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

September 5, 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 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

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

has non-degenerate marginals, then $M(\boldsymbol{s})$ is a max-stable process (de Haan and Ferreira, 2006, Chapter 9). A practical issue with max-stable processes is that 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. As such, 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) and 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.

In order for limiting max-stable or generalized Pareto processes to provide good statistical models, we require that the extremes of $Y(\boldsymbol{s})$ are well represented by these processes, i.e., adequate convergence has occurred. However, there are no guarantees on rates of convergence, and in practice, limit models may not hold well. One way to assess the validity of convergence is to assess whether the stability properties of limit models hold well: max-stable copulas are invariant to the operation of taking pointwise maxima (max-stability),
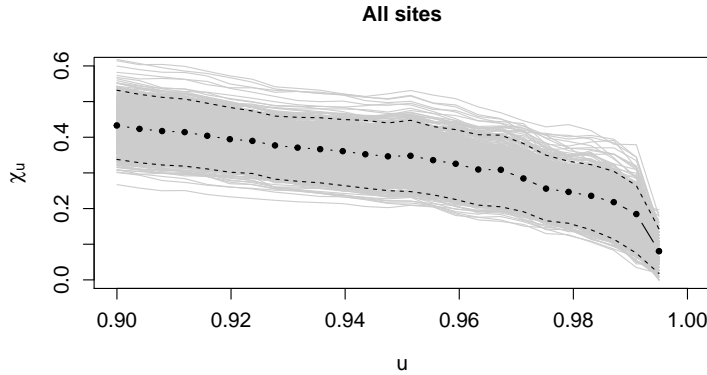
2

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.

whilst generalized Pareto copulas are invariant to conditioning on threshold exceedances of higher levels (threshold-stability). A graphical diagnostic for max-stability is presented in Gabda et al. (2012), whilst for threshold-stability, one can examine plots 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 follow a generalized Pareto process law, then this function should be constant as the quantile u tends to one (Rootzén et al., 2017). For environmental data in particular, it is much more common to see estimates of (2) decreasing as $u \to 1$, indicating that dependence weakens with level of extremeness. An example of this is given in Figure 1, for a dataset of significant wave heights, to be analyzed in §4.1.

If the limit of χ_u defined in (2) as $u \to 1$ is positive for all sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$ and all $d \geq 2$, the process $Y(\boldsymbol{s})$ is termed *asymptotically dependent*, and eventually, possibly at much higher levels, a generalized Pareto process should represent a suitable model for the data. If the limit is zero for all sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$ and all $d \geq 2$, we term the process *asymptotically independent*; in such cases no generalized Pareto model would ever be suitable. Intermediate scenarios

3

are possible, but owing to the structure of spatial data, it is common over small spatial domains to assume that the process is either asymptotically dependent or asymptotically independent, and we assume this here also. Determining a suitable model for the data usually requires distinguishing between these two scenarios, since most models exhibit only one type of dependence; choosing the incorrect class will lead to unsuitable extrapolation into the joint upper tail (Ledford and Tawn, 1997; Davison et al., 2013).

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 parametrized and requires pairwise 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 quantify our uncertainty about the dependence class 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, subject to model assumptions, 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 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 work 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 the new spatial model and 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, while 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*; a consequence of this is that for any $x \geq 1$,

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

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

where $L_W : (0, \infty) \to (0, \infty)$ is slowly varying at infinity, i.e., $L_W(ax)/L_W(x) \to 1$ as $x \to \infty$ for any $a > 0$, and $0 < \eta_W(\boldsymbol{h}) < 1$ for $\boldsymbol{h} = \boldsymbol{s}_j - \boldsymbol{s}_k \neq \boldsymbol{0}$ (Ledford and Tawn, 1996; Resnick, 2002). Note that we exclude the further possibility $\eta_W(\boldsymbol{h}) = 1$ ($\boldsymbol{h} \neq \boldsymbol{0}$), $L_W(x) \to 0$ as $x \to \infty$, because this does not arise in models that we might naturally consider for $W(\boldsymbol{s})$. The parameter $\eta_W(\boldsymbol{h})$, called the *coefficient of tail dependence*, summarizes the joint tail decay of the process $W(\boldsymbol{s})$ and it is a function of the lag vector $\boldsymbol{h}$. For simplicity, in what follows we will restrict ourselves to isotropic processes, and will therefore write $\eta_W(h)$ (or, for notational convenience, η_W, when no confusion can arise), where $h = \|\boldsymbol{h}\| = \|\boldsymbol{s}_j - \boldsymbol{s}_k\| \geq 0$ denotes the Euclidean distance between sites $\boldsymbol{s}_j, \boldsymbol{s}_k \in \mathcal{S}$. 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 defined through 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 potentially useful: when $\delta > 1/2$ then R^δ is heavier-tailed than $W^{1-\delta}$ and this induces

6

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. In practice, high-dimensional inference requires tractable densities for $W(\boldsymbol{s})$ (see §3.1), leading to the Gaussian copula as a natural choice in spatial settings. Alternative possibilities for $W(\boldsymbol{s})$ are discussed further in §2.5.

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 $\tilde{R} := \{\tilde{R}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2\}$, 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 $h = \|s_j - s_k\|$. A value of $\chi(h) > 0$ indicates asymptotic dependence for this pair of sites, whilst $\chi(h) = 0$ defines asymptotic independence. Because the process $X(s)$ has common margins with upper endpoint at infinity, the limit may be equivalently expressed as

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

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

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

9

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 $X(s)$ in (4), and then use this result in (7) and (8), combined with (9), to derive the corresponding coefficients χ and η_X characterizing tail dependence of $X(s)$ depending on the value of δ.

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

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

where L is slowly varying at infinity.

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*

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

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*

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

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).

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
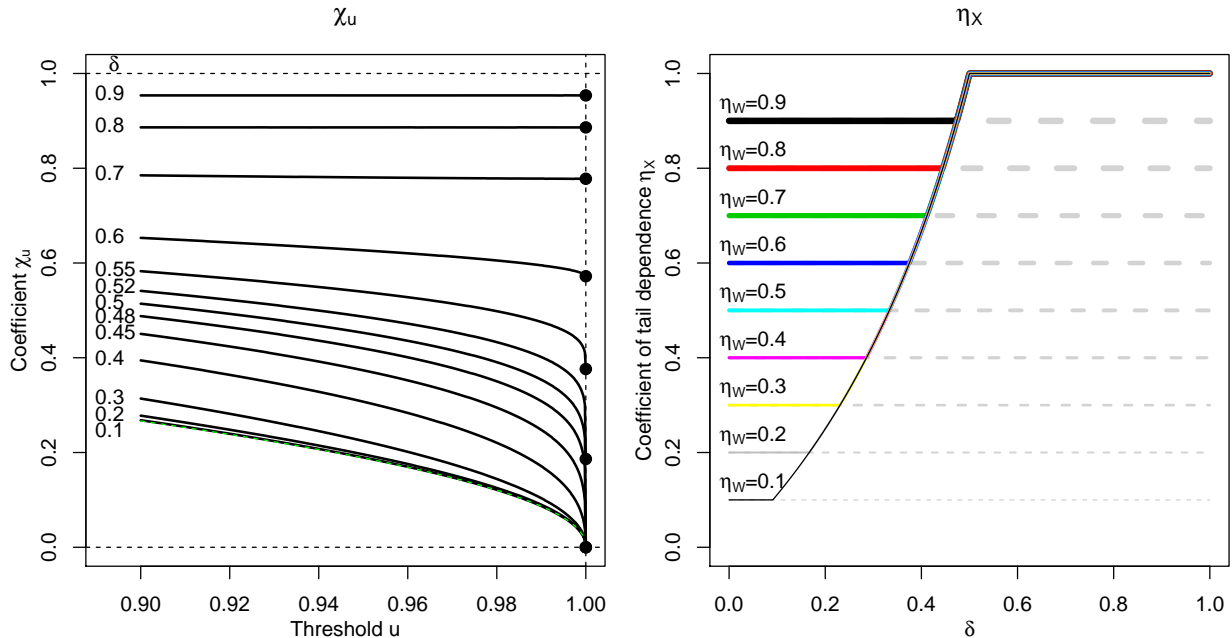
Figure 2: *Left:* Coefficient χ_u in (6) for model (4) plotted as a function of the threshold $u \in [0.9, 1]$ for $\delta = 0.1, 0.2, 0.3, 0.4, 0.45, 0.48, 0.5, 0.52, 0.55, 0.6, 0.7, 0.8, 0.9$ (top to bottom) and $(W_j, W_k)^T$ with a Gaussian copula with correlation 0.4. Black dots at $u = 1$ denote 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.

rapid for δ near 1, to rates slower than $(1 - u)$ for $1/2 < \delta < 2/3$. Note that for $\delta < 1/2$, the rate $\chi_u - \chi$ is determined 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 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.

13

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 (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 $\eta_W(h) = \{1 + \rho(h)\}/2$. In this case, the value of $\chi(h)$ in (10) needs to be calculated either by Monte Carlo or numerical integration, both of which are simple and quick.

Example 2 (Inverted max-stable process). Let $\{M(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ 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

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

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 $\eta_W^{1:d} \in [1/d, 1)$.

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. Pairwise or higher-dimensional composite likelihoods (see, e.g., Padoan et al., 2010; Varin et al., 2011; Castruccio et al., 2016) offer an alternative approach, but we do not explore this further here. 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. The latter show that the copula model defined by $\boldsymbol{X} = R\boldsymbol{W}$, where the radial variable R follows a unit scale generalized Pareto distribution with shape parameter $\xi \in \mathbb{R}$ and $\max(\boldsymbol{W}) = 1$, with R and $\boldsymbol{W}$ independent, displays asymptotic dependence for $\xi > 0$ and asymptotic independence for $\xi \leq 0$. 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 cap-

turing 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, i.e., for observations above a high marginal threshold, 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 as in Huser et al. (2017). 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

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

with contributions defined through the sets of indices $\mathcal{J}_i = \{j : U_{ij} > u_j^\star\} \subseteq \{1, \ldots, d\}$ as

$$L_i(\boldsymbol{\psi}) = \begin{cases} C(u_1^\star, \ldots, u_d^\star; \boldsymbol{\psi}), & \mathcal{J}_i = \emptyset, \\ c(U_{i1}, \ldots, U_{id}; \boldsymbol{\psi}), & \mathcal{J}_i = \{1, \ldots, d\}, \\ C_{\mathcal{J}_i}\{\max(U_{i1}, u_1^\star), \ldots, \max(U_{id}, u_d^\star); \boldsymbol{\psi}\}, & \text{otherwise.} \end{cases}$$

The set $\mathcal{J}_i$ determines whether the ith observation vector $(U_{i1}, \ldots, U_{id})^T$ has threshold exceedances in no, all, or some but not all components, respectively; therefore, these sets may be different for each likelihood contribution $i = 1, \ldots, n$. 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 $\mathcal{J}_i = \emptyset$ (i.e., the variable U_{ij} is lower than the threshold $u_j^\star$ for all $j = 1, \ldots, d$), 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 $\lambda = 0.5$
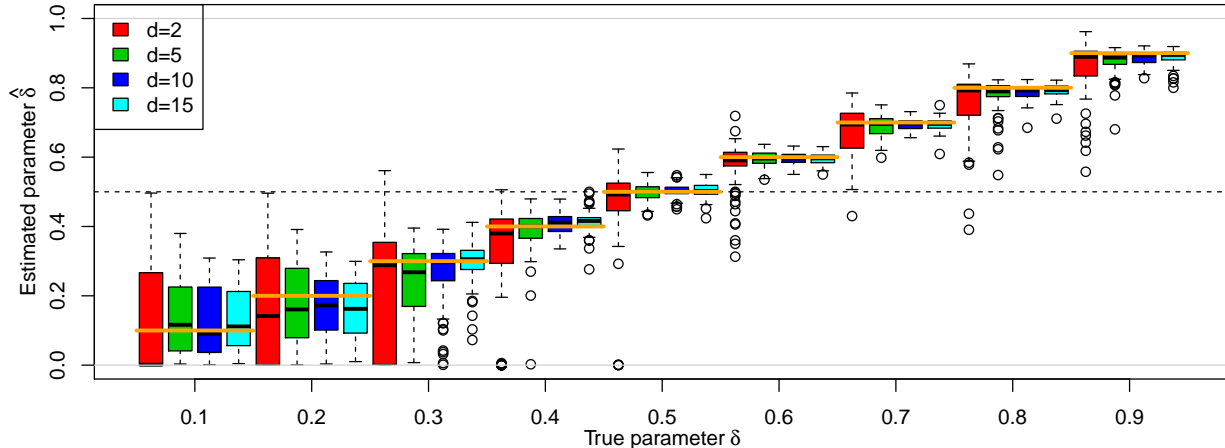
Figure 3: Boxplots for the MLE $\hat{\delta}$ based on (18) setting the thresholds as $u_j^\star = 0.95$, $j = 1, \ldots, d$, for model (4) with $\delta = 0.1, \ldots, 0.9$ (left to right). The underlying process $W(s)$ has a Gaussian copula with correlation function $\rho(s_j, s_k) = \exp\{-(\|s_j - s_k\|/\lambda)^\nu\}$ and $\lambda = 0.5$, $\nu = 1$. The process was observed at $d = 2$ (red, left), 5 (green, second left), 10 (dark blue, second right) and 15 (light blue, right) 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.

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. 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 model (4) over 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 almost useless in practice. To increase the power for discriminating between asymptotic dependence and independence, a parametric model-based approach seems sensible and our copula model (4) provides a very natural way to proceed, because the transition between the two asymptotic paradigms takes place in the interior of the parameter space. We stress, however, that the validity of such a test is reliant on modeling assumptions, and as such is

best used in conjunction with other diagnostics. Standard likelihood theory can be invoked to design tests for the null hypotheses

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

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

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

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

respectively, where $z_{1-\alpha}$ is the $(1 - \alpha)$-quantile of the standard normal distribution.

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 §3.2.1. 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 with higher dimensions, although there is little difference between $d = 5$ and $d = 10$. Comparing left and right panels, increased sample size also improves power, with a steeper transition around $\delta = 1/2$. The departure from nominal levels for the
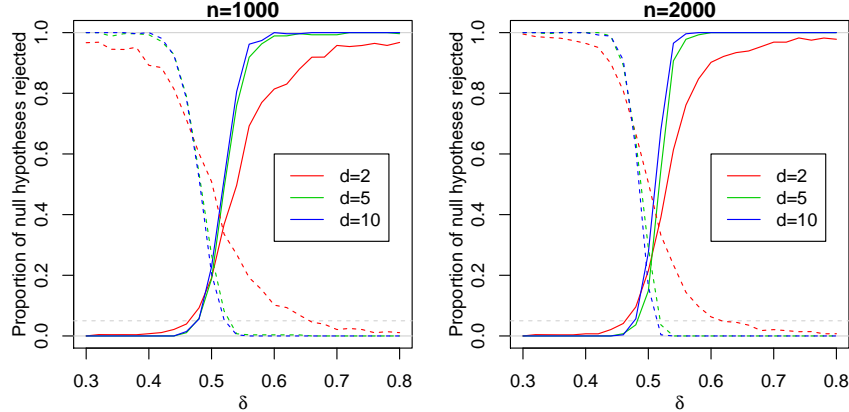
Figure 4: Proportion of times the null hypotheses H_0^{AD} (dashed) and H_0^{AI} (solid) are rejected, plotted against the true δ value, computed from 300 simulations. The horizontal dashed line at 0.05 shows the nominal level used for the tests. The model (4) was simulated at $d = 2$ (red), 5 (green), 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 $u_1^\star = \cdots = u_d^\star = 0.95$.

Type I error however suggests that the Hessian may not give a good representation of the asymptotic variance, possibly owing to numerical approximations. In Section 4.1 we suggest using bootstrap methods to calculate uncertainty.

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 energy, 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 ≈ 111km); the range of distances between sites is 0.27–2.99 units.

Model (4) was fitted by maximum likelihood based on (18) with thresholds $u_1^\star = \cdots = u_d^\star = 0.95$, assuming a Gaussian copula for the W process (Example 1); Table 1 reports the results. The uncertainty measures are based on 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 2 in the Supplementary Material shows 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

Table 1: MLEs, standard deviations (SD) of bootstrap replications, and approximate 95% confidence intervals (CI) for parameters of our model (4) (left) and the Gaussian process (right). Figures given to the longer of two decimal places or significant figures.

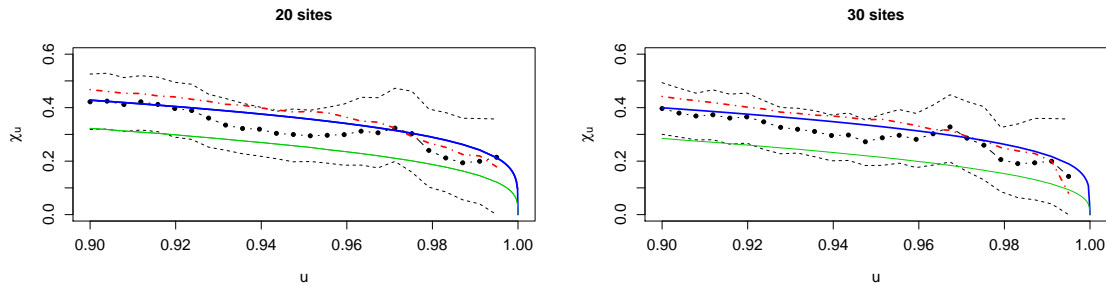|	MLE	SD	95% CI	MLE	SD	95% CI
δ	0.46	0.039	(0.36,0.54)	–	–	–
λ	3.19	0.26	(2.60,3.71)	3.84	0.17	(3.62,4.26)
ν	1.98	0.0033	(1.97,1.98)	1.97	0.0043	(1.96,1.98)

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).

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 $\delta = 0$. The maximized log-likelihood for our model was 62 units higher than for the Gaussian model, representing a clear improvement, although interpretation is difficult as 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 censored likelihood above a 95%-quantile threshold, the fit looks good on the plotted range $u \in (0.9, 1)$. The Gaussian model clearly underestimates the dependence.

The second diagnostic we consider is the distribution of the number of threshold ex-

ceedances, conditioning 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, 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 (recall Example 2). 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 with the approach of Wadsworth et al. (2017), we use the censored likelihood described therein and 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

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).

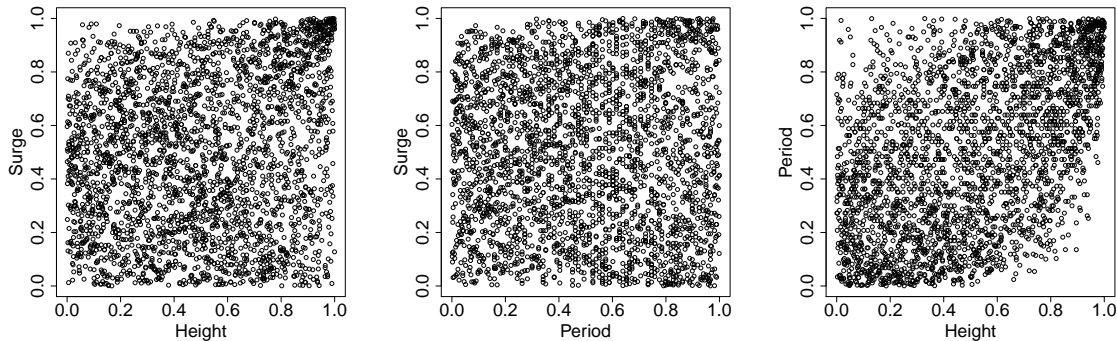|	Height–Surge	Period–Surge	Height–Period
AIC WTDE	264.1	515.4	225.7
AIC asymmetric	267.3	493.8	181.6

Figure 6: Newlyn wave data on approximate standard uniform margins. From left to right: Height-Surge, Period-Surge, Height-Period.

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.

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 rich 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; their code was used to speed up the bootstrap procedure in § 4.1. 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 $\delta > 1/3$, $\eta_X(h) > 1/2$ indicating a persistence of positive extremal association even as the lag $h \to \infty$. This is, however, a common problem with many 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 (10), $\chi(h) \geq (2\delta - 1)/\delta$. 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 in u but its limit $\chi(h)$ is small and positive. In practice however, this is not likely to be restrictive, since in our experience almost all environmental datasets display a decreasing χ_u function.

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.

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

$$= P(W_j > S, W_k > S),$$

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 $\eta_W > \delta/(1-\delta)$.

28

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

$$\int_1^{x^{1/(1-\delta)}} L_W(s)s^{(1-\delta)/\delta - 1/\eta_W - 1} \, ds = \int_1^\infty L_W(s)s^{(1-\delta)/\delta - 1/\eta_W - 1} \, ds - \int_{x^{1/(1-\delta)}}^\infty L_W(s)s^{(1-\delta)/\delta - 1/\eta_W - 1} \, ds, \tag{21}$$

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

$$E\{\min(W_j, W_k)^{(1-\delta)/\delta}\} = \int_0^\infty P\{\min(W_j, W_k)^{(1-\delta)/\delta} > t\} \, dt$$

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

Overall in Case 2 we thus have

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

$$= E\{\min(W_j, W_k)^{(1-\delta)/\delta}\}x^{-1/\delta}\{1 + o(1)\},$$

since $\eta_W(1-\delta) < \delta$. $\square$

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

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

If $\delta > 1/2$: Then $P(X_j > x) \sim \frac{\delta}{2\delta-1}x^{-1/\delta}$. We have $\delta/(1-\delta) > 1$ so we must be in Case 2, and

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

with expression (10) following as

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

29

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

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

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

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

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

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

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

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 above integral tends 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

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

Lemma 1 gives the behavior of $F_X^{-1}(u) = q\{(1 - u)^{-1}\}$, whilst the proof of Proposition 1 provides $\mathrm{P}(X_j > x, X_k > x) = \mathrm{E}\{\min(W_j, W_k)^{(1-\delta)/\delta}\}x^{-1/\delta} - L(x)x^{-1/\{\eta_W(1-\delta)\}}$, giving

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

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

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

with constant terms absorbed in to L. Since $(2\delta - 1)/(1 - \delta) < \delta/\{\eta_W(1 - \delta)\} - 1$ for $\eta_W < 1$, the result follows.

$\square$

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*, 25(4):1212–1229.

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(5):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.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.

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(1):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.

Gabda, D., Towe, R., Wadsworth, J., and Tawn, J. (2012). Discussion of "Statistical modeling of spatial extremes" by A.C. Davison, S.A. Padoan and M. Ribatet. *Statistical Science*, 27(2):189–192.

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 dependence in spatial extremes using gaussian scale mixtures. *Spatial Statistics*, 21(A):166–186.

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.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. B*, 59(2):475–499.

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(489):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. Preprint ArXiv:1705.07987.

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(2):249–268.

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

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

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. 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.