Modeling spatial processes with unknown extremal dependence class

Raphaël G. Huser

CEMSE Division, King Abdullah University of Science and Technology

and

Jennifer L. Wadsworth

Department of Mathematics and Statistics, Lancaster University

March 20, 2017

Abstract

Many environmental processes exhibit weakening spatial dependence as events become more extreme. Well-known limiting models, such as max-stable or generalized Pareto processes, cannot capture this, which can lead to a preference for models that exhibit a property known as asymptotic independence. However, weakening dependence does not automatically imply asymptotic independence, and whether the process is truly asymptotically (in)dependent is usually far from clear. The distinction is key as it can have a large impact upon extrapolation, i.e., the estimated probabilities of events more extreme than those observed. In this work, we present a single spatial model that is able to capture both dependence classes in a parsimonious manner, and with a smooth transition between the two cases. The model covers a wide range of possibilities from asymptotic independence through to complete dependence, and permits weakening dependence of extremes even under asymptotic dependence. Censored likelihood-based inference for the implied copula is feasible in moderate dimensions due to closed-form margins. The model is applied to oceanographic datasets with ambiguous true limiting dependence structure.

Keywords: asymptotic dependence and independence; censored likelihood inference; copula; threshold exceedance; spatial extremes.

1

1 Introduction

The statistical modeling of spatial extremes has received much attention since the article of Padoan et al. (2010) provided a method of inference for *max-stable processes.* The latter form an important class of models for spatial extremes, as they arise as the only non-degenerate limits of renormalized pointwise maxima of spatial stochastic processes. More precisely, let $Y_i(\boldsymbol{s})$, $i = 1, 2, \ldots$, be independent and identically distributed copies of a stochastic process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ with index set $\mathcal{S} \subset \mathbb{R}^2$. If there exist functions $a_n(\boldsymbol{s}) > 0, b_n(\boldsymbol{s})$ such that the limiting process

$$M(\boldsymbol{s}) = \lim_{n \to \infty} \max_{1 \leq i \leq n} \frac{Y_i(\boldsymbol{s}) - b_n(\boldsymbol{s})}{a_n(\boldsymbol{s})} \tag{1}$$

has non-degenerate marginals, then $M(\boldsymbol{s})$ is a max-stable process (Davison et al., 2012). In practice, to employ these asymptotic max-stable processes as a statistical model for spatial extremes, one takes the pointwise maximum of spatially-indexed observations over a large but finite block, such as the annual maximum of daily observations.

Taking such pointwise maxima over many repetitions of $Y(\boldsymbol{s})$ is one way to examine its extremal behavior. However, there are two major practical issues with max-stable processes:

(i) their d-dimensional densities (and hence the likelihood function) are difficult to evaluate, as the number of terms involved equals the dth Bell number, which grows super-exponentially with d (Castruccio et al., 2016);

(ii) a large class of processes $Y(\boldsymbol{s})$ exhibit *asymptotic independence*, i.e., the limit $M(\boldsymbol{s})$ in (1) is a white noise process, consisting of everywhere independent random variables.

Issue (i) has now received much attention. Padoan et al. (2010) used pairwise likelihoods, leading to consistent, but inefficient, estimators, whilst Huser and Davison (2013) and Castruccio et al. (2016) suggested using higher-order composite likelihoods. Einmahl

et al. (2016) proposed methodology avoiding the likelihood, but still reliant on pairwise estimation. Wadsworth and Tawn (2014) simplified the likelihood by conditioning on event times (Stephenson and Tawn, 2005), although Wadsworth (2015) and Huser et al. (2016) show how this can lead to bias in finite samples for large dimensions d and weak dependence. In recent papers, Thibaud et al. (2016) and Dombry et al. (2016) offer ways to deal with this using algorithms for missing data, but their solution is tricky to implement as it involves parallel Gibbs samplers on the space of partitions of $\{1, \ldots, d\}$, and can be computationally intensive for some popular max-stable models. Recognizing issue (i) as a consequence of taking maxima in the representation (1), spatial models for high threshold exceedances, which have simpler likelihoods, have become more appealing; see e.g., Ferreira and de Haan (2014); Wadsworth and Tawn (2014); Engelke et al. (2015); Thibaud and Opitz (2015); de Fondeville and Davison (2016). The threshold exceedance analogue of the max-stable process is known as the *generalized Pareto process*, and has a similar asymptotic dependence structure in its joint tail region.

Issue (ii) is perhaps most pertinent for applications. Since asymptotically independent processes can still exhibit strong dependence at high but finite levels, maxima over many repetitions typically exhibit dependence. The archetypal example is the Gaussian process: this is always asymptotically independent for any non-trivial correlation function $\rho(h) < 1$ with distance $h = \|\boldsymbol{s}_j - \boldsymbol{s}_k\| > 0$, $\boldsymbol{s}_j, \boldsymbol{s}_k \in \mathcal{S}$. Mistakenly identifying this asymptotically-vanishing dependence as genuine asymptotic dependence, and fitting a max-stable or generalized Pareto model, will lead to incorrect extrapolation into the joint upper tail (Davison et al., 2013). Appropriately capturing the rate of tail decay in threshold exceedances using asymptotically independent processes seems to be more helpful, although the number of available suitable models is limited. Wadsworth and Tawn (2012) introduce the *inverted max-stable process*, which they show to exhibit a flexible range of asymptotically independent behaviors, whilst

Opitz (2016) introduces Laplace random fields as an alternative to Gaussian processes.

The reason why asymptotically independent processes often represent extremes of environmental processes better than their generalized Pareto counterparts is that many environmental processes exhibit the trait that as events become more extreme, these extremes become more spatially localized. This feature is not captured by limiting asymptotic dependence models (i.e., max-stable or generalized Pareto processes) because their dependence structures are *stable* with respect to level of extremeness. That is, max-stable copulas are invariant to the operation of taking componentwise maxima, whilst generalized Pareto copulas are invariant to taking conditional threshold exceedances of higher levels. To illustrate the issue, Figure 1 shows an extremal dependence summary for some significant wave height hindcast data $Y(\boldsymbol{s})$ from the North Sea (see §4.1). The function plotted is an estimate of

$$\chi_u := \mathrm{P}\{F_1(Y_1) > u, \ldots, F_d(Y_d) > u | F_1(Y_1) > u\}, \qquad Y_j = Y(\boldsymbol{s}_j) \sim F_j, \qquad (2)$$

where the argument $\boldsymbol{s}_j$ denotes the jth spatial location; if the data followed a generalized Pareto process law, then this function should be constant, rather than decreasing, as the quantile u tends to one. This weakening of dependence does not necessarily imply asymptotic independence, however, the distinction between the two cases being whether the limit $\lim_{u \to 1} \chi_u$ is positive (asymptotic dependence) or equal to zero (asymptotic independence). A popular theoretical example that is asymptotically dependent but has decreasing extremal dependence is the Student t process with $\nu > 0$ degrees of freedom.

In practice, because asymptotic properties are always difficult to infer, it is ideal to fit spatial models encompassing both asymptotic dependence classes, and let the data speak for themselves. To our knowledge, the only instance in the literature of such a hybrid spatial extreme model is the max-mixture model of Wadsworth and Tawn (2012). However, in that model, asymptotic independence only occurs at a boundary point of the parameter space, thus inference methods allowing for this are non-regular. Moreover, the model is highly
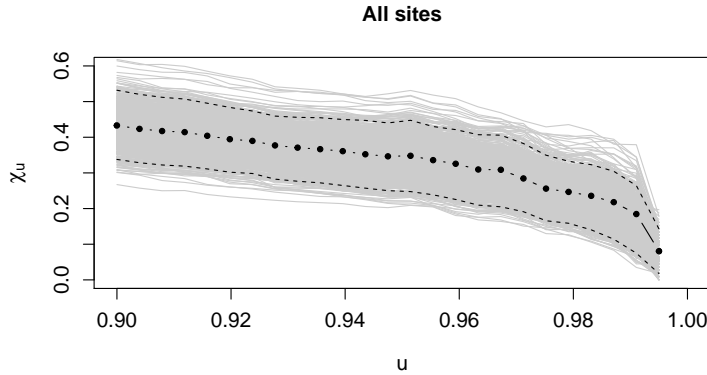
4

All sites

Figure 1: Estimate of the dependence summary χ_u in (2) (dots) for the significant wave height data of §4.1 plotted against quantile, u. Gray lines are estimates from a stationary bootstrap resampling procedure (see §4.1); dashed lines indicate the central 95% of bootstrap samples, pointwise.

parametrized and requires pairwise composite likelihood fitting methods.

In this paper, we address such deficiencies by presenting a class of spatial processes described by a small number of parameters and making a smooth transition between the two dependence paradigms. Specifically, we propose a novel class of spatial extremal models that have non-trivial asymptotically dependent and asymptotically independent submodels with the transition taking place in the interior of the parameter space. The latter property allows us to be formally uncertain about the dependence class, and to quantify this uncertainty in a simple manner. Our new spatial models can thus be viewed as similar in spirit to the generalized extreme-value (GEV) distribution in the univariate case, which was introduced by von Mises (1954) and Jenkinson (1955) as a three-parameter model combining the three limiting extreme-value types (i.e., reversed Weibull, Gumbel and Fréchet), hence providing a way to make inference without specifying the asymptotic distribution family prior to fitting the model. Furthermore, standard hypothesis testing methods can be used to assess the evidence for asymptotic dependence over asymptotic independence, if so desired.

In encompassing both extremal dependence classes, our approach has similarities with

5

the bivariate model of Wadsworth et al. (2017). However our construction here is simpler and substantially more amenable to higher-dimensional inference. Other related work that allows for both asymptotic dependence structures in a spatial setting is the Gaussian scale mixture models proposed in the recent preprint of Huser et al. (2017), but their models either make the transition at a boundary point of the parameter space, or are inflexible in their representation of asymptotic independence structures.

The paper is organized as follows. Section 2 describes our new spatial model, and studies its extremal dependence properties. Section 3 details censored likelihood inference, describes a test for the asymptotic dependence class, and presents a simulation study validating the methodology. The new model is then applied to two oceanographic datasets in Section 4, and Section 5 concludes with some discussion. All proofs are deferred to Appendix A.

2 Model

2.1 Copula-based approach

The main goal of this work is to provide flexible extremal dependence structures for spatial processes. As such, we take a copula-based approach and seek the construction of flexible families of copulas for spatial extremal dependence. For a process with marginal distribution functions $X_j \sim F_j$, the d-dimensional copula function C, is defined as

$$C(u_1, \ldots, u_d) = \mathrm{P}\{F_1(X_1) \leq u_1, \ldots, F_d(X_d) \leq u_d\}.$$

When the margins F_j, $j = 1, \ldots, d$, are continuous, which will be the case throughout this paper, the copula is unique (Sklar, 1959), and represents a multivariate distribution function with standard uniform margins. In §2.2, we describe construction of a model whose copula displays interesting extremal dependence properties. Details of likelihood calculations of the copula of the model we introduce are presented in Section 3.

2.2 Construction

Let $\{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2\}$ be a stationary spatial process with standard Pareto margins, and displaying asymptotic independence with *hidden regular variation*; i.e., for any $x \geq 1$,

$$\mathrm{P}\{W(\boldsymbol{s}_j) > x\} = x^{-1},$$

$$\mathrm{P}\{W(\boldsymbol{s}_j) > x, W(\boldsymbol{s}_k) > x\} = L_W(x) x^{-1/\eta_W(h)}, \quad k \neq j, \tag{3}$$

where $L_W : (0, \infty) \to (0, \infty)$ is a slowly varying function at infinity, i.e., $L_W(ax)/L_W(x) \to 1$ as $x \to \infty$ for any $a > 0$, and $0 < \eta_W(h) < 1$ (Ledford and Tawn, 1996; Resnick, 2002). Note that we exclude the further possibility $\eta_W(h) = 1, L_W(x) \to 0$ as $x \to \infty$, because this does not arise in models that we might naturally consider for $W(\boldsymbol{s})$. The parameter $\eta_W(h)$, called the *coefficient of tail dependence*, summarizes the joint tail decay of the process $W(\boldsymbol{s})$ and it is in general a function of $h = \|\boldsymbol{s}_j - \boldsymbol{s}_k\|$, the distance between sites $\boldsymbol{s}_j, \boldsymbol{s}_k \in \mathcal{S}$. (If the process W is not isotropic, then h should be interpreted as the vector $\boldsymbol{s}_j - \boldsymbol{s}_k$.) For notational convenience, we will write η_W for $\eta_W(h)$ when no confusion can arise. Examples of models satisfying (3) include marginally transformed Gaussian processes and inverted max-stable processes; see §2.5 for more details.

With $W(\boldsymbol{s})$ as described, let R be an independent standard Pareto random variable. Our spatial dependence model is the random field constructed as

$$X(\boldsymbol{s}) = R^\delta W(\boldsymbol{s})^{1-\delta}, \quad \delta \in [0, 1]. \tag{4}$$

The following simple observation highlights why the parsimonious model defined in (4) is useful: when $\delta > 1/2$ then R^δ is heavier-tailed than $W^{1-\delta}$ and this induces asymptotic dependence; when $\delta < 1/2$ the converse is true, and this induces asymptotic independence. These facts are formalized in §2.3.

Construction (4) has superficial similarities with the Gaussian scale mixture models studied by Huser et al. (2017), who multiply a Gaussian random field by a random effect that

determines the extremal dependence properties. However, in (4) the latent process $W(\boldsymbol{s})$ does not have Gaussian margins, resulting in a very different construction in practice, and need not have a Gaussian copula structure, which yields a much wider class of models.

Remark 1. *Representation* (4) *is convenient to study the asymptotic dependence properties of the process* $X(\boldsymbol{s})$ *using the theory of regular variation; see* §2.3. *However, as the copula structure is invariant with respect to monotone marginal transformations, there is an infinite number of ways to characterize the copula stemming from* $X(\boldsymbol{s})$, *some of which may be computationally more attractive or have appealing interpretations. For example, taking the logarithm on both sides of* (4), *we obtain an additive structure*

$$\tilde{X}(\boldsymbol{s}) := \delta\tilde{R} + (1 - \delta)\tilde{W}(\boldsymbol{s}), \tag{5}$$

where $\tilde{R} := \log(R) \sim Exp(1)$ *is independent of* $\tilde{W}(\boldsymbol{s}) := \log\{W(\boldsymbol{s})\}$, *also with* $Exp(1)$ *margins. In Sections* 3 *and* 4, *copula and likelihood computations are based on expression* (5).

The variable R in (4) or equivalently the variable $\tilde{R}$ in (5), may be interpreted in various ways, shedding light on the extremal behavior of $X(\boldsymbol{s})$. For example, by writing $R := \{R(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2\}$ or $\tilde{R} := \{\tilde{R}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2\}$, it can be seen as a random process indexed by $\mathcal{S}$ with perfect dependence, so the representation in (5) implies that $\tilde{X}(\boldsymbol{s})$ can be interpreted as a mixture between perfect dependence and asymptotic independence. This contrasts with Coles and Pauli (2002), who constructed hybrid bivariate models using a certain type of mixture between asymptotic dependence and complete independence.

Alternatively, R or $\tilde{R}$ may be interpreted as an unobserved latent random factor impacting simultaneously the whole region $\mathcal{S}$, hence affecting the joint tail characteristics, and making a link with the common factor copula models for spatial data introduced by Krupskii et al. (2017). One major difference with our approach, however, is that here $\tilde{R}$ and $\tilde{W}(\boldsymbol{s})$ are both on the unit exponential scale, whereas the location mixture copula models of Krupskii

et al. (2017) assume that $\tilde{W}(\boldsymbol{s})$ is a Gaussian process and that both components in (5) are weighted equally, corresponding to $\delta = 1/2$. Consequently, their exponential factor model always displays asymptotic dependence. Other distributions for the random factor were investigated in Krupskii et al. (2017), but they all yield copulas with (non-trivial) asymptotic dependence lying on the boundary of, or at a single point in, the parameter space.

We next study the dependence properties of model (4) for $\delta \in (0, 1)$ noting the simple interpretations at the endpoints of the parameter space: it is clear from (4) or (5) that perfect dependence arises as $\delta \to 1$, whilst the copula of $W(\boldsymbol{s})$ is recovered as $\delta \to 0$.

2.3 Dependence properties

Owing to the simple construction of this process, it is sufficient to study bivariate dependence to make more general conclusions. Comments on higher-dimensional dependence will be made throughout the remainder of the section.

To examine the dependence properties of the process (4), we relate the behavior of the bivariate joint survivor function on the diagonal, $\mathrm{P}(X_j > x, X_k > x)$, to the marginal survivor function, $\mathrm{P}(X_j > x)$, where for simplicity we write $X_j = X(\boldsymbol{s}_j)$ and so forth. We focus on a bivariate version of the dependence measure (2),

$$\chi_u(h) := \mathrm{P}\{F_j(X_j) > u \mid F_k(X_k) > u\}, \qquad \text{and} \qquad X_j \sim F_j, X_k \sim F_k, \qquad (6)$$

and its limit $\chi(h) := \lim_{u \to 1} \chi_u(h)$, with $h = \|\boldsymbol{s}_j - \boldsymbol{s}_k\|$. A value of $\chi(h) > 0$ indicates asymptotic dependence, whilst $\chi(h) = 0$ defines asymptotic independence. Because the process $X(\boldsymbol{s})$ has common margins with upper endpoint at infinity, the limit may be equivalently expressed as

$$\chi(h) := \lim_{x \to \infty} \mathrm{P}(X_j > x, X_k > x)/\mathrm{P}(X_j > x). \qquad (7)$$

When $\chi(h) = 0$, alternative measures are needed to discriminate between the different levels of dependence exhibited by asymptotically independent distributions. A widely satisfied

assumption, already made for the process $W(s)$ in (3) (modulo the restriction made on the coefficient of tail dependence), is

$$P(X_j > x, X_k > x) = L_X\{P(X_j > x)^{-1}\}P(X_j > x)^{1/\eta_X(h)}, \tag{8}$$

where L_X is slowly varying at infinity, and $\eta_X(h) \in (0,1]$ is the coefficient of tail dependence for the process $X(s)$. When $\eta_X(h) = 1$ and $L_X(x) \to \chi(h) > 0$ as $x \to \infty$, the pair of variables $(X_j, X_k)^T$ are asymptotically dependent, else they are asymptotically independent and the value of $\eta_X(h)$ summarizes the strength of extremal dependence in the joint upper tail. For notational convenience, the dependence on distance h in $\chi(h)$ and $\eta_X(h)$ may be omitted when no confusion can arise.

2.3.1 Marginal distribution

The marginal distribution of the process (4) may be established for $\delta \neq 1/2$ as follows:

$$
\begin{aligned}
1 - F_X(x) = P(X_j > x) &= P\{W_j > x^{1/(1-\delta)}R^{-\delta/(1-\delta)}\} \\
&= P(R > x^{1/\delta}) + x^{-1/(1-\delta)}\int_1^{x^{1/\delta}} r^{\delta/(1-\delta)-2}\,\mathrm{d}r \\
&= \frac{\delta}{2\delta-1}x^{-1/\delta} - \frac{1-\delta}{2\delta-1}x^{-1/(1-\delta)}, \qquad x \geq 1, \delta \neq 1/2. \tag{9}
\end{aligned}
$$

The case $\delta = 1/2$ may either be established independently, or as a limit, from which we get

$$1 - F_X(x) = P(X_j > x) = x^{-2}\{2\log(x) + 1\} \qquad x \geq 1, \delta = 1/2;$$

this is the survival function of a log-Gamma random variable with rate and shape parameters both equal to two. Notice that margins are here available in closed form, unlike the Gaussian scale mixture model of Huser et al. (2017), or the bivariate model of Wadsworth et al. (2017). Since the copula is the object of interest in all of the above cases, this makes model (4) computationally more appealing.

2.3.2 Joint distribution

We now derive the joint survivor function of a pair of variables $(X_j, X_k)^T$ from the process $X(\boldsymbol{s})$ in (4), and then use this result in (7), combined with (9), to derive the corresponding coefficients χ and η_X characterizing tail dependence of $X(\boldsymbol{s})$ depending on the value of δ.

Proposition 1. *With definitions and notation as established above, the joint survivor function of* (4) *satisfies*

$$P(X_j > x, X_k > x) = \begin{cases} L(x)x^{-1/\{(1-\delta)\eta_W\}}, & \text{if } \eta_W \geq \delta/(1-\delta), \\ \mathrm{E}\{\min(W_j, W_k)^{(1-\delta)/\delta}\}x^{-1/\delta} - L(x)x^{-1/\{(1-\delta)\eta_W\}}, & \text{if } \eta_W < \delta/(1-\delta), \end{cases}$$

where L represents a generic slowly varying (at infinity) function.

Corollary 1. *If $\delta > 1/2$, the pair $(X_j, X_k)^T$ is asymptotically dependent with*

$$\chi = \mathrm{E}\left\{\min(W_j, W_k)^{(1-\delta)/\delta}\right\}\frac{2\delta - 1}{\delta} = \mathrm{E}\left(\min\left[\frac{W_j^{(1-\delta)/\delta}}{\mathrm{E}\{W_j^{(1-\delta)/\delta}\}}, \frac{W_k^{(1-\delta)/\delta}}{\mathrm{E}\{W_k^{(1-\delta)/\delta}\}}\right]\right) > 0. \quad (10)$$

If $\delta \leq 1/2$, the pair $(X_j, X_k)^T$ is asymptotically independent, i.e., $\chi = 0$. Furthermore, the coefficient of tail dependence for the process (4) *is*

$$\eta_X = \begin{cases} 1, & \text{if } \delta \geq 1/2, \\ \delta/(1-\delta), & \text{if } \eta_W/(1+\eta_W) < \delta < 1/2, \\ \eta_W, & \text{otherwise.} \end{cases} \quad (11)$$

Remark 2. *Analogous dependence summaries in d dimensions are simple to establish using the same techniques of proof as for Proposition 1 and Corollary 1. Specifically, letting $\eta_X^{1:d}$ and $\eta_W^{1:d}$ denote d-dimensional counterparts of the coefficient of tail dependence, defined using the d-dimensional joint survivor function, then expression (11) still holds with η_X and η_W replaced by $\eta_X^{1:d}$ and $\eta_W^{1:d}$. The d-dimensional analogue of χ generalizes expression (10), and is discussed in Remark 3.*

The case $\delta = 1/2$ is of particular interest, since it represents a boundary between asymptotic dependence and asymptotic independence: according to Corollary 1 we have asymptotic

independence ($\chi = 0$), but the coefficient of tail dependence η_X attains its boundary value of 1. In this case, we therefore have $L_X(x) \to 0$ as $x \to \infty$ in (8). Furthermore, the model has the appealing property that $\chi \searrow 0$ as $\delta \searrow 1/2$ and $\eta_X \searrow \eta_W$ as $\delta \searrow \eta_W/(1 + \eta_W)$. As noted at the end of §2.2, as $\delta \searrow 0$, the dependence structure of the W process is recovered. Our model $X(\boldsymbol{s})$ in (4) hence provides a smooth interpolation from the asymptotically independent submodel $W(\boldsymbol{s})$ and perfect dependence, as the parameter δ varies in the unit interval, and it transits through non-trivial asymptotically independent and asymptotically dependent submodels.

2.4 Further dependence properties under asymptotic dependence

Here, we outline the connection to other well-known measures of dependence in the case of asymptotic dependence. We focus firstly on a limiting measure, namely the so-called *exponent function*, $V : (0, \infty)^d \to (0, \infty)$ defined for all $x_1, \ldots, x_d > 0$ by

$$V(x_1, \ldots, x_d) = \lim_{t \to \infty} t \left(1 - \mathrm{P} \left[X_1 \leq F_X^{-1} \left\{ 1 - (tx_1)^{-1} \right\}, \ldots, X_d \leq F_X^{-1} \left\{ 1 - (tx_d)^{-1} \right\} \right] \right),$$

which describes the joint dependence of the associated max-stable or generalized Pareto process; see Davison et al. (2012), Cooley et al. (2012), Segers (2012) or Davison and Huser (2015) for recent reviews on max-stable models. We then examine the *sub-asymptotic* behavior under asymptotic dependence, i.e., the mode of convergence towards such limiting structures, which is important in practice for modeling extreme events at observable levels.

Proposition 2. *For $X(\boldsymbol{s})$ as given in model (4) and $\delta > 1/2$,*

$$V(x_1, \ldots, x_d) = \mathrm{E} \left[\max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{\mathrm{E}\{W_j^{(1-\delta)/\delta}\} x_j} \right] = \frac{2\delta - 1}{\delta} \mathrm{E} \left\{ \max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{x_j} \right\}.$$

Remark 3. *It follows from Proposition 2 that the d-dimensional extremal coefficient, $\theta^{1:d} = V(1, \ldots, 1)$, is*

$$\theta^{1:d} = \mathrm{E} \left[\max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{\mathrm{E}\{W_j^{(1-\delta)/\delta}\}} \right] = \frac{2\delta - 1}{\delta} \mathrm{E} \left\{ \max_{j=1,\ldots,d} W_j^{(1-\delta)/\delta} \right\} \in [1, d]. \tag{12}$$

Furthermore the d-dimensional version of χ, $\chi^{1:d}$, in (6) equals

$$\lim_{u \to 1} P\{F_X(X_1) > u, \ldots, F_X(X_d) > u \mid F_X(X_1) > u\} = \frac{2\delta - 1}{\delta} \mathrm{E}\left\{\min_{j=1,\ldots,d} W_j^{(1-\delta)/\delta}\right\},$$

which can be demonstrated directly or by using inclusion-exclusion relationships between these parameters, such as outlined in Rootzén et al. (2017, Section 6).

The behavior of $\chi_u - \chi = \mathrm{P}\{F_X(X_j) > u \mid F_X(X_k) > u\} - \chi$ as $u \to 1$, i.e., the rate at which χ_u converges to its limit χ, determines the flexibility of a process for capturing sub-asymptotic extremal dependence. Proposition 3 demonstrates that the parameterization of model (4) gives flexibility in this rate, meaning that dependence can weaken above the level used for fitting, whilst still allowing for the possibility of asymptotic dependence.

Proposition 3. *For $\delta > 1/2$,*

$$\chi_u - \chi = \chi \frac{1-\delta}{\delta} \left(\frac{\delta}{2\delta - 1}\right)^{(1-2\delta)/(1-\delta)} (1-u)^{(2\delta-1)/(1-\delta)}\{1 + o(1)\}, \qquad u \to 1. \qquad (13)$$

For comparison, generalized Pareto processes have $\chi_u \equiv \chi$ for all u above a certain level (Rootzén et al., 2017), whilst all max-stable processes have $\chi_u - \chi \asymp (1 - u)$, as $u \to 1$. However, as χ is a dependence measure on the scale of the observations rather than maxima, it is less useful in the context of max-stable processes, where the summary (12) is typically used instead. From Proposition 3, we observe a wide range of convergence rates, from very rapid for δ near 1, to rates slower than $(1 - u)$ for $1/2 < \delta < 2/3$. Note that for $\delta \leq 1/2$, the rate $\chi_u - \chi$ is given by the coefficient of tail dependence, η_X; recall (8) and (11).

Figure 2 illustrates the flexibility in extremal dependence structures, by plotting χ_u in (6) as a function of the threshold u and the limit quantity $\chi = \lim_{u \to 1} \chi_u$ in (10), for a range of values of $\delta \in (0, 1)$, and $(W_j, W_k)^T$ following a Gaussian copula with correlation parameter 0.4. Figure 2 also displays the coefficient of tail dependence η_X defined in (8) and (11) as a function of δ for $\eta_W = 0.1, \ldots, 0.9$. The smooth transition from asymptotic dependence to
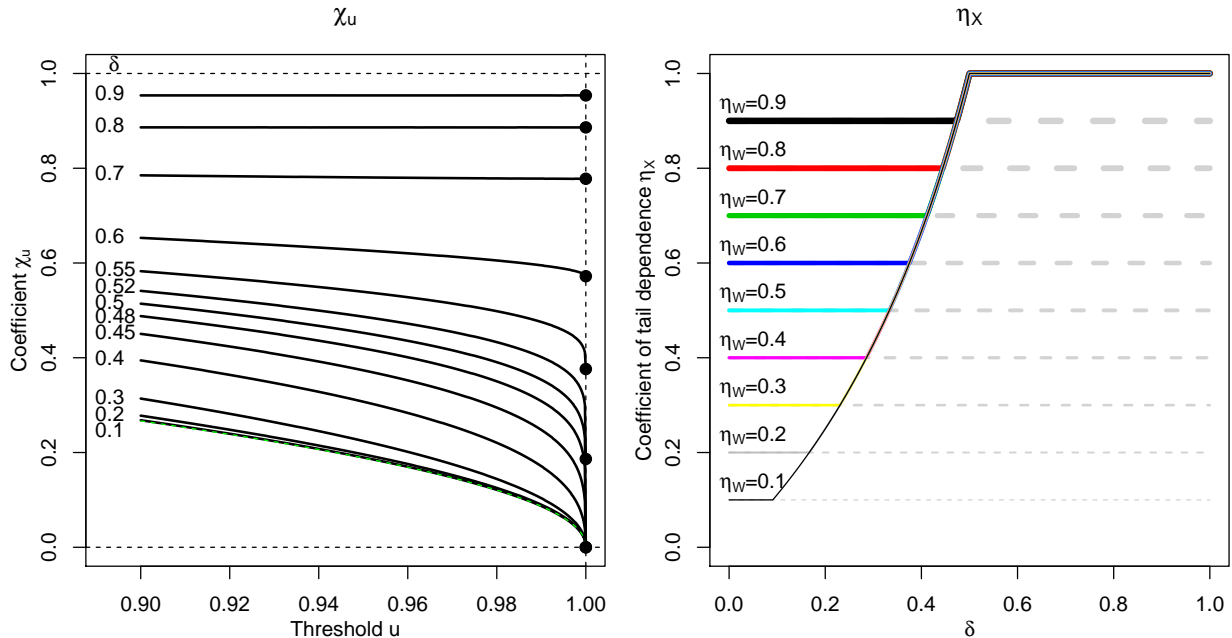
Figure 2: *Left:* Coefficient χ_u in (6) for model (4) plotted as a function of the threshold $u \in [0.9, 1]$ for $\delta = 0.1, 0.2, 0.3, 0.4, 0.45, 0.48, 0.5, 0.52, 0.55, 0.6, 0.7, 0.8, 0.9$ (top to bottom) and $(W_j, W_k)^T$ with a Gaussian copula with correlation 0.4. Black dots at $u = 1$ denote the limit quantities $\chi = \lim_{u \to 1} \chi_u$, and the green dashed curve (superimposed with $\delta = 0.1$) corresponds to the vector $(W_j, W_k)^T$, i.e., the case $\delta = 0$. *Right:* Coefficient of tail dependence η_X in (11) (solid colored curves) as a function of $\delta \in (0, 1)$ for $\eta_W = 0.1, \ldots, 0.9$ (thin to thick). The value $\eta_W = 0.7$ corresponds to the choice of $(W_j, W_k)^T$ on the left panel.

asymptotic independence taking place at $\delta = 1/2$ can be clearly seen from these two plots. Moreover, as is intuitive, the right panel of Figure 2 shows that the process $X(s)$ in (4) cannot reach lower levels of dependence than its underlying $W(s)$ process.

2.5 Example models

We conclude this section with some concrete suggestions for the W process that may be useful in certain applications, such as those described in Section 4.

Example 1 (Inverted max-stable process)**.** Let $\{M(s) : s \in \mathcal{S}\}$ be a stationary max-stable process with extremal coefficient function $\theta(h) \in (1, 2]$, and marginal distribution functions

$G_{\boldsymbol{s}}$, $\boldsymbol{s} \in \mathcal{S}$. Then the process

$$W(\boldsymbol{s}) = 1/G_{\boldsymbol{s}}\{M(\boldsymbol{s})\}$$

has an inverted max-stable copula (Ledford and Tawn, 1996; Wadsworth and Tawn, 2012), Pareto margins, and coefficient of tail dependence $\eta_W(h) = 1/\theta(h)$. The value of $\chi(h)$ in (10) can be calculated as

$$\chi(h) = \frac{2\delta - 1}{\delta}\mathrm{E}\{\min(W_j, W_k)^{(1-\delta)/\delta}\} = \frac{2\delta - 1}{1 - (1 - \delta)\{1 + \eta_W(h)\}}. \tag{14}$$

For this class of processes $\eta_W(h) \in [1/2, 1)$ for $h > 0$, thus the range of $\chi(h)$ values can be established for each δ. Moreover, the d-dimensional quantity $\chi^{1:d}$ takes the same form as in equation (14) with $\eta_W(h)$ replaced by $\eta_W^{1:d} \in [1/d, 1)$.

Example 2 (Gaussian process). Let $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ be a stationary Gaussian process with correlation function $\rho(h)$, and standard Gaussian marginal distribution, denoted Φ. Then

$$W(\boldsymbol{s}) = 1/[1 - \Phi\{Z(\boldsymbol{s})\}]$$

has a Gaussian copula, Pareto margins, and coefficient of tail dependence $\eta_W(h) = \{1 + \rho(h)\}/2$. The value of χ in (10) needs to be calculated either by Monte Carlo or numerical integration, both of which are simple and quick.

In what follows, we will principally take $W(\boldsymbol{s})$ to have a Gaussian copula because the resulting density is much simpler in high dimensions than that of the inverted max-stable process, which suffers the same explosion in the number of terms as a max-stable process density. Outside of a spatial context however, other dependence structures may be preferred.

Example 3 (Non-spatial model). We remark that non-spatial use of the model (4) is also possible, replacing the process $W(\boldsymbol{s})$ with an asymptotically independent random vector $\boldsymbol{W} = (W_1, \ldots, W_d)^T$ with pairwise coefficients of tail dependence $\eta_W^{jk} < 1$, $j < k \in \{1, \ldots, d\}$.

For multivariate models in dimension d greater than two some care is required however, as model (4) allows only for d-wise asymptotic dependence (i.e., $\chi^{1:d} > 0$), or d-wise asymptotic independence (i.e., $\chi^{jk} = 0$, for all $j < k \in \{1, \ldots, d\}$). Such assumptions are natural in the context of spatial processes, but often less so for genuinely multivariate data. For dimension $d = 2$ however, where $\chi^{1:2} > 0$ is the complement of $\chi^{1:2} = 0$, model (4) offers an interesting alternative to that of Wadsworth et al. (2017) for bivariate data. One advantage of model (4) is that a version with an asymmetric dependence structure is simpler to implement, by selecting an asymmetric bivariate distribution for the copula of $\boldsymbol{W}$. We illustrate the improvement this can offer in §4.2.

3   Inference and Simulation

3.1   Censored likelihood

We wish to fit the dependence structure of model (4) to the extremes of spatial processes. Since the dependence characteristics of the model are tailored towards appropriately capturing extremal dependence, we use a censored likelihood, which prevents low values from affecting the estimation of the extremal dependence structure. Such an approach is now standard in inference for multivariate and spatial extremes, although different censoring schemes have been adopted; see e.g. Smith et al. (1997), Wadsworth and Tawn (2012) and Huser et al. (2016, 2017). We assume that we are working with a W process that has a density, so that this is also true for the copula.

Assume that n independent replicates of a random process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2\}$ are observed at d spatial locations, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d \in \mathcal{S}$. Denote the ith replicate at the jth location by Y_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, d$. We assume that in its joint tail region, the process $Y(\boldsymbol{s})$ has the same copula as our model $X(\boldsymbol{s})$ defined in (4), but with possibly different marginal distributions $F_{\boldsymbol{s}}$. To estimate the dependence structure, we first transform the margins to

uniform independently at each site $\boldsymbol{s}_j$, $j = 1, \ldots, d$. In Section 4, we use the semi-parametric procedure of Coles and Tawn (1991), whereby the distribution function is estimated using the asymptotically-motivated generalized Pareto distribution above a high marginal threshold, and the empirical distribution function below that threshold. The resulting variables are denoted $U_{ij} = \hat{F}_{\boldsymbol{s}_j}(Y_{ij})$. An alternative is to use the empirical distribution function throughout. This two-step approach is common practice in the copula literature and provides consistent inference for the copula under mild regularity conditions (see, e.g., Joe, 2015).

The second step is to estimate the copula parameters using the transformed data based on a censored likelihood. When fitting the copula stemming from model (4), the parameters to be estimated are $\boldsymbol{\psi} = (\delta, \boldsymbol{\psi}_W^T)^T \in \Psi = (0, 1) \times \Psi_W \subset \mathbb{R}^p$, where $\boldsymbol{\psi}_W$ is a $(p-1)$-dimensional vector of parameters describing the $W(\boldsymbol{s})$ process. Using the alternative representation $\tilde{X}(\boldsymbol{s}) = \delta \tilde{R} + (1-\delta)\tilde{W}(\boldsymbol{s})$ in (5), the resulting copula C and its density c are

$$C(u_1, \ldots, u_d; \boldsymbol{\psi}) = F_{\tilde{X}}^{1:d}\{F_{\tilde{X}}^{-1}(u_1), \ldots, F_{\tilde{X}}^{-1}(u_d)\}, \tag{15}$$

$$c(u_1, \ldots, u_d; \boldsymbol{\psi}) = f_{\tilde{X}}^{1:d}\{F_{\tilde{X}}^{-1}(u_1), \ldots, F_{\tilde{X}}^{-1}(u_d)\} \prod_{j=1}^{d} \left[f_{\tilde{X}}\{F_{\tilde{X}}^{-1}(u_j)\} \right]^{-1}, \tag{16}$$

where $F_{\tilde{X}}(x) = F_X(e^x)$ and $f_{\tilde{X}}(x) = f_X(e^x)e^x$ with $f_X(x) = \mathrm{d}F_X(x)/\mathrm{d}x$, easily obtained in closed form through (9). The functions $F_{\tilde{X}}$ and $f_{\tilde{X}}$ are the marginal distribution and density, respectively, stemming from the $\tilde{X}(\boldsymbol{s})$ process observed at the sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$, whilst

$$F_{\tilde{X}}^{1:d}(x_1, \ldots, x_d) = \int_0^{r_\delta^\star} F_{\tilde{W}} \left\{ (1-\delta)^{-1}(x_1 - \delta r), \ldots, (1-\delta)^{-1}(x_d - \delta r) \right\} e^{-r} \mathrm{d}r,$$

$$f_{\tilde{X}}^{1:d}(x_1, \ldots, x_d) = (1-\delta)^{-d} \int_0^{r_\delta^\star} f_{\tilde{W}} \left\{ (1-\delta)^{-1}(x_1 - \delta r), \ldots, (1-\delta)^{-1}(x_d - \delta r) \right\} e^{-r} \mathrm{d}r$$

represent the joint distribution function and density, respectively, of this process. Here, $r_\delta^\star = \min(x_1, \ldots, x_d)/\delta$, and $F_{\tilde{W}}, f_{\tilde{W}}$ denote the joint distribution and density, respectively, for the $\tilde{W}(\boldsymbol{s})$ process. The partial derivatives of the copula $C(u_1, \ldots, u_d; \boldsymbol{\psi})$ with respect to

any set of variables $\mathcal{J} \subset \{1, \ldots, d\}$ of cardinality $d_{\mathcal{J}}$ may be expressed as

$$
\begin{aligned}
C_{\mathcal{J}}(u_1, \ldots, u_d; \boldsymbol{\psi}) &= \frac{\partial^{d_{\mathcal{J}}}}{\prod_{j \in \mathcal{J}} \partial u_j} C(u_1, \ldots, u_d; \boldsymbol{\psi}) \\
&= F_{\tilde{X}, \mathcal{J}}^{1:d} \{F_{\tilde{X}}^{-1}(u_1), \ldots, F_{\tilde{X}}^{-1}(u_d)\} \prod_{j \in \mathcal{J}} \left[f_{\tilde{X}} \{F_{\tilde{X}}^{-1}(u_j)\} \right]^{-1},
\end{aligned} \tag{17}
$$

where

$$
F_{\tilde{X}, \mathcal{J}}^{1:d}(x_1, \ldots, x_d) = (1 - \delta)^{-d_{\mathcal{J}}} \int_0^{r_\delta^\star} F_{\tilde{W}, \mathcal{J}} \left\{ (1 - \delta)^{-1}(x_1 - \delta r), \ldots, (1 - \delta)^{-1}(x_d - \delta r) \right\} e^{-r} \mathrm{d}r,
$$

with $F_{\tilde{W}, \mathcal{J}}(x_1, \ldots, x_d) = \partial^{d_{\mathcal{J}}} F_{\tilde{W}}(x_1, \ldots, x_d) / \prod_{j \in \mathcal{J}} \partial x_j$. When the process $W(\boldsymbol{s})$ is based on a Gaussian copula, partial derivatives in (17) involve the multivariate Gaussian distribution in dimension $d - d_{\mathcal{J}}$. Although the unidimensional integrals appearing in (15), (16) and (17) cannot be expressed in closed form, they can nevertheless be accurately approximated using standard finite integration or (quasi) Monte Carlo methods. To estimate the parameters $\boldsymbol{\psi} = (\delta, \boldsymbol{\psi}_W^T)^T \in \Psi$, while avoiding influence of non-extreme data below high marginal thresholds $u_1^\star, \ldots, u_d^\star$, we maximize the censored log likelihood function defined as

$$
\ell(\boldsymbol{\psi}) = \sum_{i=1}^n \log\{L_i(\boldsymbol{\psi})\}, \tag{18}
$$

with contributions

$$
L_i(\boldsymbol{\psi}) = \begin{cases} C(u_1^\star, \ldots, u_d^\star; \boldsymbol{\psi}), & U_{ij} \leq u_j^\star, j = 1, \ldots, d, \\ c(U_{i1}, \ldots, U_{id}; \boldsymbol{\psi}), & U_{ij} > u_j^\star, j = 1, \ldots, d, \\ C_{\mathcal{J}_i}\{\max(U_{i1}, u_1^\star), \ldots, \max(U_{id}, u_d^\star); \boldsymbol{\psi}\}, & U_{ij} > u_j^\star, j \in \mathcal{J}_i \text{ and } U_{ij} \leq u_j^\star, j \in \mathcal{J}_i^c, \end{cases}
$$

where the set of indices $\mathcal{J}_i \subset \{1, \ldots, d\}$ and its complement $\mathcal{J}_i^c = \{1, \ldots, d\} \setminus \mathcal{J}_i$ may be different for each replicate $i = 1, \ldots, n$. The estimator maximizing (18) over Ψ is denoted by $\hat{\boldsymbol{\psi}}$. The performance of this inference approach is assessed in our simulation study §3.2 and it is used in the application in §4.1.

Another possible censoring scheme is to use either the fully censored contribution $C(u_1^\star, \ldots, u_d^\star; \boldsymbol{\psi})$ in (18) if all variables U_{ij} are lower than the threshold $u_j^\star$, $j = 1, \ldots, d$,

or the completely uncensored contribution $c(U_{i1}, \ldots, U_{id}; \boldsymbol{\psi})$ otherwise. This was used by Wadsworth and Tawn (2012), Opitz (2016) and Wadsworth et al. (2017), and is adopted in the example of §4.2, where we compare fits of bivariate models.

3.2 Simulation study

3.2.1 Parameter estimation

To assess the performance of the maximum censored likelihood estimator $\hat{\boldsymbol{\psi}}$ defined through (18), we simulated data from the copula defined by (4) at $d = 2, 5, 10, 15$ locations uniformly generated in the unit square, $[0,1]^2$. We sampled $n = 1000$ independent replicates at these locations, and considered the scenarios $\delta = 0.1, \ldots, 0.9$ (from asymptotic independence to dependence) with $W(\boldsymbol{s})$ defined by a Gaussian copula structure with powered exponential correlation function $\rho(\boldsymbol{s}_j, \boldsymbol{s}_k) = \exp\{-(\|\boldsymbol{s}_j - \boldsymbol{s}_k\|/\lambda)^\nu\}$, $\lambda > 0, \nu \in (0, 2]$. Setting $\lambda = 0.5$ and $\nu = 1$, we then estimated $\boldsymbol{\psi} = (\delta, \lambda, \nu)^T$ by maximizing (18) with marginal thresholds $u_1^\star = \cdots = u_d^\star = 0.95$, giving about 50 exceedances at each location. For identifiability reasons, we fixed $\nu = 1$ when $d = 2$. Because the process is almost perfectly dependent when δ approaches unity, this creates numerical difficulties: to deal with this issue, we increase the relative precision in the R function `integrate` used for the calculation of the integrals in (15), (16) and (17) for larger values of δ. Figure 3 shows boxplots of estimated parameters $\hat{\delta}$ based on 100 independent simulations.

Overall, the estimation procedure works as expected, with boxplots for δ approximately centered around the true value, though a small bias appears for $\delta = 0.8, 0.9$, which is due to numerical instabilities and difficulties in identifying all three parameters in such strong dependence scenarios, despite the higher numerical precision; recall Figure 2. As is typical for a bounded parameter, the asymptotic normality of $\hat{\delta}$ looks to hold well when δ is not too close to 0 and 1, but the distribution displays some asymmetry near to the endpoints 0 and 1. Estimation seems to be easier when $\delta \approx 1/2$, which leads to small bias and variability.
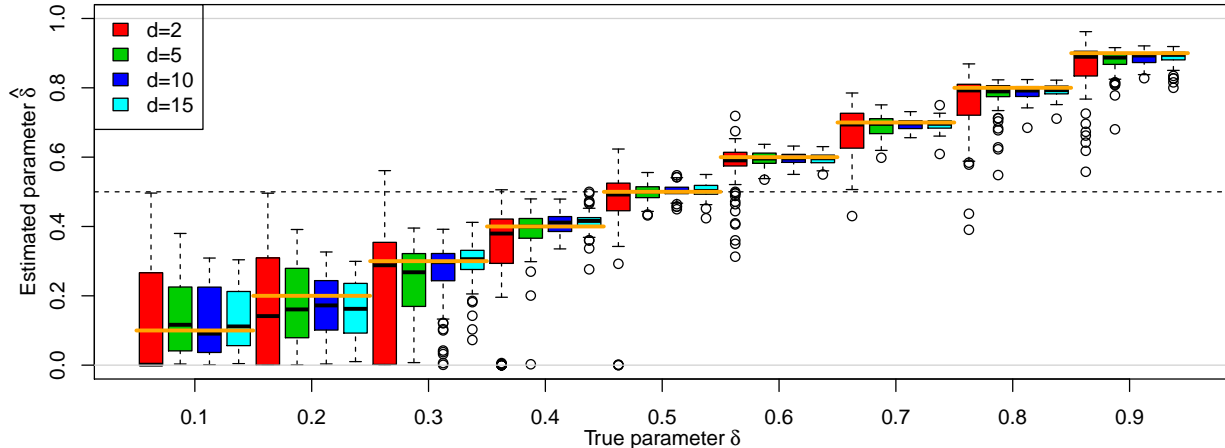
Figure 3: Boxplots for the MLE $\hat{\delta}$ based on (18) setting the thresholds as $u_j^\star = 0.95$, $j = 1, \ldots, d$, for model (4) with $\delta = 0.1, \ldots, 0.9$ (left to right). The underlying process $W(\boldsymbol{s})$ has a Gaussian copula with correlation function $\rho(\boldsymbol{s}_j, \boldsymbol{s}_k) = \exp\{-(\|\boldsymbol{s}_j - \boldsymbol{s}_k\|/\lambda)^\nu\}$ and $\lambda = 0.5$, $\nu = 1$. The process was observed at $d = 2$ (red), 5 (green), 10 (dark blue) and 15 (light blue) random locations in $[0,1]^2$, and $n = 1000$ replicates were simulated. True values are indicated by the orange horizontal segments, and the boundary between asymptotic dependence and independence is indicated by the horizontal dashed line at $\delta = 0.5$. Boxplots are based on 100 independent simulations.

As $\delta \to 0$, the copula structure of $X(\boldsymbol{s})$ converges to that of the latent process $W(\boldsymbol{s})$, here a Gaussian copula, and therefore low values such as $\delta = 0.1, 0.2$ yield very similar dependence structures, leading to higher variability. Boxplots of $\hat{\lambda}$ and $\hat{\nu}$ (see Supplementary Material) suggest that results are better in the asymptotic independence case when $\delta \leq 0.5$. For larger values of δ, the range λ is more variable and the smoothness parameter ν is slightly more biased, owing to the very strong dependence. However, in practice, one could restrict the parameter δ to $0 \leq \delta \leq 0.8$, say, as $\delta > 0.8$ is very unlikely to occur in applications. For all parameters δ, λ and ν, but particularly for ν, the fit improves significantly when more locations are available.

3.2.2 Testing the dependence class

A major advantage of our model (4) over all currently available models for spatial extremes is that we do not need to explicitly determine whether the data exhibit asymptotic dependence

or asymptotic independence in order to select an appropriate class of models. However, since so much effort has previously been placed on determining the appropriate dependence class, we present the details and simulation experiments of a model-based test for this here. Coles et al. (1999) suggest using nonparametric estimators of the measure χ_u (defined slightly differently to (6)) and its counterpart $\bar{\chi}_u$, but when the threshold u increases to unity, the associated uncertainty inflates dramatically. This renders any test based on these nonparametric estimators useless in practice. To increase the power for discriminating between asymptotic dependence and independence, a parametric model-based approach seems more sensible and our copula model (4) provides a very natural way to proceed, because the transition between the two asymptotic paradigms takes place at $\delta = 1/2$, i.e., in the interior of the parameter space. Hence, standard likelihood theory can be invoked to design tests for the null hypotheses

$$H_0^{\mathrm{AD}} : \delta > 1/2 \text{ (asympt. dependence)} \qquad vs \quad H_A^{\mathrm{AI}} \quad : \delta \leq 1/2 \text{ (asympt. independence)};$$

$$H_0^{\mathrm{AI}} : \delta \leq 1/2 \text{ (asympt. independence)} \qquad vs \quad H_A^{\mathrm{AD}} \quad : \delta > 1/2 \text{ (asympt. dependence)}.$$

Let $\hat{\boldsymbol{\psi}} = (\hat{\delta}, \hat{\lambda}, \hat{\nu})^T$ be the maximum likelihood estimator (MLE). We suggest using asymptotic normality of $\hat{\boldsymbol{\psi}}$ to test for H_0^{AD} or H_0^{AI}, an assumption that should hold true if n is large and δ is not too close to its boundaries 0 and 1. In particular, denoting the estimated variance of $\hat{\delta}$ by $\hat{v}_\delta$, the power of these tests at the level $100 \times (1 - \alpha)\%$ can be computed as

$$\mathrm{P}(\text{reject } H_0^{\mathrm{AD}} \mid H_A^{\mathrm{AI}} \text{ holds}) = \mathrm{P}(\hat{\delta} < 0.5 - \sqrt{\hat{v}_\delta} z_{1-\alpha} \mid \delta \leq 0.5), \qquad (19)$$

$$\mathrm{P}(\text{reject } H_0^{\mathrm{AI}} \mid H_A^{\mathrm{AD}} \text{ holds}) = \mathrm{P}(\hat{\delta} > 0.5 + \sqrt{\hat{v}_\delta} z_{1-\alpha} \mid \delta > 0.5), \qquad (20)$$

respectively, where $z_{1-\alpha}$ is the $(1 - \alpha)$-quantile of the standard normal distribution. As $n \to \infty$, the two power curves are symmetric, but for finite n when the asymptotic normality assumption might not hold exactly, they might look slightly asymmetric.
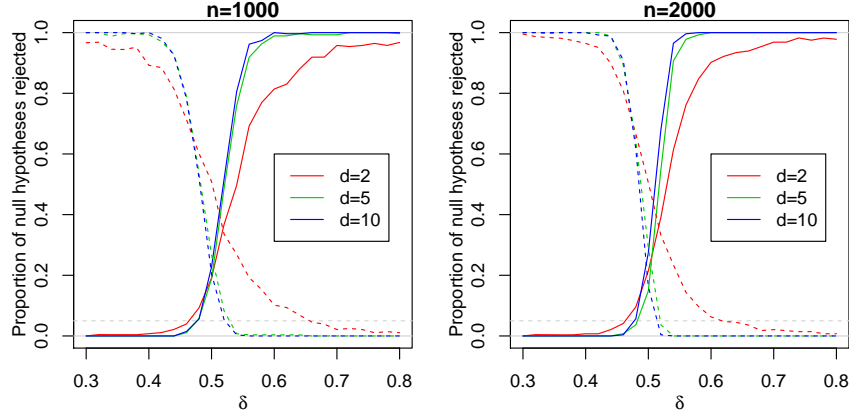
Figure 4: Proportion of times the null hypotheses H_0^{AD} (dashed) and H_0^{AI} (solid) are rejected, plotted against the true δ value, computed from 300 simulations. The horizontal dashed line at 0.05 shows the nominal level used for the tests. The model (4) was simulated at $d = 2$ (red), 5 (green), and 10 (blue) sites in $[0,1]^2$ with $n = 1000$ (left), 2000 (right) replicates, using an underlying Gaussian copula for $W(\boldsymbol{s})$ with correlation function $\rho(\boldsymbol{s}_j, \boldsymbol{s}_k) = \exp\{-(\|\boldsymbol{s}_j - \boldsymbol{s}_k\|/\lambda)^{\nu}\}$ and $\lambda = 0.5$, $\nu = 1$. Parameters were estimated by maximum likelihood based on (18) with marginal thresholds $u_1^{\star} = \cdots = u_d^{\star} = 0.95$.

To compute the power curves (19) and (20), we drew 300 simulations of model (4) at $d = 2, 5, 10$ locations in $[0,1]^2$ with $n = 1000, 2000$ independent replicates, under the same setting as above. Range $\lambda = 0.5$ and smoothness $\nu = 1$ were fixed, and we considered a sequence $\delta \in [0.3, 0.8]$ in steps of 0.02, estimating all parameters using the MLE based on (18) with marginal thresholds $u_1^{\star} = \cdots = u_d^{\star} = 0.95$. The Hessian matrix at the MLE was used in order to compute $\hat{v}_{\delta}$ as the $(1,1)$-entry of the reciprocal Fisher information.

Figure 4 displays the proportion of null hypotheses rejected (i.e., the power curves (19) and (20) when the corresponding null hypotheses are false), estimated using the 300 simulations and plotted as a function of $\delta \in [0.3, 0.8]$. As expected, for all dimensions, the power to reject asymptotic dependence (respectively asymptotic independence) improves as $\delta \to 0$ (respectively $\delta \to 1$), and the curves for the two tests seem fairly symmetric for $d = 5, 10$, as the increased information in these dimensions means asymptotic normality holds well. When $d = 2$, the power is not very high for a wide range of δ values owing to the large variability in estimating δ, but it improves significantly for higher dimensions; the transition at $\delta = 0.5$

is indeed much steeper when $d = 5, 10$. Furthermore, the type-I error is quite large when $d = 2$ but it is much closer to the nominal level at $\delta = 0.5$ when $d = 5, 10$. The comparison of the left and right panels in Figure 4 suggests that the power further increases with larger sample sizes n, as expected. In summary, these numerical results suggest that it is possible to discriminate between asymptotic dependence and asymptotic independence within our model, provided δ is not too close to 0.5, and that the power improves with larger d and n.

4   Oceanographic applications

4.1   Hindcast significant wave height data

Wadsworth and Tawn (2012) considered modeling the extremes of the winter observations of a hindcast dataset of significant wave height, a measure of ocean surface roughness, from the North Sea. Calculating the coefficient of tail dependence $\eta(h)$ for the wave height process, they suggested that there was evidence for asymptotic independence of the process, although strong spatial dependence between sites. Figure 1 suggests a high degree of ambiguity in what the appropriate extremal dependence structure should be, since the summary χ_u is decreasing as u increases, but not necessarily to a value of zero. This ambiguous situation is replicated throughout numerous applications, and demonstrates the necessity for a model such as model (4) that can handle both scenarios.

Measurements of the hindcast are recorded at three-hourly intervals, yielding eight observations per day, over a period of 31 years. In total the dataset of winter (December, January, February) wave heights consists of 22376 observations at 50 locations. Margins are transformed to uniform using the semiparametric transformation of Coles and Tawn (1991). The data are strongly temporally dependent and so we subsample to extract one realization per day, giving 2797 observations. The resulting data still exhibit temporal dependence, but this thinning eases the computational burden of model fitting, whilst the information loss

should be small. Finally, we select a subset of 20 sites to fit the model to, whilst using all data for validation of the fit. Distance is measured in units of latitude (one unit $\approx$ 111km); the range of distances between sites is 0.27–2.99 units.

Maximum likelihood estimates and measures of uncertainty are displayed on the left side of Table 1. The uncertainty measures are based on re-estimation of the model for 200 bootstrap samples, created using the stationary bootstrap (Politis and Romano, 1994). This procedure relies on sampling blocks of geometric length; we sampled using an average length of 14 days, although any blocks that reached the end of February (i.e., the end of one winter) were curtailed, so that observations within a block are always consecutive. Figure 8 in the Supplementary Material indicates that this bootstrap procedure captures the temporal dependence in the extremes adequately.

The MLE of δ indicates asymptotic independence, although the 95% bootstrap confidence interval includes values above 0.5, meaning that firm conclusions about the asymptotic dependence class are difficult to draw; this further highlights the need for models that can incorporate both scenarios. Whilst asymptotic independence is indicated, the value of δ suggests that that our model is more suited than a simple Gaussian model. To reinforce this, we also fit a Gaussian model, using the same censored likelihood scheme, with results reported on the right side of Table 1. Although the Gaussian model is nested within the model we fit, testing is non-standard as it occurs at the boundary of the parameter space, i.e., for $\delta = 0$. However, the maximized log-likelihood for our model was 62 units higher than for the Gaussian model, an appreciable improvement although there is no explicit accounting for temporal dependence in the likelihood.

To assess the fit of the model, we consider two diagnostics. Figure 5 displays the fitted value of χ_u, as defined in (2), for the subset of sites included in the model fit (left panel) and the subset of sites excluded from the fit (right panel). Although the model was fitted using

Table 1: MLEs, standard deviations (SD) of bootstrap replications, and approximate 95% confidence intervals (CI) for parameters of our model (left) and the Gaussian process (right). Figures given to the longer of two decimal places or significant figures.

	MLE	SD	95% CI	MLE	SD	95% CI
δ	0.46	0.039	(0.36,0.54)	–	–	–
λ	3.19	0.26	(2.60,3.71)	3.84	0.17	(3.62,4.26)
ν	1.98	0.0033	(1.97,1.98)	1.97	0.0043	(1.96,1.98)
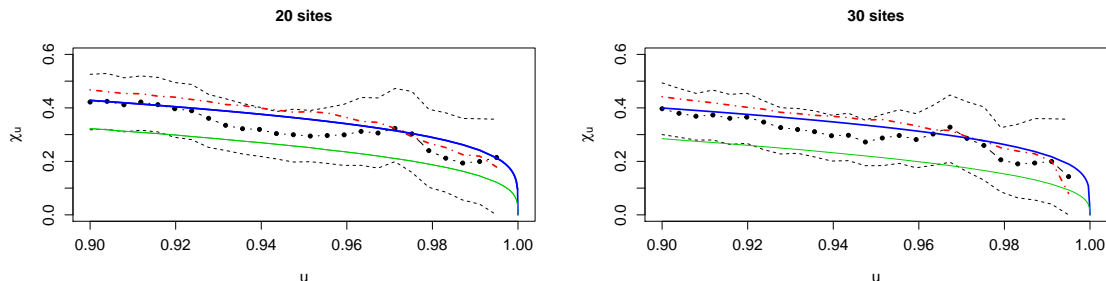
Figure 5: Estimates of χ_u for the hindcast wave height data. Central black dots: empirical estimate of χ_u from the temporally thinned data; dashed lines: approximate 95% confidence intervals based on the stationary bootstrap procedure described in the text; dot-dash red line: empirical estimate of χ_u from all data; thick solid blue line: fit from our model; thin solid green line: fit from the Gaussian model. Left: data to which the model was fitted (from 20 sites); right: data to which the model was not fitted (from 30 sites).

censored likelihood above a 95%-quantile threshold, the fit looks good on the plotted range $u = (0.9, 1)$. The Gaussian model clearly underestimates the dependence.

The second diagnostic we consider is the distribution of the number of threshold exceedances, conditionally upon having at least one exceedance. The Supplementary Material contains histograms of the distribution from our data sample and from the fitted model, and suggests that the fitted model appears to capture this distribution quite well.

4.2   Newlyn oceanographic data

We fit a bivariate version of model (4), as discussed in Example 3, to the Newlyn oceanographic data analyzed in Wadsworth et al. (2017) to illustrate an asymmetric construction, and to compare with the symmetric models fitted therein. The data, shown in Figure 6,

Table 2: AIC for bivariate copula fits to the Newlyn data, using the symmetric model of Wadsworth et al. (2017) (first row) and using our model with the asymmetric inverted Dirichlet model (second row).

	Height–Surge	Period–Surge	Height–Period
AIC WTDE	264.1	515.4	225.7
AIC asymmetric	267.3	493.8	181.6

comprise 2894 observations of wave height, surge and period, and we analyze them pairwise, transforming to uniformity again using the semiparametric transformation of Coles and Tawn (1991). To generate an asymmetric model, we assume that the copula of $(W_1, W_2)^T$ is that of an inverted Dirichlet max-stable distribution. The bivariate Dirichlet max-stable distribution (Coles and Tawn, 1991) has exponent function

$$V(x_1, x_2) = \frac{1}{x_1}\left\{1 - \mathrm{Be}\left(\frac{\alpha x_1}{\alpha x_1 + \beta x_2}; \alpha + 1, \beta\right)\right\} + \frac{1}{x_2}\mathrm{Be}\left(\frac{\alpha x_1}{\alpha x_1 + \beta x_2}; \alpha, \beta + 1\right), \qquad \alpha, \beta > 0,$$

where $\mathrm{Be}(\cdot, a, b)$ is the Beta distribution function with shape parameters a and b. The bivariate inverted max-stable distribution with Pareto margins has joint survivor function $\mathrm{P}(W_1 > w_1, W_2 > w_2) = \exp[-V\{(\log w_1)^{-1}, (\log w_2)^{-1}\}]$. To ensure consistency in the fitting, we use the censored likelihood described in Wadsworth et al. (2017) and described at the end of §3.1 for both models. That is, we use the full density contribution when either variable is above a censoring threshold, which is set to the 95%-quantile in each margin. Table 2 gives the Akaike Information Criterion (AIC) for the model of Wadsworth et al. (2017) and our asymmetric model; improvements are seen for pairs involving wave period, which shows a more asymmetric dependence structure than height and surge. One limitation of this choice for $(W_1, W_2)^T$ is that it cannot exhibit negative dependence, and as such, the model is less flexible when it comes to accounting for dependence structures with weak asymptotic dependence (i.e., with small but positive χ). This does not appear to be an issue for these asymptotically independent pairs, but alternative choices for $(W_1, W_2)^T$ such as the skew bivariate normal (Azzalini and Dalla Valle, 1996) could be used to overcome this.
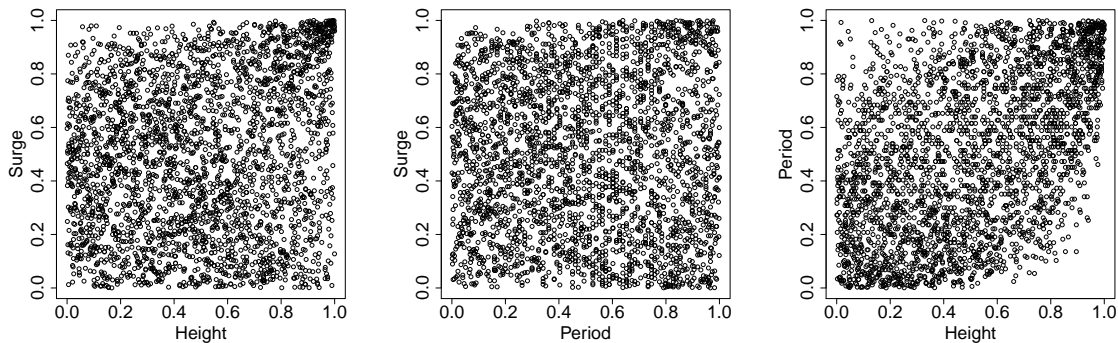
Figure 6: Newlyn wave data on approximate standard uniform margins. From left to right: Height-Surge, Period-Surge, Height-Period.

5   Discussion

Motivated by deficiencies in existing frameworks for modeling spatial extremes, we presented a parsimonious model that is able to capture the sub-asymptotic dependence behavior of spatial processes. Importantly, both extremal dependence classes are captured, with non-trivial structures within each class, and a smooth transition between paradigms at the interior of the parameter space.

Inference for model (4) is feasible in moderate dimensions, but computationally intensive when W has a Gaussian copula, owing to the need to integrate expressions involving a multivariate Gaussian distribution function. However, new quasi-Monte Carlo algorithms, such as those used by de Fondeville and Davison (2016), and the associated R package `mvPot`, have the potential to increase scalability. With the exception of the specific model used in de Fondeville and Davison (2016), truly high-dimensional inference for spatial extreme-value models has yet to be achieved, and our model is competitive with others in this aspect.

There are two notable limitations of the model (4). The first of these is that for certain values of δ, independence between sites will never be attained as the lag $h \to \infty$. This is, however, in common with a majority of models for spatial extremes. Consequently, the model is more suitable for smaller spatial regions or data for which this is not an issue. The

second limitation concerns the link between δ and the limiting value of $\chi(h)$ for $\delta > 1/2$. Since $W(\boldsymbol{s}) \geq 1$ we have $\min(W_j, W_k) \geq 1$ and consequently from (7), $\chi(h) \geq (2\delta - 1)/\delta$. As can be observed from Figure 2 and equation (13), for values of δ near 1, the process (4) behaves similarly to a generalized Pareto process. However, model (4) would be unable to capture a weakly dependent generalized Pareto process, i.e., one for which $\chi_u(h)$ is constant but its limit $\chi(h)$ is small and positive. In practice however, this is not very restrictive, since in our experience almost all environmental datasets display a decreasing χ_u function.

Acknowledgements

Code and data   Code for fitting the models described is available as Supplementary Material and at `http://www.lancaster.ac.uk/~wadswojl/SpatialADAI`. The NEXTRA hindcast data analyzed in § 4.1 are subject to restrictions. Access may be granted for academic purposes by members of the North European Storm Study User Group (NUG); requests can be made using the details at `http://www.oceanweather.com/metocean/next/index.html`. The Newlyn wave data analyzed in § 4.2 are available as Supplementary Material.

A   Proofs

Proof of Proposition 1.

$$\mathrm{P}(X_j > x, X_k > x) = \mathrm{P}(W_j > x^{1/(1-\delta)} R^{-\delta/(1-\delta)}, W_k > x^{1/(1-\delta)} R^{-\delta/(1-\delta)})$$
$$= \mathrm{P}(W_j > S, W_k > S),$$

where $S = x^{1/(1-\delta)}R^{-\delta/(1-\delta)}$, so that S has support $(0, x^{1/(1-\delta)})$, and Lebesgue density $f_S(s) = \frac{1-\delta}{\delta}s^{(1-\delta)/\delta-1}x^{-1/\delta}$ on this interval. Using assumption (3), we have

$$P(W_j > S, W_k > S) = \frac{1-\delta}{\delta}x^{-1/\delta}\int_0^1 s^{(1-\delta)/\delta-1}\,ds + \frac{1-\delta}{\delta}x^{-1/\delta}\int_1^{x^{1/(1-\delta)}} L_W(s)s^{(1-\delta)/\delta-1/\eta_W-1}\,ds$$

$$= x^{-1/\delta} + \frac{1-\delta}{\delta}x^{-1/\delta}\int_1^{x^{1/(1-\delta)}} L_W(s)s^{(1-\delta)/\delta-1/\eta_W-1}\,ds.$$

Consider the behavior of $\int_1^{x^{1/(1-\delta)}} L_W(s)s^{(1-\delta)/\delta-1/\eta_W-1}\,ds$, which is convergent since we have a well defined probability. We will apply Karamata's Theorem (Resnick, 2006, Theorem 2.1) and so distinguish between the cases when the index of regular variation is $\gtreqless -1$. The notation $g \in RV_\rho$ denotes that a function g is regularly varying at infinity with index $\rho \in \mathbb{R}$.

Case 1: $(1-\delta)/\delta - 1/\eta_W - 1 \geq -1$ i.e., $\eta_W \geq \delta/(1-\delta)$. By Karamata's Theorem $\int_1^x L(s)s^\theta\,ds \in RV_{\theta+1}$ when $\theta \geq -1$. Thus

$$\int_1^{x^{1/(1-\delta)}} L_W(s)s^{(1-\delta)/\delta-1/\eta_W-1}\,ds = \tilde{L}(x)x^{1/\delta-1/\{\eta_W(1-\delta)\}},$$

where $\tilde{L}$ is a new SV function, using also a result on composition of regularly varying functions (Resnick, 2006, Prop. 2.6 (iv)). Overall in Case 1 we thus have

$$P(X_j > x, X_k > x) = L(x)x^{-1/\{\eta_W(1-\delta)\}},$$

for some slowly varying function L, noting that terms of order $x^{-1/\delta}$ are absorbed into L when $\eta_W > \delta/(1-\delta)$.

Case 2: $(1-\delta)/\delta - 1/\eta_W - 1 < -1$ i.e., $\eta_W < \delta/(1-\delta)$. By Karamata's Theorem $\int_x^\infty L(s)s^\theta\,ds \in RV_{\theta+1}$ when $\theta < -1$. We have

$$\int_1^{x^{1/(1-\delta)}} L_W(s)s^{(1-\delta)/\delta-1/\eta_W-1}\,ds = \int_1^\infty L_W(s)s^{(1-\delta)/\delta-1/\eta_W-1}\,ds - \int_{x^{1/(1-\delta)}}^\infty L_W(s)s^{(1-\delta)/\delta-1/\eta_W-1}\,ds,$$

$$(21)$$

and so the second term on the right-hand side the is regularly varying of index $1/\delta - 1/\{\eta_W(1-\delta)\}$. The first term on the right-hand side of expression (21) is established by noting that

$$
\begin{aligned}
\mathrm{E}\{\min(W_j, W_k)^{(1-\delta)/\delta}\} &= \int_0^\infty \mathrm{P}\{\min(W_j, W_k)^{(1-\delta)/\delta} > t\}\,\mathrm{d}t \\
&= 1 + \int_1^\infty L_W\{t^{\delta/(1-\delta)}\}t^{-\delta/\{\eta_W(1-\delta)\}}\,\mathrm{d}t = 1 + \frac{1-\delta}{\delta}\int_1^\infty L_W(s)s^{(1-\delta)/\delta - 1/\eta_W - 1}\,\mathrm{d}s.
\end{aligned}
$$

Overall in Case 2 we thus have

$$
\mathrm{P}(X_j > x, X_k > x) = \mathrm{E}\{\min(W_j, W_k)^{(1-\delta)/\delta}\}x^{-1/\delta} - L(x)x^{-1/\{\eta_W(1-\delta)\}},
$$

where, as in Case 1, the factor $(1-\delta)/\delta$ is absorbed into the slowly varying function L of the second term. □

Proof of Corollary 1. Since X has common margins and upper endpoint infinity, the extremal dependence class is determined by the limit

$$
\chi = \lim_{x \to \infty} \frac{\mathrm{P}(X_j > x, X_k > x)}{\mathrm{P}(X_j > x)}.
$$

If $\delta > 1/2$ Then $\mathrm{P}(X_j > x) \sim \frac{\delta}{2\delta - 1}x^{-1/\delta}$. We have $\delta/(1-\delta) > 1$ so we must be in Case 2, and since $\eta_W < 1$ the dominant order in the joint survivor function is also $x^{-1/\delta}$. Thus

$$
\chi = \mathrm{E}\{\min(W_j, W_k)^{(1-\delta)/\delta}\}\frac{2\delta - 1}{\delta} > 0,
$$

with expression (10) following as

$$
\mathrm{E}\{W_j^{(1-\delta)/\delta}\} = \int_1^\infty w^{(1-\delta)/\delta - 2}\,\mathrm{d}w = \frac{\delta}{2\delta - 1}. \tag{22}
$$

If $\delta = 1/2$ Consider the relation (8), and note that since $\mathrm{P}(X_j > x)^{-1}$ is regularly varying with limit infinity, then the composition $L_X\{\mathrm{P}(X_j > x)^{-1}\} =: L^*(x)$ is slowly varying at infinity; cf. Resnick (2006, Prop. 2.6(iv)). For $\delta = 1/2$, we have

$$
\mathrm{P}(X_j > x) = x^{-2}\{2\log(x) + 1\}, \quad \mathrm{P}(X_j > x, X_k > x) = \mathrm{E}\{\min(W_j, W_k)\}x^{-2} - L(x)x^{-2/\eta_W},
$$

since $\eta_W < 1$ by assumption, which puts us in Case 2. We thus have

$$P(X_j > x, X_k > x) = L^*(x)P(X_j > x),$$

so that $\eta_X = 1$ and

$$L^*(x) \sim \frac{E\{\min(W_j, W_k)\}}{\{2\log(x) + 1\}} \to 0, \quad x \to \infty.$$

If $\delta < 1/2$ Then $P(X_j > x) \sim \frac{1-\delta}{1-2\delta} x^{-1/(1-\delta)}$. If $\eta_W \geq \delta/(1-\delta)$ then we are in Case 1 and the survivor function is $L(x)x^{-1/\{\eta_W(1-\delta)\}}$. Otherwise if $\eta_W < \delta/(1-\delta)$ we are in Case 2 and the survivor function decays like $x^{-1/\delta}$. In both cases this leads to $\chi = 0$ with coefficient of tail dependence,

$$\eta_X = \begin{cases} \delta/(1-\delta) & \text{if } \eta_W < \delta/(1-\delta) \\ \eta_W & \text{otherwise.} \end{cases}$$

$\square$

During the proofs of Propositions 2 and 3, we will need results on the quantile function $q(t) := F_X^{-1}\{1 - 1/t\}$, which we give in the following Lemma.

Lemma 1. *For $\delta > 1/2$, the marginal quantile function $q(t) = F_X^{-1}(1 - 1/t)$ satisfies*

$$q(t) = \left(\frac{\delta}{2\delta - 1}\right)^{\delta} t^{\delta} \left[1 - (1 - \delta)\left(\frac{\delta}{2\delta - 1}\right)^{(1-2\delta)/(1-\delta)} t^{(1-2\delta)/(1-\delta)}\{1 + o(1)\}\right], \quad t \to \infty.$$

Proof. The quantile function is obtained by solving $1 - F_X\{q(t)\} = t^{-1}$, which leads to

$$\frac{\delta}{2\delta - 1}q(t)^{-1/\delta}\left\{1 - \frac{1-\delta}{\delta}q(t)^{(1-2\delta)/\{\delta(1-\delta)\}}\right\} = t^{-1}$$

and thus

$$q(t)\left\{1 - \frac{1-\delta}{\delta}q(t)^{(1-2\delta)/\{\delta(1-\delta)\}}\right\}^{-\delta} = \left(\frac{\delta}{2\delta - 1}\right)^{\delta} t^{\delta}. \tag{23}$$

Since $q(t) \to \infty$ as $t \to \infty$, and $(1 - 2\delta)/\{\delta(1 - \delta)\} < 0$ for $\delta > 1/2$, expression (23) leads to

$$q(t) = \left(\frac{\delta}{2\delta - 1}\right)^{\delta} t^{\delta}\{1 + o(1)\}, t \to \infty,$$

which can be fed back into (23) to give the result claimed. $\square$

Proof of Proposition 2. When $\delta > 1/2$, the exponent function $V(x_1, \ldots, x_d)$ is obtained from the limit

$$V(x_1, \ldots, x_d) = \lim_{t \to \infty} t(1 - P[X_1 \le F_X^{-1}\{1 - (tx_1)^{-1}\}, \ldots, X_1 \le F_X^{-1}\{1 - (tx_d)^{-1}\}]).$$

Using Lemma 1, we have $q(tx) = \left(\frac{\delta}{2\delta-1}\right)^\delta (tx)^\delta \{1 + o(1)\}$, and so

$$
\begin{aligned}
1 - P\{X_1 \le q(tx_1), \ldots, X_1 \le q(tx_d)\} &= P\left\{\max_{j=1,\ldots,d} \frac{X_j}{q(tx_j)} > 1\right\} \\
&= P\left[\max_{j=1,\ldots,d} \frac{R^\delta W_j^{1-\delta}}{\left(\frac{\delta}{2\delta-1}\right)^\delta (tx_j)^\delta \{1 + o(1)\}} > 1\right] \\
&= \int_0^1 P\left[\max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{\left(\frac{\delta}{2\delta-1}\right) x_j \{1 + o(1)\}} > tu\right] du \\
&= \frac{1}{t} \int_0^t P\left[\max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{\left(\frac{\delta}{2\delta-1}\right) x_j \{1 + o(1)\}} > z\right] dz.
\end{aligned}
$$

For sufficiently large t, an integrable function of the form $P\left[K \max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{\left(\frac{\delta}{2\delta-1}\right) x_j} > z\right]$, $1 < K < \infty$, dominates the integrand over $(0, \infty)$ and thus the integral in the final line converges to

$$\int_0^\infty P\left\{\max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{\left(\frac{\delta}{2\delta-1}\right) x_j} > z\right\} dz = E\left\{\max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{x_j}\right\} \left(\frac{2\delta - 1}{\delta}\right),$$

and hence

$$
\begin{aligned}
\lim_{t \to \infty} t[1 - P\{X_1 \le q(tx_1), \ldots, X_d \le q(tx_d)\}] &= E\left\{\max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{x_j}\right\} \left(\frac{2\delta - 1}{\delta}\right) \\
&= E\left[\max_{j=1,\ldots,d} \frac{W_j^{(1-\delta)/\delta}}{E\{W_j^{(1-\delta)/\delta}\} x_j}\right],
\end{aligned}
$$

the final line following by equation (22). $\qquad\square$

Proof of Proposition 3. The function

$$\chi_u = P\{F_X(X_j) > u \mid F_X(X_k) > u\} = \frac{P\{X_j > F_X^{-1}(u), X_k > F_X^{-1}(u)\}}{1 - u}.$$

32

Lemma 1 gives the behavior of $F_X^{-1}(u) = q\{(1-u)^{-1}\}$, and combining this with Proposition 1 we have

$$\frac{P\{X_j > F_X^{-1}(u), X_k > F_X^{-1}(u)\}}{1-u} = E\{\min(W_j, W_k)^{(1-\delta)/\delta}\}\frac{F_X^{-1}(u)^{-1/\delta}}{1-u} + L\{F_X^{-1}(u)\}\frac{F_X^{-1}(u)^{-1/\{\eta_W(1-\delta)\}}}{1-u}$$

$$= \chi\left[1 + \frac{1-\delta}{\delta}\left(\frac{\delta}{2\delta-1}\right)^{(1-2\delta)/(1-\delta)}(1-u)^{(2\delta-1)/(1-\delta)}\{1+o(1)\}\right]$$

$$+ L\{(1-u)^{-1}\}(1-u)^{\delta/\{\eta_W(1-\delta)\}-1}\{1+o(1)\}.$$

Since $(2\delta-1)/(1-\delta) < \delta/\{\eta_W(1-\delta)\} - 1$ for $\eta_W < 1$, the result follows.

□

B   Supplementary Material
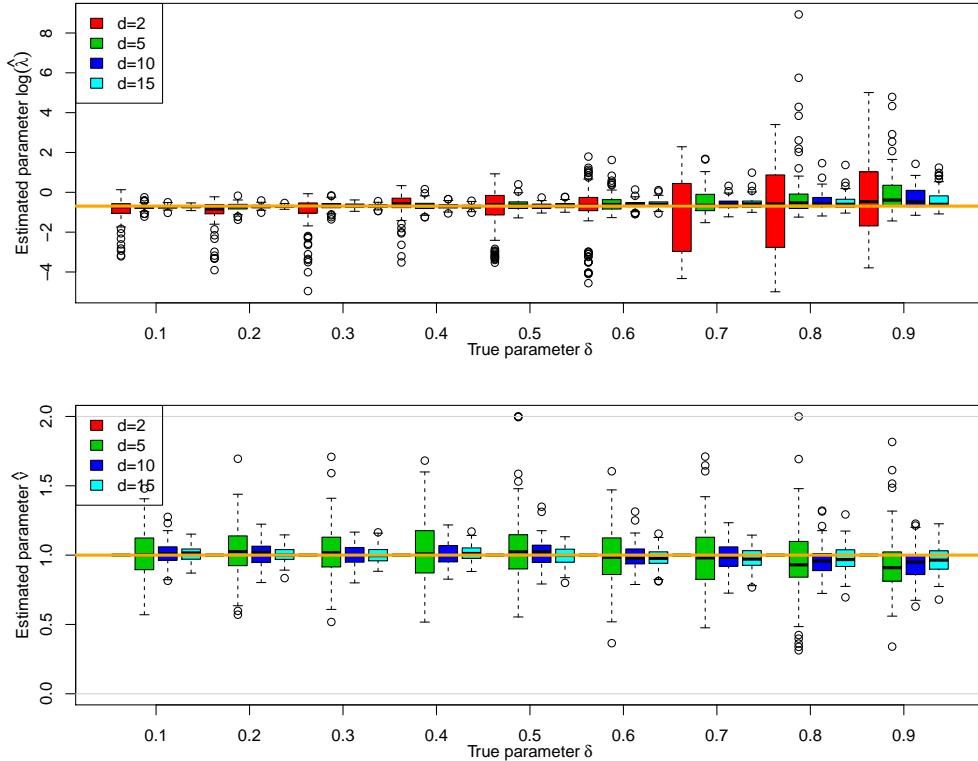
B.1   Supporting information for Section 3

Figure 7: Boxplots for the MLEs $\log(\hat{\lambda})$ and $\hat{\nu}$, estimated concurrently with $\hat{\delta}$ as in Figure 3 of Section 3.2.

B.2 Supporting information for Section 4

B.2.1 Bootstrap procedure

To demonstrate that the stationary bootstrap procedure described in § 4.1 adequately reproduces the temporal dependence in the extremes, we consider a spatial extension of the *extremal index* for univariate time series. For a stationary time series $\{X_t\}$, the extremal index, $\theta \in [0, 1]$, can be defined as

$$\theta = \lim_{n \to \infty} \mathrm{P}(X_2 \leq u_n, \ldots, X_{p_n} \leq u_n | X_1 > u_n),$$

where $p_n = o(n)$ and u_n is a series such that $n\{1 - F(u_n)\} \to \tau \in (0, \infty)$. The extremal index describes the degree of temporal clustering of extremes, with $1/\theta$ the limiting mean cluster size. A popular estimator for θ is the so-called Runs Estimator (Smith and Weissman, 1994). The estimate is formed by taking the reciprocal of the mean cluster size, whereby threshold exceedances are determined the be part of different clusters (the same cluster) if they are separated by a run of at least m (fewer than m) consecutive non-exceedances.

In our application we have a time series of spatial processes $\{X_t(s)\}$, which, as we consider winter months only, may reasonably be deemed stationary. In analogy to the univariate case, we define clusters of spatial threshold exceedances as follows. A realization of the process is deemed to be a "threshold exceedance" if the observation at any site exceeds a given threshold. Clusters are then defined as sequences of threshold exceedances separated by a run of at least m non-exceedances, and θ as the reciprocal mean cluster size. Figure 8 displays a histogram of estimated θs, using a value of $m = 1$, from 200 bootstrap samples, along with that from the original dataset of 50 sites temporally thinned to one observation per day. The threshold value used was the 95%-quantile, as in the model fit. The agreement between the original and bootstrap samples indicates that the temporal structure of the extremes is adequately reproduced.
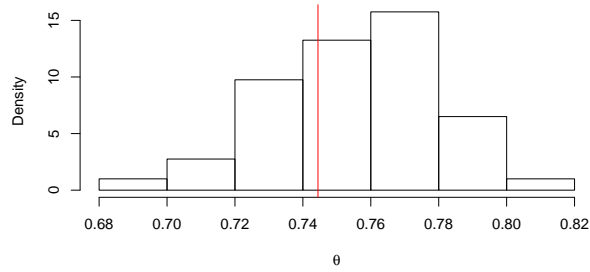
Figure 8: Estimates of the extremal index from the time series of spatial processes using the stationary bootstrap sampling procedure described in § 4.1. The vertical line is the value from the original sample.
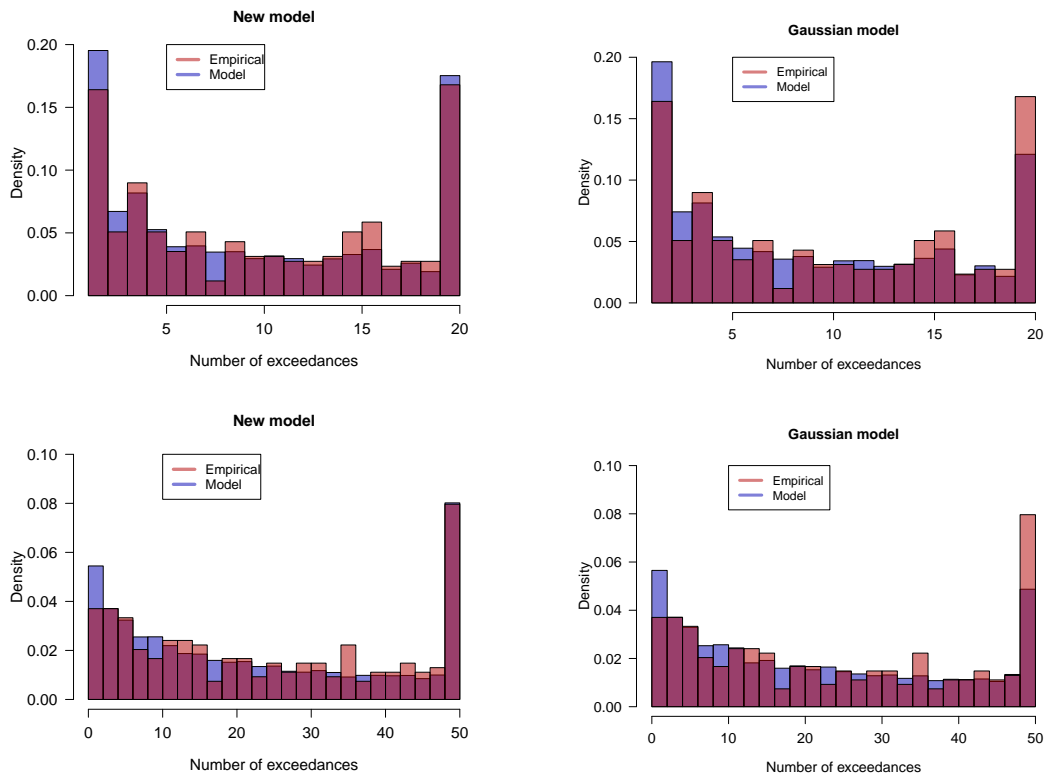
B.2.2 Additional model fit diagnostics

Figure 9: Distribution of the number of exceedances, given at least one exceedance of the 95%-quantile threshold. These histograms are based on the data at the 20 sites used for fitting the model (top) or all 50 sites (bottom). Model-based quantities are calculated for our new model (left) and the Gaussian model (right) by simulating 10^5 values from the fitted dependence models.

References

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726.

Castruccio, S., Huser, R., and Genton, M. G. (2016). High-order Composite Likelihood Inference for Max-Stable Distributions and Processes. *Journal of Computational and Graphical Statistics*. To appear.

Coles, S. G., Heffernan, J., and Tawn, J. A. (1999). Dependence Measures for Extreme Value Analyses. *Extremes*, 2(4):339–365.

Coles, S. G. and Pauli, F. (2002). Models and inference for uncertainty in extremal dependence. *Biometrika*, 89(1):183–196.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. B*, 53(2):377–392.

Cooley, D. S., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt-Shamseldin, E. C., Omolo, B. O., and Sun, Y. (2012). A Survey Of Spatial Extremes: Measuring Spatial Dependence And Modeling Spatial Effects. *REVSTAT*, 10(1):135–165.

Davison, A., Huser, R., and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci.*, 45:511–529.

Davison, A. C. and Huser, R. (2015). Statistics of Extremes. *Annual Review of Statistics and its Application*, 2:203–235.

Davison, A. C., Padoan, S., and Ribatet, M. (2012). Statistical Modelling of Spatial Extremes (with Discussion). *Statistical Science*, 27(2):161–186.

de Fondeville, R. and Davison, A. C. (2016). High-dimensional peaks-over-threshold inference for the Brown–Resnick process. Preprint ArXiv:1605.08558.

Dombry, C., Genton, M., Huser, R., and Ribatet, M. (2016). Full likelihood inference for max-stable data. Submitted.

Einmahl, J., Kiriliouk, A., Krajina, A., and Segers, J. (2016). An m-estimator of spatial tail dependence. *J. Roy. Statist. Soc. B*, 78:275–298.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler–
Reiss distributions and Brown–Resnick processes. *J. Roy. Statist. Soc. B*, 77:239–265.

Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards
application and simulation. *Bernoulli*, 20(4):1717–1737.

Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick
process. *Biometrika*, 100(2):511–518.

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

Huser, R., Opitz, T., and Thibaud, E. (2017). Bridging asymptotic independence and de-
pendence in spatial extremes using gaussian scale mixtures. Submitted.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum)
values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*,
81(348):158–171.

Joe, H. (2015). *Dependence Modeling with Copulas*. Chapman & Hall/CRC, Boca Raton,
FL.

Krupskii, P., Huser, R., and Genton, M. G. (2017). Factor copula models for replicated
spatial data. *J. Am. Stat. Assoc.* To appear.

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

Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on laplace
random fields. *Spatial Statistics*, 16:1–18.

Padoan, S., Ribatet, M., and Sisson, S. (2010). Likelihood-based inference for max-stable
processes. *J. Am. Stat. Assoc.*, 105:263–277.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *J. Am. Stat. Assoc.*,
89(428):1303–1313.

Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic
independence. *Extremes*, 5(4):303–336.

Resnick, S. I. (2006). *Heavy Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2017). Multivariate generalized Pareto distributions: parameterizations, representations, and properties. In preparation.

Segers, J. (2012). Max-stable models for multivariate extremes. *Revstat*, 10(1):61–82.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'institut de Statistique de l'Université de Paris*, 8:229–231.

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84:249–268.

Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *J. Roy. Statist. Soc. B*, pages 515–528.

Stephenson, A. and Tawn, J. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92(1):213–227.

Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J. (2016). Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. Preprint arXiv:1506.07836.

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

von Mises, R. (1954). La distribution de la plus grande de n valeurs. In *Selected Papers*, volume II, pages 271–294. American Mathematical Society, Providence, RI.

Wadsworth, J., Tawn, J., Davison, A., and Elton, D. (2017). Modelling across extremal dependence classes. *J. Roy. Statist. Soc. B*, 79:149–175.

Wadsworth, J. L. (2015). On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika*, 102(3):705–711.

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

Wadsworth, J. L. and Tawn, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15.