbf{s})$ with kernel $k(\mathbf{s}, \mathbf{s}^*) = \mathbf{1}(\|\mathbf{s} - \mathbf{s}^*\| < 0.25)$, $k(\mathbf{s}, \mathbf{s}^*) = \exp(-10\|\mathbf{s} - \mathbf{s}^*\|^{1/2})$, $k(\mathbf{s}, \mathbf{s}^*) = \exp(-20\|\mathbf{s} - \mathbf{s}^*\|)$, and $k(\mathbf{s}, \mathbf{s}^*) = \exp(-50\|\mathbf{s} - \mathbf{s}^*\|^2)$ (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 second and third panels, respectively.

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 how inference may be performed for the Cauchy convolution process $Z(\mathbf{s})$ observed at d locations $\mathbf{s}_1, \dots, \mathbf{s}_d$, for the modified mixture process $\tilde{Z}(\mathbf{s})$, and for their corresponding extreme-value limit, $Z_{\text{EV}}(\mathbf{s})$.

3 Inference

3.1 General setting and marginal estimation

Throughout this section, we assume that $\{\mathbf{y}_i = (y_{i1}, \dots, y_{id})^\top\}_{i=1}^n$ are n i.i.d. realizations of a stationary process $Y(\mathbf{s})$ measured at d spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_d \in \mathbb{R}^2$, and whose dependence structure (i.e., copula) is the same as either (i) the Cauchy convolution process $Z(\mathbf{s})$ defined as in (2) with $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$ using some parametric kernel function k (Section 3.2); or (ii) the extended mixture model $\tilde{Z}(\mathbf{s})$ defined as in (6) (Section 3.3), 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 $Y(\mathbf{s})$ and that $\boldsymbol{\theta}_F$ is the vector of marginal

parameters. In other words, the process $Y(\mathbf{s})$ has the same dependence properties (both in the bulk and the tails) as $Z(\mathbf{s})$ or $\tilde{Z}(\mathbf{s})$ or $Z_{\text{EV}}(\mathbf{s})$, 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 $\hat{\boldsymbol{\theta}}_F$. 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 (Varin et al., 2011). The data can then be transformed to the uniform $\text{Unif}(0, 1)$ scale using the probability integral transform. More precisely, pseudo-uniform scores $\{\mathbf{u}_i = (u_{i1}, \dots, u_{id})^\top\}_{i=1}^n$ can be obtained by setting $u_{ij} = F(y_{ij}; \hat{\boldsymbol{\theta}}_F)$.

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. We focus here on the models $Z(\mathbf{s})$ in (2) and $\tilde{Z}(\mathbf{s})$ in (6), while composite likelihood inference for $Z_{\text{EV}}(\mathbf{s})$ is discussed in the Supplementary Material.

3.2 Parameter estimation for the Cauchy convolution process model $Z(\mathbf{s})$ defined as 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 $Z(\mathbf{s})$ in (2) with $W(d\mathbf{s}^*) \sim_{\text{i.i.d.}} \text{Cauchy}(d\mathbf{s}^*)$ using some parametric kernel function $k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K)$, where $\boldsymbol{\theta}_K$ is the vector of dependence parameters controlling the kernel function. While the joint likelihood function for the process $Z(\mathbf{s})$ 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 $\{\mathbf{z}_i^* = (z_{i1}^*, \dots, z_{id}^*)^\top\}_{i=1}^n$ with $z_{ij}^* = F_c^{-1}(u_{ij})$, where $F_c^{-1}(q) = \tan\{\pi(q - 0.5)\}$ is the quantile function of the standard Cauchy distribution. By stability of the Cauchy distribution, the marginal distribution of the process convolution $Z(\mathbf{s})$ in (2) are Cauchy with scale parameter $c(\mathbf{s}; \boldsymbol{\theta}_K) = \int_{\mathbb{R}^2} k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) d\mathbf{s}^*$. This implies that $Z^*(\mathbf{s}) := Z(\mathbf{s})/c(\mathbf{s}; \boldsymbol{\theta}_K) \sim \text{Cauchy}(1)$ and that $\{\mathbf{z}_i^*\}_{i=1}^n$ can be treated as pseudo-observations from $Z^*(\mathbf{s})$. Moreover, we have

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

where

$$c_{jk}(\boldsymbol{\theta}_K) = \int_{\mathbb{R}^2} |\zeta(\mathbf{s}_j, \mathbf{s}^*) - \zeta(\mathbf{s}_k, \mathbf{s}^*)| d\mathbf{s}^*, \quad \zeta(\mathbf{s}, \mathbf{s}^*) \equiv \zeta(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) = \frac{k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K)}{\int_{\mathbb{R}^2} k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) d\mathbf{s}^*}. \quad (7)$$

For each pair of sites $\{\mathbf{s}_j, \mathbf{s}_k\}$, 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 \hat{c}_{jk} satisfies the equation $\sum_{i=1}^n \hat{c}_{jk}^2 / (\hat{c}_{jk}^2 + y_{ijk}^{*2}) = n/2$, whose positive root can be easily found using numerical routines. The estimator \hat{c}_{jk} is a consistent and asymptotically normal estimator of c_{jk} . Alternatively, the median of absolute values, $\text{median}\{|y_{1jk}|, \dots, |y_{njk}|\}$, 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). To estimate the vector of parameters $\boldsymbol{\theta}_K$, we can then use a least squares approach and compute

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

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., $\omega_{jk} = \mathbf{1}(\|\mathbf{s}_j - \mathbf{s}_k\| \leq \delta_{\max})$ 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 $\boldsymbol{\theta}_K$ (Millar, 1984).

3.3 Parameter estimation for the extended model $\tilde{Z}(\mathbf{s})$ defined as 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 $\tilde{Z}(\mathbf{s})$ in (6). In addition to the parametric kernel function $k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K)$ 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 $Z_{\text{EV}}(\mathbf{s})$ 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 $\{\tilde{Z}(\mathbf{s}_j), \tilde{Z}(\mathbf{s}_k)\}$. 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 $\{\mathbf{s}_j, \mathbf{s}_k\}$ as

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

where $u \approx 1$ is a high threshold on the uniform scale. The empirical estimator $\hat{\lambda}_{jk}$ is consistent as $n \rightarrow \infty$ and $u \equiv u_n \rightarrow 1$ such that $n(1-u_n) \rightarrow \infty$. Many other valid non-parametric estimators of the tail dependence coefficient exist. In particular, as the copula stemming from $\tilde{Z}(\mathbf{s})$ is reflection-symmetric, it would be possible to combine information from the lower and upper tails to estimate $\lambda_{jk}(\boldsymbol{\theta}_K)$ but here we restrict ourselves to the empirical estimator (9), which is quite standard and better reflects the upper tail. On the other hand, from Propositions 1 and 3, we have

$$\begin{aligned} \lambda_{jk}(\boldsymbol{\theta}_K) &= 2 - \ell(1, 1) = 2 - \int_{\mathbb{R}^2} \max\{\zeta(\mathbf{s}_j, \mathbf{s}^*), \zeta(\mathbf{s}_k, \mathbf{s}^*)\} d\mathbf{s}^* \\ &= 1 - \int_{\zeta(\mathbf{s}_j, \mathbf{s}^*) > \zeta(\mathbf{s}_k, \mathbf{s}^*)} \zeta(\mathbf{s}_j, \mathbf{s}^*) d\mathbf{s}^* + 1 - \int_{\zeta(\mathbf{s}_k, \mathbf{s}^*) > \zeta(\mathbf{s}_j, \mathbf{s}^*)} \zeta(\mathbf{s}_k, \mathbf{s}^*) d\mathbf{s}^* \end{aligned}$$

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

where $\ell(w_1, w_2)$ is the stable tail dependence function of $\{\tilde{Z}(\mathbf{s}_j), \tilde{Z}(\mathbf{s}_k)\}^\top$, and $\zeta(\mathbf{s}, \mathbf{s}^*) \equiv \zeta(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) = k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) / \int_{\mathbb{R}^2} k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) d\mathbf{s}^*$. 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:

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

where $\hat{\lambda}_{jk}$ 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 $Z(\mathbf{s})$, 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

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

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 $\{\tilde{\mathbf{z}}_i = (\tilde{z}_{i1}, \dots, \tilde{z}_{id})^\top\}_{i=1}^n$, where $\tilde{z}_{ij} = F^{-1}(u_{ij}; 1, \beta_0)$. If β_0 is the true value of β , then the vectors $\{\tilde{\mathbf{z}}_i\}_{i=1}^n$ can be considered as pseudo-observations from the process $\tilde{Z}(\mathbf{s})$ in (6). Now, notice that for any two sites $\{\mathbf{s}_j, \mathbf{s}_k\}$, we have

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

where $c_{jk}(\boldsymbol{\theta}_K)$ is given in (7) and $\beta_{jk}^+(\boldsymbol{\theta}_G) = \beta_0\{2+2\rho_G(\delta_{jk}; \boldsymbol{\theta}_G)\}^{1/2}$, $\beta_{jk}^-(\boldsymbol{\theta}_G) = \beta_0\{2-2\rho_G(\delta_{jk}; \boldsymbol{\theta}_G)\}^{1/2}$, $\delta_{jk} = \|\mathbf{s}_j - \mathbf{s}_k\|$. 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

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

where $\omega_{jk}^+ \geq 0$ and $\omega_{jk}^- \geq 0$ are some non-negative weights associated to each pair of sites $\{\mathbf{s}_j, \mathbf{s}_k\}$ 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 (Millar, 1984).

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 and 3.3, before illustrating our proposed methodology by application to real temperature data in Section 4.2.

We use the algorithms presented in Section 2.4 to simulate datasets comprised of $n = 100, 200, 500$ replicates of the Cauchy convolution process $Z(\mathbf{s})$ in (2) for some kernel function $k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K)$,

its mixture process extension $\tilde{Z}(\mathbf{s})$ in (6), and the extreme-value limit $Z_{\text{EV}}(\mathbf{s})$ in (5), at $d = 9, 16, 25$ sites on the regular grid $\{(j/(m+1), k/(m+1))^\top\}_{j,k=1}^m \subset [0, 1]^2$, for $m = 3, 4, 5$. To illustrate the methods, we consider here the compactly supported kernel function $k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) = (1 - \|\mathbf{s} - \mathbf{s}^*\|/r)_+^\eta$ with true kernel parameters chosen as $\boldsymbol{\theta}_K = (\eta, r)^\top = (1, 0.25)^\top$. For the mixture process $\tilde{Z}(\mathbf{s})$ 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

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

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

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

where $g_{jk}(\phi; r) = 1 - \max\{0, 1 + \|\mathbf{s}_j - \mathbf{s}_k\|/(2r \sin \phi)\}$. 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 (i.e., n replicates of the process $Z(\mathbf{s})$, or $\tilde{Z}(\mathbf{s})$, or $Z_{\text{EV}}(\mathbf{s})$), we assume that the marginal distributions are unknown, and for simplicity, we estimate them using the empirical distribution in this simulation study (separately at each site). Dependence (i.e., copula) parameters are then estimated in a second step based on the inference methods described in Sections 3.2, 3.3, or by a pairwise likelihood approach with equal weights, respectively. Here, we

simply set the weights to $\omega_{jk}^+ = \omega_{jk}^- = 1$ for all pairs of sites $\{\mathbf{s}_j, \mathbf{s}_k\}$ in (8), (11), and (13), though this may not be the optimal choice. Parameters to be estimated are the kernel parameters $\boldsymbol{\theta}_K = (\eta, r)^\top$ for $Z(\mathbf{s})$ and $Z_{\text{EV}}(\mathbf{s})$, and $(\eta, r, \theta_G, \beta)^\top$ for the mixture process $\tilde{Z}(\mathbf{s})$. 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 $\Delta_{\max} = \frac{1}{N} \sum_{i=1}^N \max_{x \in (0,1)} |k(\mathbf{s}_0, \mathbf{s}(x); \boldsymbol{\theta}_K) - k(\mathbf{s}_0, \mathbf{s}(x); \hat{\boldsymbol{\theta}}_{K,i})|$, and $\Delta_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N \int_0^1 |k(\mathbf{s}_0, \mathbf{s}(x); \boldsymbol{\theta}_K) - k(\mathbf{s}_0, \mathbf{s}(x); \hat{\boldsymbol{\theta}}_{K,i})| dx$, where $\mathbf{s}_0 = (0, 0)^\top$ and $\mathbf{s}(x) = (0, x)^\top$, while $\boldsymbol{\theta}_K = (1, 0.25)^\top$ denotes the true kernel parameters and $\hat{\boldsymbol{\theta}}_{K,i}$ is its estimate for the i -th simulation ($i = 1, \dots, N$). Hence, Δ_{\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 $(\Delta_{\max}, \Delta_{\text{avg}})$ 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 3 shows the estimated kernel profiles along the x -axis, i.e., $k(\mathbf{s}_0, \mathbf{s}(x), \hat{\boldsymbol{\theta}}_{K,i})$ with $0 < x < 1$, for each simulated dataset of $Z(\mathbf{s})$ and $Z_{\text{EV}}(\mathbf{s})$ with $d = 25$. 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.

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 20 stations, namely

Table 1: Performance metrics (i.e., RMSE of estimated parameters and $(\Delta_{\max}, \Delta_{\text{avg}})$) calculated for the inference schemes detailed in Sections 3.2, 3.3, and for a pairwise likelihood method for data simulated from the processes $Z(\mathbf{s})$ (top panel), $\tilde{Z}(\mathbf{s})$ (middle panel), and $Z_{\text{EV}}(\mathbf{s})$ (bottom panel), respectively. The true kernel function is chosen as $k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) = (1 - \|\mathbf{s} - \mathbf{s}^*\|/r)_+^\eta$, with parameters $\boldsymbol{\theta}_K = (\eta, r)^\top = (1, 0.25)^\top$. For the mixture process $\tilde{Z}(\mathbf{s})$, 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 $d = 9, 16, 25$ locations, with $n = 100, 200, 500$ independent replicates.

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

Sample size	RMSE for $\boldsymbol{\theta}_K^\top = (\eta, r)$ (top) and $(\Delta_{\max}, \Delta_{\text{avg}})$ (bottom)		
	$d = 9$	$d = 16$	$d = 25$
$n = 100$	(3.85, 0.25)	(3.26, 0.19)	(2.93, 0.17)
	(0.37, 0.05)	(0.19, 0.03)	(0.19, 0.03)
$n = 200$	(2.47, 0.17)	(1.58, 0.10)	(1.08, 0.06)
	(0.29, 0.04)	(0.13, 0.02)	(0.12, 0.02)
$n = 500$	(1.14, 0.08)	(0.47, 0.03)	(0.38, 0.02)
	(0.15, 0.02)	(0.08, 0.01)	(0.06, 0.01)

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

Sample size	RMSE for $(\eta, r, \theta_G, \beta)$ (top) and $(\Delta_{\max}, \Delta_{\text{avg}})$ (bottom)		
	$d = 9$	$d = 16$	$d = 25$
$n = 100$	(3.08, 0.17, 0.78, 1.18)	(1.88, 0.11, 0.83, 1.25)	(1.71, 0.10, 0.86, 1.32)
	(0.25, 0.03)	(0.17, 0.02)	(0.17, 0.02)
$n = 200$	(1.67, 0.10, 0.71, 1.02)	(0.83, 0.05, 0.69, 1.02)	(0.41, 0.03, 0.68, 1.01)
	(0.19, 0.03)	(0.14, 0.02)	(0.12, 0.01)
$n = 500$	(0.48, 0.04, 0.47, 0.70)	(0.32, 0.03, 0.41, 0.61)	(0.25, 0.03, 0.37, 0.55)
	(0.14, 0.02)	(0.12, 0.01)	(0.10, 0.01)

Data simulated from the extreme-value process $Z_{\text{EV}}(\mathbf{s})$ in (5), with inference based on a pairwise likelihood.

Sample size	RMSE for $\boldsymbol{\theta}_K^\top = (\eta, r)$ (top) and $(\Delta_{\max}, \Delta_{\text{avg}})$ (bottom)		
	$d = 9$	$d = 16$	$d = 25$
$n = 100$	(3.53, 0.23)	(2.04, 0.14)	(1.59, 0.11)
	(0.30, 0.04)	(0.14, 0.02)	(0.11, 0.02)
$n = 200$	(2.76, 0.18)	(1.23, 0.08)	(0.84, 0.06)
	(0.21, 0.03)	(0.11, 0.02)	(0.08, 0.01)
$n = 500$	(1.59, 0.10)	(0.41, 0.03)	(0.30, 0.02)
	(0.12, 0.02)	(0.07, 0.01)	(0.05, 0.01)

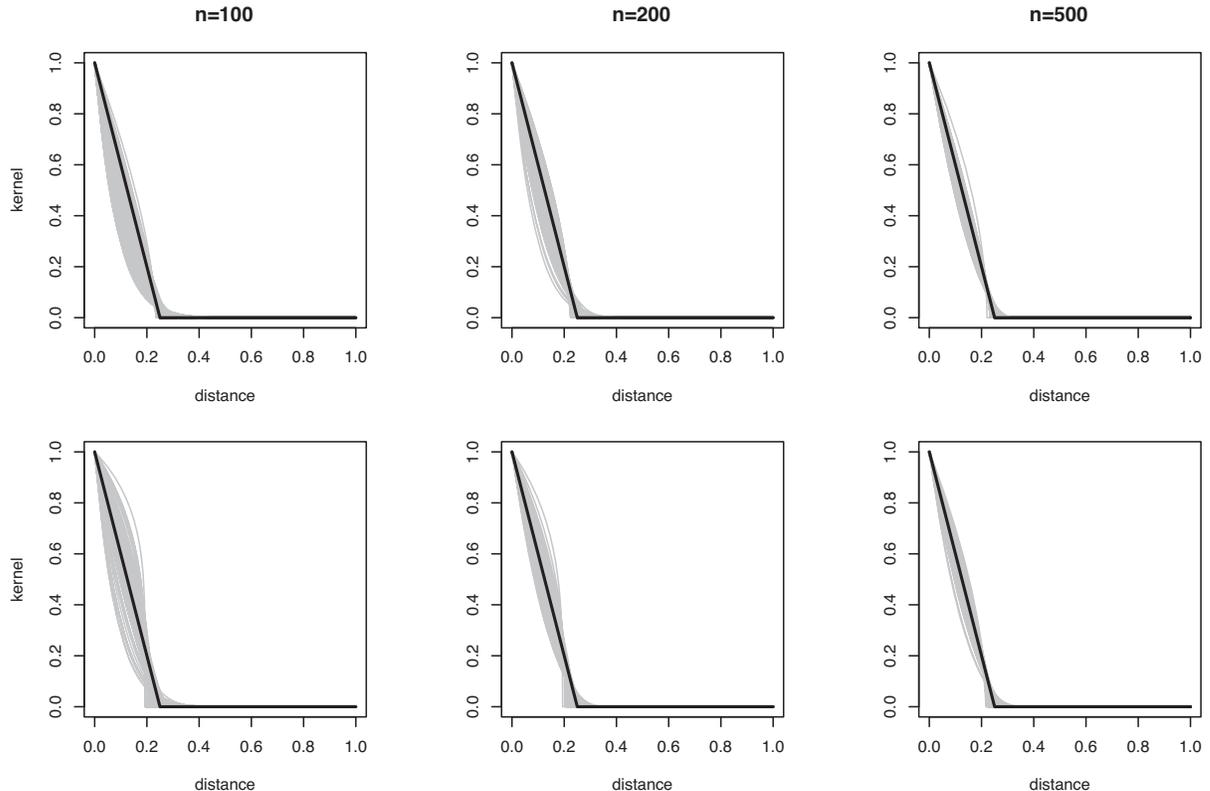


Figure 3: Estimated kernel profiles $k(\mathbf{s}_0, \mathbf{s}(x); \hat{\boldsymbol{\theta}}_{K,i})$, $0 < x < 1$, $i = 1, \dots, N$ (light gray lines), based on $N = 500$ simulated processes $Z(\mathbf{s})$ (top row) and extreme-value limit $Z_{EV}(\mathbf{s})$ (bottom row), with $n = 100$ (left column), 200 (middle column) and 500 (right column) independent replicates. Black thick solid lines show the true kernels $k(\mathbf{s}_0, \mathbf{s}(x; \boldsymbol{\theta}_K))$ with $\boldsymbol{\theta}_K = (1, 0.25)^\top$.

Acme, Ada, Ardmore2, Byars, Centrahoma, Chikasha, Durant, Fittstown, Ketchum Ranch, Lane, Madill, Minco, Newport, Norman, Pauls Valley, Ringling, Shawnee, Tishomingo, Washington, and Waurika. 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(3) 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, 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

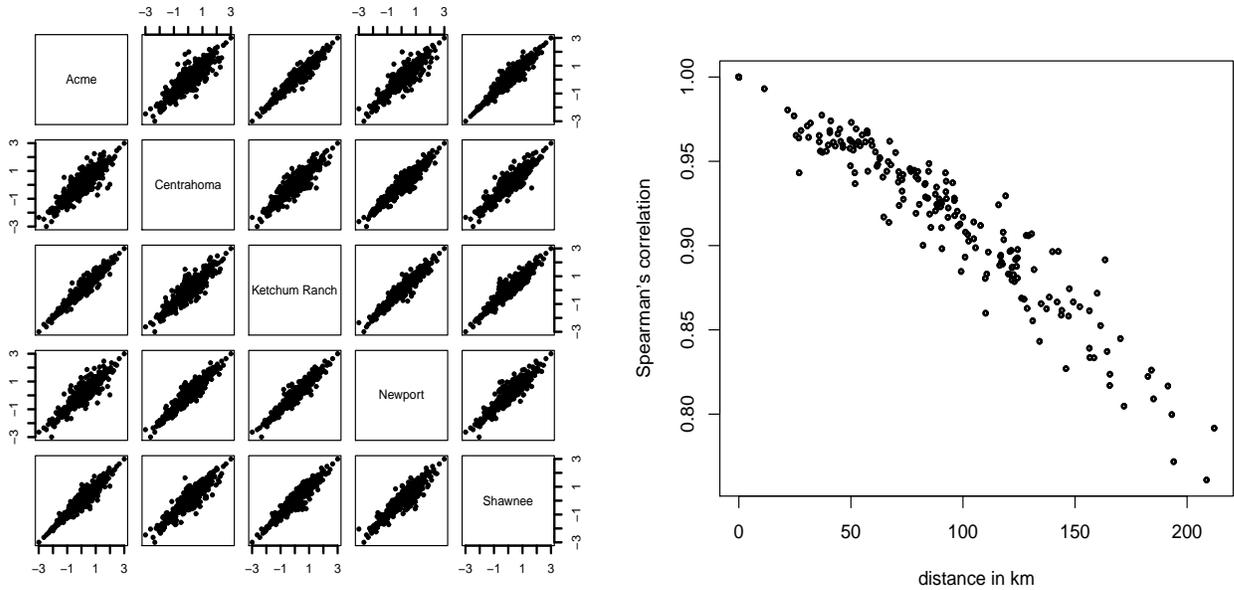


Figure 4: *Left*: Bivariate scatter plots of normal scores based on the temperature data residuals (obtained from preliminary marginal fits) for some selected pairs of stations. *Right*: Spearman's correlation estimates for all pairs of stations, plotted as a function of distance between stations.

Figure 4 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 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 4 shows Spearman's correlation estimates for all pairs of stations, plotted as a function of distance between stations. 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.

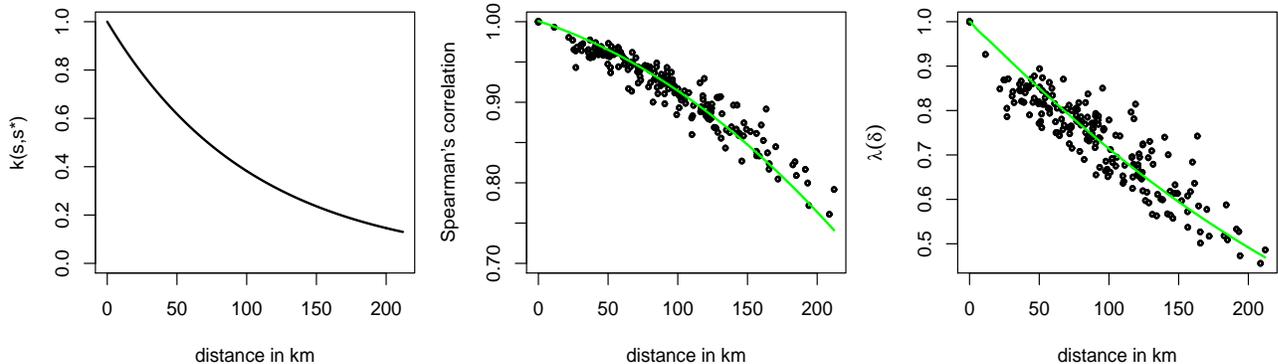


Figure 5: *Left*: Fitted kernel function $k(\mathbf{s}, \mathbf{s}^*; \hat{\boldsymbol{\theta}}_K) = (1 - \|\mathbf{s} - \mathbf{s}^*\|/\hat{r})_+^{\hat{\eta}}$ with $\hat{\boldsymbol{\theta}}_K = (\hat{\eta}, \hat{r})^\top = (53742, 5586)^\top$. *Middle*: Empirical estimates of Spearman's correlation for all pairs of stations (black dots) and fitted Spearman's correlation function $S_\rho(\delta)$ (solid green line), plotted as a function of distance δ between stations. *Right*: Empirical upper tail dependence coefficient estimates for all pairs of stations (black dots) and fitted tail dependence function $\lambda(\delta)$ (solid green line), plotted as a function of distance δ between stations.

All the stations are located in a rather small geographical region in the Southern part of Oklahoma State, and the largest distance between any two stations is only 212 km. We can therefore reasonably assume that the data are stationary over space. We select the kernel function $k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) = (1 - \|\mathbf{s} - \mathbf{s}^*\|/r)_+^\eta$, $\boldsymbol{\theta}_K = (\eta, r)^\top \in (0, \infty)^2$, and estimate its parameters using the inference approach described in Section 3.3. Notice that although this kernel is compactly supported on $[0, r]$ with $r > 0$ representing the dependence range, it reduces to the exponential kernel $k(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_K) = \exp(-\xi\|\mathbf{s} - \mathbf{s}^*\|)$ when $\eta := \eta(r)$ such that $\eta(r) \rightarrow \infty$ and $\eta(r)/r \rightarrow \xi > 0$ as $r \rightarrow \infty$. The estimated kernel parameters are $\hat{\boldsymbol{\theta}}_K = (\hat{\eta}, \hat{r})^\top = (53742, 5586)^\top$ with the estimated range $\hat{r} > 0$ expressed in units of 1000 km. Therefore, as expected, our results here imply fairly long range dependence for this temperature dataset, with $k(\mathbf{s}, \mathbf{s}^*) \approx e^{-9.62\|\mathbf{s} - \mathbf{s}^*\|}$; the fitted kernel function is plotted on the left panel of Figure 5, 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 keeping it fixed a priori. We then estimate the remaining parameters of the mixture process (6) with the underlying Gaussian process specified with a powered exponential correlation function $\rho_G(\delta) = \{1 - \tau\mathbb{I}(\delta = 0)\} \exp(-\theta_G\delta^\alpha)$, $\delta > 0$, 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 $\boldsymbol{\theta}_K = (53742, 5586)^\top$ as estimated above, and then estimate the remaining parameters $(\theta_G, \alpha, \tau, \beta)^\top$ using the least squares approach (13) described in Section 3.3. We obtain $\hat{\theta}_G = 2.97$ (0.86), $\hat{\alpha} = 1.98$ (0.26), $\hat{\tau} = 0.035$ (0.011), and $\hat{\beta} = 1.80$ (0.32). The standard errors calculated using the bootstrap are shown in parentheses. The fitted model thus corresponds to a Cauchy convolution model mixed with a very smooth Gaussian process. Since $\hat{\beta}$ is positive and quite far from zero, the fitted mixture process $\tilde{Z}(\mathbf{s})$ turns out to be quite different from the Cauchy convolution model $Z(\mathbf{s})$, providing increased flexibility to capture the behavior in the bulk of the distribution. Furthermore, notice that the estimated nugget effect $\hat{\tau}$ is very small, indicating negligible (yet statistically significant) small-scale variability.

To assess the goodness-of-fit for the mixture process (6), we then compute empirical and model-based Spearman's correlation estimates $S_\rho(\delta)$ for all pairs of sites $\{\mathbf{s}_j, \mathbf{s}_k\}$ and plot them in the middle panel of Figure 5 as a function of distance $\delta = \|\mathbf{s}_j - \mathbf{s}_k\|$. Model-based estimates are obtained by Monte Carlo using a large number of simulations from the fitted model. Similarly, the right panel of Figure 5 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 shows 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 $\lambda(\delta)$).

5 Conclusion

In this paper, we have proposed a Cauchy kernel convolution copula model for non-Gaussian spa-

tial data and studied its dependence properties, with a particular eye on its tail behavior. In particular, with compactly support 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 extended model, 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 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 (1990) max-stable process. 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 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 sub-asymptotic levels. 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 interesting to study how to

modify our model to capture rougher tail dependence structures of Brown–Resnick type.

References

- Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44, 182–198.
- Bárdossy, A., 2006. Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research* 42.
- Bárdossy, A., Li, J., 2008. Geostatistical interpolation using copulas. *Water Resources Research* 44.
- Bopp, G., Shaby, B.A., Huser, R., 2020. A hierarchical max-infinitely divisible spatial model for extreme precipitation. *Journal of the American Statistical Association* To appear.
- 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.
- Calder, C., Cressie, N. A., 2007. Some topics in convolution-based spatial modeling. *Bulletin of the International Statistical Institute* 62, 132–139.
- 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.
- 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.
- Davison, A.C., Gholamrezaee, M.M., 2012. Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences* 468, 581–608.
- Davison, A.C., Huser, R., 2015. Statistics of Extremes. *Annual Review of Statistics and its Application* 2, 203–235.
- 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.
- Davison, A.C., Padoan, S., Ribatet, M., 2012. Statistical Modelling of Spatial Extremes (with Discussion). *Statistical Science* 27, 161–186.
- Engelke, S., Hitz, A.S., 2020. Graphical models for extremes. *Journal of the Royal Statistical Society, Series B (with Discussion)* 82, 871–932.
- 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.
- 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.

- Fasen, V., 2005. Extremes of regularly varying Lévy-driven mixed moving average processes. *Advances in Applied Probability* 37(4), 993–1014.
- de Fondeville, R., Davison, A.C., 2018. High-dimensional peaks-over-threshold inference. *Biometrika* 105, 575–592.
- Gneiting, T., 2002. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* 97, 590–600.
- 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.
- 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.
- 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*.
- de Haan, L., 1984. A Spectral Representation for Max-stable Processes. *Annals of Probability* 12, 1194–1204.
- Hazra, A., Huser, R., 2020. Estimating high-resolution Red sea surface temperature hotspots, using a low-rank semiparametric spatial model. *Annals of Applied Statistics* To appear.
- 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.
- Hua, L., 2017. On a bivariate copula with both upper and lower full-range tail dependence. *Insurance: Mathematics and Economics* 73, 94–104.
- 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.
- Huser, R., Dombry, C., Ribatet, M., Genton, M.G., 2019. Full likelihood inference for max-stable data. *Stat* 8, e218.
- Huser, R., Genton, M.G., 2016. Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological and Environmental Statistics* 21, 470–491.
- Huser, R., Opitz, T., Thibaud, E., 2017. Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*. 21, 166–186.
- Huser, R., Opitz, T., Thibaud, E., 2020. Max-infinitely divisible models and inference for spatial extremes. *Scandinavian Journal of Statistics* To appear.
- Huser, R., Wadsworth, J.L., 2019. Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association* 114, 434–444.
- Huser, R., Wadsworth, J.L., 2020. Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics* To appear.

- Hüsler, J., Reiss, R. D., 1989. Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters* 7, 283–286.
- 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.
- Kabluchko, Z., Schlather, M., de Haan, L., 2009. Stationary max-stable fields associated to negative definite functions. *Annals of Probability* 37, 2042–2065.
- Krupskii, P., Genton, M. G., 2019. A copula model for non-Gaussian multivariate spatial data. *Journal of Multivariate Analysis* 169, 264–277.
- Krupskii, P., Huser, R., Genton, M. G., 2018. Factor copula models for replicated spatial data. *Journal of the American Statistical Association* 521, 467–479.
- Krupskii, P., Joe, H., 2013. Factor copula models for multivariate data. *Journal of Multivariate Analysis* 120, 85–101.
- Kurowicka, D., Joe, H., 2011. *Dependence Modeling: Vine Copula Handbook*. World Scientific, Singapore.
- Ledford, A.W., Tawn, J.A., 1996. Statistics for near independence in multivariate extreme values. *Biometrika* 83, 169–187.
- Marshall, W. A., Olkin, I., 1967. A multivariate exponential distribution. *Journal of the American Statistical Association* 62, 30–44.
- Millar, P. W., 1984. A general approach to the optimality of minimum distance estimators. *Transactions of the American Mathematical Society* 286, 377–418.
- 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.
- 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.
- Opitz, T., 2017. Spatial random field models based on lévy indicator convolutions. arXiv preprint arXiv:1710.06826.
- Paciorek, C. J., Schervish, M. J., 2006. Spatial modelling using a new class of nonstationary-covariance functions. *Environmetrics* 17, 483–506.
- 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.
- Rootzén, H., 1978. Extremes of moving average of stable processes. *The Annals of Probability* 6(5), 847–869.
- Rootzén, H., Segers, J., Wadsworth, J.L., 2018. Multivariate peaks over thresholds models. *Extremes* 21, 115–145.
- Sato, K., 1999. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, UK.

- Schlather, M., 2002. Models for Stationary Max-Stable Random Fields. *Extremes* 5, 33–44.
- Segers, J., 2012. Max-stable models for multivariate extremes. *REVSTAT* 10, 61–82.
- Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Institute of Statistics of the University of Paris 8, 229–231.
- Smith, R., 1990. Max-stable processes and spatial extremes. Department of Mathematics, University of Surrey .
- Strokorb, K., Ballani, F., Schlather, M., 2015. Tail correlation functions of max-stable processes. *Extremes* 18, 241–271.
- Su, J., Hua, L., 2017. A general approach to full-range tail dependence copulas. *Insurance: Mathematics and Economics* 77, 49–64.
- Thibaud, E., Mutzner, R., Davison, A.C., 2013. Threshold modeling of extreme spatial rainfall. *Water Resources Research* 49, 4633–4644.
- Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. *Statistica Sinica* 21, 5–42.
- Vettori, S., Huser, R., Genton, M.G., 2019. Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics* 75, 831–841.
- 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.
- Wadsworth, J.L., Tawn, J.A., 2012. Dependence modelling for spatial extremes. *Biometrika* 99, 253–272.
- Wadsworth, J.L., Tawn, J.A., 2019. Higher-dimensional spatial extremes via single-site conditioning. arXiv preprint 1912.06560.
- 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.
- 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.