The Annals of Statistics
2021, Vol. 49, No. 5, 2552–2576
https://doi.org/10.1214/20-AOS2046

RANK-BASED ESTIMATION UNDER ASYMPTOTIC DEPENDENCE AND INDEPENDENCE, WITH APPLICATIONS TO SPATIAL EXTREMES

BY MICHAËL LALANCETTE[1,*], SEBASTIAN ENGELKE[2] AND STANISLAV VOLGUSHEV[1,†]

[1]*Department of Statistical Sciences, University of Toronto,* [*]*lalancette@utstat.toronto.edu;* [†]*stanislav.volgushev@utoronto.ca*
[2]*Research Center for Statistics, University of Geneva,* *sebastian.engelke@unige.ch*

Multivariate extreme value theory is concerned with modeling the joint tail behavior of several random variables. Existing work mostly focuses on asymptotic dependence, where the probability of observing a large value in one of the variables is of the same order as observing a large value in all variables simultaneously. However, there is growing evidence that asymptotic independence is equally important in real world applications. Available statistical methodology in the latter setting is scarce and not well understood theoretically. We revisit nonparametric estimation and introduce rank-based M-estimators for parametric models that simultaneously work under asymptotic dependence and asymptotic independence, without requiring prior knowledge on which of the two regimes applies. Asymptotic normality of the proposed estimators is established under weak regularity conditions. We further show how bivariate estimators can be leveraged to obtain parametric estimators in spatial tail models, and again provide a thorough theoretical justification for our approach.

1. Introduction. Assessing the frequency of extreme events is crucial in many different fields such as environmental sciences, finance and insurance. The most severe risks are often associated to a combination of extreme values of several different variables or the joint occurrence of an extreme phenomenon across different spatial locations. Statistical methods for accurate modeling of such multivariate or spatial dependencies between rare events is provided by extreme value theory. Applications include the analysis of extreme flooding (Asadi, Davison and Engelke (2015), Engelke and Hitz (2020), Keef, Tawn and Svensson (2009)), risk diversification between stock returns (Poon, Rockinger and Tawn (2004), Zhou (2010)) and climate extremes (Westra and Sisson (2011), Zscheischler and Seneviratne (2017)).

Extremal dependence between largest observations of two random variables X and Y with distribution functions F_1 and F_2, respectively, can take many different forms. A classical assumption to measure and model this dependence is multivariate regular variation (cf., Resnick (1987)), which is equivalent to the existence of the stable tail dependence function

$$\ell(x, y) := \lim_{t\downarrow 0} \frac{1}{t}\mathbb{P}\big(F_1(X) \geq 1 - tx \text{ or } F_2(Y) \geq 1 - ty\big), \quad x, y \in [0, \infty); \tag{1.1}$$

see Huang (1992) and de Haan and Ferreira (2006). This condition allows a first broad classification regarding extremal dependence of bivariate random vectors into two different regimes. If $\ell(x, y) = x + y$, X and Y are said to be asymptotically independent; in this case the joint exceedance probability is negligible compared to the marginal exceedance probabilities. Otherwise, a stronger form of extremal dependence, called asymptotic dependence,

Received August 2020; revised December 2020.
MSC2020 subject classifications. Primary 62G32; secondary 62F12, 62G20, 62G30.
Key words and phrases. Multivariate extremes, asymptotic independence, inverted max-stable distribution, spatial process, M-estimation.

holds and the joint exceedance probability is of the same order as the probability of one of the components exceeding a high threshold.

Most of the existing probabilistic and statistical theory deals with asymptotic dependence. A variety of methods exists, including nonparametric estimation (Einmahl and Segers (2009), Guillotte, Perron and Segers (2011), Huang (1992)), bootstrap procedures (Bücher and Dette (2013), Peng and Qi (2008)), parametric approaches including likelihood estimation (de Haan, Neves and Peng (2008), Dombry, Engelke and Oesting (2017), Ledford and Tawn (1996), Padoan, Ribatet and Sisson (2010)) and M-estimation (Einmahl, Krajina and Segers (2008), Engelke et al. (2015)). See also Einmahl, Krajina and Segers (2012), Einmahl et al. (2016) for inference in the d-dimensional and spatial setting. There is a rich literature on multivariate tail models (see, for instance, Gumbel (1960), Hüsler and Reiss (1989), Tawn (1988), among many others) and generalizations to spatial domains (Brown and Resnick (1977), Schlather (2002), Smith (1990)).

Recent studies have shown that in many applications such as spatial precipitation (Le et al. (2018)) and significant wave height (Wadsworth and Tawn (2012)), dependence tends to become weaker for the largest observations and asymptotic independence is therefore the more appropriate regime. In this case, the stable tail dependence function in (1.1) does not contain information on the degree of asymptotic independence and is therefore not suitable for statistical modeling. A remedy to this problem was proposed by Ledford and Tawn (1996), Ledford and Tawn (1997) who introduced a more flexible condition on the joint exceedance probabilities. Their setting implies the existence of

$$c(x, y) := \lim_{t \downarrow 0} \frac{1}{q(t)} \mathbb{P}\big(F_1(X) \geq 1 - tx, F_2(Y) \geq 1 - ty\big), \quad x, y \in [0, \infty), \tag{1.2}$$

where q is a suitable measurable function that makes the limit nontrivial. Necessarily, q is regularly varying at zero with index $1/\eta \in [1, \infty)$. The residual tail dependence coefficient η describes the strength of residual dependence in the tail and $\eta < 1$ implies asymptotic independence. One speaks about positive and negative association between extremes if $\eta > 1/2$ and $\eta < 1/2$, respectively. Early works focus on estimating the degree of asymptotic independence η and various estimators have been proposed and studied (Draisma et al. (2004), Ledford and Tawn (1997), Peng (1999)). A more complete description of the residual extremal dependence structure is given by the function c in equation (1.2); in fact, the value of η can be deduced from c (see Section 2 below). Several parametric families exist for multivariate (e.g., Weller and Cooley (2014)) and spatial applications (e.g., Wadsworth and Tawn (2012)). Other statistical approaches for modeling asymptotic independence are also related to this function, including hidden regular variation (Heffernan and Resnick (2007), Resnick (2002)) and the conditional extreme value model (Heffernan and Tawn (2004)). Note that equation (1.2) includes the asymptotic dependence case if $\lim_{t \downarrow 0} q(t)/t > 0$, and the function $c(x, y) \propto x + y - \ell(x, y)$ then contains the same information as ℓ.

Since it is typically not known a priori whether asymptotic dependence or independence is present in a data set, recent parametric models are able to capture both regimes as different sub-sets of the parameter space (e.g., Engelke, Opitz and Wadsworth (2019), Huser, Opitz and Thibaud (2017), Huser and Wadsworth (2019), Ramos and Ledford (2009), Wadsworth et al. (2017)). Most of the literature in this domain is concerned with constructing parametric models, and estimation is usually based on censored likelihood and discussed informally while no theoretical treatment of the corresponding estimators is provided. Moreover, it is typically assumed that extreme observations used to construct estimators already follow a limiting model, and the bias which results from this type of approximation is ignored.

The present paper is motivated by a lack of generic estimation methods that work under both asymptotic dependence and independence and have a thorough theoretical justification.

We first revisit a nonparametric, rank-based estimator of the function c in equation (1.2) which appeared in Draisma et al. (2004) and provide a new fundamental result on its asymptotic properties, which completely removes any smoothness assumptions on c. This result is the crucial building block for the second major contribution of this paper: a new M-estimation framework that is applicable across dependence classes.

M-estimators for the stable tail dependence function ℓ have been proposed by Einmahl, Krajina and Segers (2008, 2012), Einmahl et al. (2016). Under asymptotic dependence, c can be recovered from ℓ, and thus any method for estimating ℓ also yields an estimator for c. However, this is no longer true under asymptotic independence. Estimators of ℓ can therefore not be used to fit statistical models with asymptotic independence or models bridging both dependence classes. We define a new class of M-estimators based on c for parametric extreme value models that can be applied regardless of the dependence class. A major challenge under asymptotic independence is due to the fact that the scaling function q is unknown. Additionally, c loses some of the regularity (such as concavity) that it enjoys under asymptotic dependence. Nevertheless, we are able to prove asymptotic normality of our estimators under weak regularity conditions, which are shown to be satisfied for popular models such as the class of inverted max-stable distributions (see Wadsworth and Tawn (2012)).

The challenges described above become even greater for spatial data. Even at the level of pairwise distributions, real data can exhibit asymptotic dependence at locations that are close but asymptotic independence at locations that are far apart. This necessitates models that can incorporate both, asymptotic dependence and independence at the same time. Estimation in such models calls for methods that can deal with both regimes simultaneously, and we show that our findings in the bivariate case can be leveraged to construct estimators in this setting.

In Section 2, we provide the necessary background on asymptotic dependence and independence for bivariate distributions, discuss an extension to the spatial setting and provide several examples. The estimation methodology is introduced in Section 3, while theoretical results are collected in Section 4. The methodology is illustrated via simulation studies in Section 5, while an application to extreme rainfall data is given in Section 6. All proofs are deferred to the Supplementary Material (Lalancette, Engelke and Volgushev (2021)). All references to sections, results and equations starting with the letter "S" are pointing to this online supplement. The R code can be found at https://github.com/mic-lalancette/rank-based.

2. Multivariate extreme value theory.

2.1. *Bivariate models.* Let (X, Y) be a bivariate random vector with joint distribution function F and marginal distribution functions F_1 and F_2, respectively. There is a variety of approaches to describe the joint tail behavior of (X, Y).

The assumption of multivariate regular variation (cf., Resnick (1987)) is classical in extreme value theory and the stable tail dependence function in (1.1) has been extensively studied. Its margins are normalized, $\ell(x, 0) = \ell(0, x) = x$, and it satisfies $x \vee y \le l(x, y) \le x + y$ for all $x, y \in [0, \infty)$. Moreover, it is a convex and homogeneous function of order one, the latter meaning that $\ell(tx, ty) = t\ell(x, y)$ for all $t > 0$. The importance of stable tail dependence functions stems from their connection to max-stable distributions. A bivariate random vector (Z_1, Z_2) has max-stable dependence with standard uniform margins iff its distribution function is given by

$$\mathbb{P}(Z_1 \le x, Z_2 \le y) = \exp\{-\ell(-\log x, -\log y)\}, \quad x, y \in [0, 1], \tag{2.1}$$

where ℓ is the stable tail dependence function pertaining to (Z_1, Z_2). Note that any max-stable distribution associated with ℓ satisfies equation (1.1) with that same ℓ, this follows after a simple Taylor expansion. Two examples of max-stable distributions (equivalently, stable tail dependence functions) that will repeatedly appear in the present paper are as follows.

(i) The bivariate Hüsler–Reiss distribution (Engelke et al. (2015), Hüsler and Reiss (1989)) is defined by

$$\ell(x, y) = x\Phi\left(\lambda + \frac{\log x - \log y}{2\lambda}\right) + y\Phi\left(\lambda + \frac{\log y - \log x}{2\lambda}\right),$$

where Φ is the standard normal distribution function and $\lambda \in [0, \infty]$ parametrizes between perfect independence ($\lambda = \infty$) and dependence ($\lambda = 0$).

(ii) The asymmetric logistic distribution (Tawn (1988)) is given by

$$\ell(x, y) = (1 - \nu)x + (1 - \phi)y + \left(\nu^r x^r + \phi^r y^r\right)^{1/r}, \quad \nu, \phi \in [0, 1], r \geq 1.$$

Note that $\nu = \phi = 1$ yields the classical logistic model (Gumbel (1960)).

While multivariate regular variation and max-stability have been very popular due to their nice theoretical properties, they are not informative under asymptotic independence, which limits their use in many applications.

Assumption (1.2) allows for more flexible tail models since the limiting function c is nontrivial even under asymptotic independence and contains information on the structure of residual extremal dependence in the vector (X, Y). For the sake of identifiability, we scale q such that $c(1, 1) = 1$. We will refer to c and q as the survival tail function and the scaling function associated to (X, Y). It turns out that q has to be regularly varying of order $1/\eta \in [1, \infty)$ and that c is a homogenous funcion of order $1/\eta$, that is,

$$c(tx, ty) = t^{1/\eta} c(x, y), \quad t > 0;$$

see, for example, Draisma et al. (2004) or Lemma S2 in the online supplement. Note that the extremal dependence coefficient (see Coles, Heffernan and Tawn (1999)) can be defined as $\chi := \lim_{t\downarrow 0} q(t)/t$. Asymptotic independence is then equivalent to $\chi = 0$, or $q(t) = o(t)$, whereas asymptotic dependence corresponds to $\chi > 0$.

Furthermore, the homogeneity property of c implies a spectral representation. More precisely, there exists a finite measure H on $[0, 1]$ such that

$$c(x, y) = \int_{[0,1]} \left(\frac{x}{1 - w} \wedge \frac{y}{w}\right)^{1/\eta} H(dw), \quad x, y \in [0, \infty);$$

see Theorem 1 in Ramos and Ledford (2009) or Lemma S6 in the online supplement.

We provide several examples that illustrate the concepts discussed above without going too deeply into technical details. A more thorough discussion of the corresponding theory is given throughout Section 4.

EXAMPLE 1 (Domain of attraction of max-stable distributions). Suppose that (X, Y) satisfies equation (1.1) for a stable tail dependence function ℓ which is not everywhere equal to $x + y$. Then equation (1.2) holds with $q(t) = \chi t$ and $c(x, y) = (x + y - \ell(x, y))/\chi$, where the extremal dependence coefficient χ is positive. We further note that equation (1.1) holds whenever (X, Y) is in the max domain of attraction of a max-stable random vector Z satisfying equation (2.1); see de Haan and Ferreira (2006) for a definition and additional details.

EXAMPLE 2 (Inverted max-stable distributions). The family of inverted max-stable distributions (e.g., Wadsworth and Tawn (2012), Definition 2) is parametrized by all stable tail dependence functions. More precisely, let G be the distribution function of a bivariate distribution with max-stable dependence, uniform margins and stable tail dependence function ℓ. A random vector (X, Y) with uniform marginal distributions is said to have an inverted max-stable distribution with stable tail dependence ℓ if $(1 - X, 1 - Y) \sim G$. Assuming that ℓ

satisfies a quadratic expansion (see Example 8), the law of (X, Y) satisfies equation (1.2) with

$$q(t) = t^{\ell(1,1)}, \qquad c(x, y) = x^{\dot{\ell}_1(1,1)} y^{\dot{\ell}_2(1,1)},$$

where $\dot{\ell}_j$ denotes the jth directional partial derivative of ℓ from the right, $j = 1, 2$. Any such stable tail dependence function satisfies $\ell(1, 1) = \dot{\ell}_1(1, 1) + \dot{\ell}_2(1, 1) \in (1, 2]$ and, therefore, this is an asymptotically independent model with $\eta = 1/\ell(1, 1)$.

Any existing parametric class of stable tail dependence functions can be used to define a parametric class of inverted max-stable distributions. In particular, we consider the two families discussed earlier:

(i) Provided that $\lambda > 0$, the inverted Hüsler–Reiss distribution has

$$q(t) = t^{2\theta}, \qquad c(x, y) = (xy)^{\theta}, \tag{2.2}$$

where $\theta := \Phi(\lambda) \in (1/2, 1]$.

(ii) The inverted asymmetric logistic distribution has

$$q(t) = t^{\theta_1+\theta_2}, \qquad c(x, y) = x^{\theta_1} y^{\theta_2}, \tag{2.3}$$

where $\theta_1 := 1 - \nu + \nu^r(\nu^r + \phi^r)^{1/r-1}$ and $\theta_2 := 1 - \phi + \phi^r(\nu^r + \phi^r)^{1/r-1}$. Note that by suitable choices of the parameters r, ν, ϕ any value of $(\theta_1, \theta_2) \in (0, 1]^2$ such that $\theta_1 + \theta_2 \in (1, 2]$ can be obtained.

EXAMPLE 3 (A random scale construction). Bivariate random scale constructions are a popular way of creating distributions with rich extremal dependence structures; see Engelke, Opitz and Wadsworth (2019) and references therein for an overview. They are random vectors of the form $(X, Y) = R(W_1, W_2)$ where the radial variable R is assumed independent of the angular variables W_j, $j \in \{1, 2\}$. This motivates the following model with parameters $\alpha_R, \alpha_W > 0$:

$$(X, Y) = R(W_1, W_2), \quad R \sim \text{Pareto}(\alpha_R), W_j \sim \text{Pareto}(\alpha_W), \tag{2.4}$$

where W_1, W_2 are independent and a Pareto(α) distribution has distribution function $1 - x^{-\alpha}$ for $x \geq 1$. By Example 9 below, (X, Y) satisfies equation (1.2) with functions q and c depending only on the value of the ratio $\lambda := \alpha_R/\alpha_W$. In particular, we obtain asymptotic dependence if $\lambda < 1$ and asymptotic independence otherwise. Detailed expressions for q and c are provided in Example 9.

2.2. *Spatial models*. Spatial extreme value theory is an extension of multivariate extremes to continuous index sets. It is particularly useful for modeling extremes of natural phenomena over spatial domains and examples include heavy rainfall, high wind speeds and heatwaves (e.g., Davison and Gholamrezaee (2012), Le et al. (2018)).

Let $\mathcal{T}$ be a spatial domain (typically a subset of $\mathbb{R}^2$) and $Y = \{Y(u) : u \in \mathcal{T}\}$ be a stochastic process whose extremal behavior we are interested in. We impose the condition in equation (1.2) on all bivariate margins of Y so that for each pair $s = (u, u')$ of locations, and all $x, y \in [0, \infty)$ the limit

$$c^{(s)}(x, y) := \lim_{t \downarrow 0} \frac{1}{q^{(s)}(t)} \mathbb{P}\big(F^{(u)}(Y(u)) \geq 1 - tx, F^{(u')}(Y(u')) \geq 1 - ty\big) \tag{2.5}$$

exists and is nontrivial; here $F^{(u)}$ is the distribution function of $Y(u)$. Similar to the bivariate case, $q^{(s)}$ must be regularly varying with index $1/\eta^{(s)} \in [1, \infty)$ and $c^{(s)}$ is homogeneous of order $1/\eta^{(s)}$.

In applications, spatial extreme value theory can be linked to multivariate extreme value theory through the fact that spatial processes are usually measured at a finite set of locations. However, generic multivariate models do not take into account the additional structure arising from spatial features of the domain. Statistical models for processes, in contrast, make use of geographical information to construct structured, low-dimensional parametric models (see, e.g., Engelke and Ivanovs (2021)).

Similarly to max-stable distributions in equation (2.1), max-stable processes play an important role in modeling spatial extremes. The stochastic process $Z = \{Z(u) : u \in \mathcal{T}\}$ is called max-stable if all its finite dimensional distributions are max-stable, which implies in particular that for each pair $s = (u, u')$, the bivariate margin $(Z(u), Z(u'))$ satisfies equation (2.1) with stable tail dependence function $\ell^{(s)}$. Hence equation (2.5) follows for any max-stable process Z for which $(Z(u), Z(u'))$ are not independent for all $u, u' \in \mathcal{T}$.

Brown–Resnick processes (Brown and Resnick (1977)) provide an important subclass of max–stable processes. A Brown–Resnick process $\mathcal{B} = \{\mathcal{B}(u) : u \in \mathcal{T}\}$ is parametrized by a variogram function $\gamma : \mathcal{T}^2 \to \mathbb{R}_+$, and any pair $(\mathcal{B}(u), \mathcal{B}(u'))$ is a bivariate Hüsler–Reiss distribution with parameter $\lambda = \sqrt{\gamma(u, u')}/2$ (Hüsler and Reiss (1989)). Parametric models can be constructed by imposing a parametric form for γ. An example when $\mathcal{T} \subset \mathbb{R}^d$ is the fractal family of variograms given by $\gamma(s) = (\|s_1 - s_2\|/\beta)^\alpha$, where $s = (s_1, s_2)$, $\|\cdot\|$ is the Euclidean norm and $\alpha \in (0, 2]$, $\beta > 0$ are the model parameters (Kabluchko, Schlather and de Haan (2009)). We next discuss two classes of processes for which equation (2.5) holds.

EXAMPLE 4 (Domain of attraction of max-stable processes). A process $Y = \{Y(u) : u \in \mathcal{T}\}$ is in the max-domain of attraction of the max-stable process Z if there exist sequences of continuous functions $a_n, b_n : \mathcal{T} \to \mathbb{R}$ such that

$$\Big\{\max_{i=1,\dots,n} Y_i(\cdot) - a_n(\cdot)\Big\}/b_n(\cdot) \rightsquigarrow Z(\cdot), \quad n \to \infty \tag{2.6}$$

for i.i.d. copies $Y_1, Y_2, \dots$ of the process Y where weak convergence takes place in the space of continuous functions on $\mathcal{T}$ equipped with the supremum norm; see de Haan and Lin (2001) and Chapter 9 of de Haan and Ferreira (2006) for the special case $\mathcal{T} = [0, 1]$.

Equation (2.6) implies that any pair $(Y(u), Y(u'))$ with $u \neq u' \in \mathcal{T}$ is in the max-domain of attraction of the pair $(Z(u), Z(u'))$. If every such pair is not independent, equation (2.5) holds for all $s = (u, u')$ by the discussion in Example 1.

While max-stable processes allow for flexible spatial dependence structures, they can only be used as models for asymptotic dependence. This often violates the characteristics observed in real data, especially for locations $u, u' \in \mathcal{T}$ that are far apart. To model data in such cases, asymptotically independent spatial models have been constructed that satisfy equation (2.5) and where the residual tail dependence coefficients $\eta^{(s)}$ vary with the distance between the pair s of locations. Most of the models are identifiable from the bivariate margins so that statistical methods for $c^{(s)}$ will provide estimators for spatial tail dependence parameters; see Section 3.3 for the methodology. A broad class of asymptotically independent stochastic processes are the inverted max-stable processes (Wadsworth and Tawn (2012)).

EXAMPLE 5 (Inverted max-stable processes). Let $Z = \{Z(u) : u \in \mathcal{T}\}$ be a process with max-stable dependence, uniform margins and bivariate tail dependence functions $\ell^{(s)}$. The process $Y = \{1 - Z(u) : u \in \mathcal{T}\}$ is called inverted max-stable. For a pair $s \in \mathcal{T}^2$, assuming that $\ell^{(s)}$ satisfies the smoothness condition mentioned in Example 2, Y satisfies equation (2.5) with

$$q^{(s)}(t) = t^{\ell^{(s)}(1,1)}, \qquad c^{(s)}(x, y) = x^{\dot{\ell}_1^{(s)}(1,1)} y^{\dot{\ell}_2^{(s)}(1,1)},$$

so that $\eta^{(s)} = 1/\ell^{(s)}(1, 1)$ is a (usually nonconstant) function on $\mathcal{T}^2$. In particular, if a Brown–Resnick process is parametrized by a variogram function $\gamma : \mathcal{T}^2 \to \mathbb{R}_+$ then the corresponding inverted Brown–Resnick process has $1/\eta^{(s)} = 2\Phi(\sqrt{\gamma(s)}/2)$.

3. Estimation. In this section, we present the proposed estimators. First, we recall the nonparametric estimator of a survival tail function from Draisma et al. (2004) in Section 3.1. Using this as building block, we construct M-estimators for bivariate survival tail functions (Section 3.2) and leverage those estimators to introduce methodology for spatial tail estimation (Section 3.3).

3.1. *Nonparametric estimators of survival tail functions.* Recall that (X, Y) is a random vector with joint distribution function F that satisfies equation (1.2), and assume that its marginal distribution functions F_1 and F_2 are continuous. Denoting by Q the joint distribution function of $(1 - F_1(X), 1 - F_2(Y))$, we can rephrase equation (1.2) as

$$\frac{Q(tx, ty)}{q(t)} = c(x, y) + O\big(q_1(t)\big), \quad x, y \in [0, \infty), \tag{3.1}$$

for some function $q_1(t) \to 0$ as $t \to 0$. Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent samples from F. Since F_1, F_2 are unknown, the observations $(1 - F_1(X_i), 1 - F_2(Y_i))$ are not available and cannot be used to construct a feasible estimator of Q. A typical solution to this problem is to replace F_j by its empirical counterpart $\widehat{F}_j$, which leads to the estimator

$$\widehat{Q}_n(x, y) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{n\widehat{F}_1(X_i) \geq n + 1 - \lfloor nx \rfloor, n\widehat{F}_2(Y_i) \geq n + 1 - \lfloor ny \rfloor\}; \tag{3.2}$$

see Drees and Huang (1998), Huang (1992), Einmahl, Krajina and Segers (2008, 2012) among others for related approaches in the estimation of stable tail dependence functions.

Given $\widehat{Q}_n$ and the expansion in equation (3.1), an intuitive plug-in estimator of the function c is given by

$$\widehat{c}_n(x, y) = \frac{\widehat{Q}_n(kx/n, ky/n)}{q(k/n)}, \tag{3.3}$$

where we set $t = k/n$ in equation (3.1) for an intermediate sequence $k = k_n$ such that $k \to \infty$, $k/n \to 0$. Note, however, that this estimator is infeasible under asymptotic independence since the function q is in general unknown. A simple remedy is to recall that we considered the normalization $c(1, 1) = 1$ and construct a ratio type estimator

$$\widetilde{c}_n(x, y) := \frac{\widehat{c}_n(x, y)}{\widehat{c}_n(1, 1)} = \frac{\widehat{Q}_n(kx/n, ky/n)}{\widehat{Q}_n(k/n, k/n)} \tag{3.4}$$

to cancel out the unknown scaling factor $q(k/n)$. This leads to a fully nonparametric estimator of c, which is interesting in its own right. Some comments on the asymptotic properties of this estimator will be provided in Section 4.1.1.

REMARK 1. In practice, and especially in a spatial context, it is sometimes appropriate to select directly the effective number of observations used for estimating c (Wadsworth and Tawn (2012)). This can be achieved by selecting $k = \widehat{k}$ such that $nQ_n(\widehat{k}/n, \widehat{k}/n) = m$ for a given value m. This leads to a data-dependent parameter $\widehat{k}$ which will also be covered by our theory.

3.2. *M-estimation in (bivariate) parametric model classes.* While the nonparametric estimators from the previous section possess attractive theoretical properties, they have certain practical drawbacks. For any finite sample size n they are neither continuous nor homogeneous, hence they are not admissible survival tail functions. Additionally, it is difficult to use purely nonparametric estimators in spatial settings. A solution to this problem, which also yields easily interpretable estimators, is to fit parametric models.

In what follows, assume that c belongs to a family $\{c_\theta : \theta \in \Theta\}$, where $\Theta \subseteq \mathbb{R}^p$ and the true parameter $\theta_0 \in \Theta$ is unknown. Our aim is to leverage the nonparametric estimators from Section 3.1 to construct an estimator for θ_0. For stable tail dependence functions which are only informative under asymptotic dependence such a program was carried out in Einmahl, Krajina and Segers (2008), Einmahl, Krajina and Segers (2012). A direct application of the corresponding ideas in our setting would be to estimate θ through

$$\breve{\theta} := \operatorname*{arg\,min}_{\theta\in\Theta} \left\| \int_{[0,T]^2} g(x,y)c_\theta(x,y)\,dx\,dy - \int_{[0,T]^2} g(x,y)\widetilde{c}_n(x,y)\,dx\,dy \right\|,$$

for an integrable vector-valued weight function $g : \mathbb{R}^2 \to \mathbb{R}^q$, where $\|\cdot\|$ denotes the Euclidean norm. However, as we will discuss in Remark 5, the use of $\widetilde{c}_n$ would place unnecessarily strong assumptions on the function c in the case of asymptotic dependence. Hence we propose to consider the following alternative approach. Define

$$\Psi_n^*(\theta,\zeta) := \zeta \int_{[0,T]^2} g(x,y)c_\theta(x,y)\,dx\,dy - \int_{[0,T]^2} g(x,y)\widehat{Q}_n(kx/n,ky/n)\,dx\,dy \tag{3.5}$$

and let

$$(\widehat{\theta}_n,\widehat{\zeta}_n) := \operatorname*{arg\,min}_{\theta\in\Theta,\zeta>0} \left\|\Psi_n^*(\theta,\zeta)\right\|. \tag{3.6}$$

To understand the rationale of this approach, note that $\widehat{c}_n(x,y)$ is proportional to $\widehat{Q}_n(kx/n,ky/n)$ but the proportionality constant involves q and is thus unknown. We thus essentially propose to add this unknown normalization factor as an additional scale parameter ζ. More precisely, write μ_{L} for the Lebesgue measure on $[0,T]^2$, let

$$\Psi_n(\theta,\sigma) = \sigma \int g c_\theta \, d\mu_{\mathrm{L}} - \int g\widehat{c}_n \, d\mu_{\mathrm{L}}$$

and note that Ψ_n^* and Ψ_n are linked through $\Psi_n^*(\theta,\zeta) = q(k/n)\Psi_n(\theta,\zeta/q(k/n))$. Thus $(\widehat{\theta}_n,\widehat{\zeta}_n)$ minimizes $\|\Psi_n^*\|$ if and only if $(\widehat{\theta}_n,\widehat{\zeta}_n/q(k/n))$ minimizes $\|\Psi_n\|$. Furthermore, under suitable assumptions on g and Θ we have $\sigma\int g c_\theta \, d\mu_{\mathrm{L}} = \int g c_{\theta_0}\, d\mu_{\mathrm{L}}$ if and only if $\theta = \theta_0$ and $\sigma = 1$. Hence, if $\widehat{c}_n$ is close to c_{θ_0}, it is expected that $\widehat{\theta}_n$ will be close to θ_0 and that $\widehat{\zeta}_n/q(k/n)$ will be approximately 1.

Note that equation (3.5) only involves an integral of $\widehat{Q}_n$ while $\widetilde{c}_n$ also involves pointwise evaluation of this function. Since integration acts as smoothing, it can be expected that studying Ψ_n^* will require less regularity conditions than working with $\breve{\theta}$; see Remark 5 for additional details.

3.3. *Parametric estimation for spatial tail models.* In this section, we show how the bivariate estimation procedures discussed earlier can be leveraged to obtain two different estimators for parametric spatial models, which can include both asymptotic dependence and independence. Assume that we observe n independent copies $Y_1,\dots,Y_n$ of a spatial process Y at a finite set of locations $u_1,\dots,u_d \in \mathcal{T}$. Denote the corresponding observations by $X_1,\dots,X_n$ where $X_i = (X_i^{(1)},\dots,X_i^{(d)}) := (Y_i(u_1),\dots,Y_i(u_d))$ are independent copies of the random vector $X = (X^{(1)},\dots,X^{(d)}) := (Y(u_1),\dots,Y(u_d)) \in \mathbb{R}^d$; see Einmahl et al. (2016) for a similar framework.

Let $\mathcal{P}$ denote the set of all subsets of $\{1, \dots, d\}$ of size 2 interpreted as ordered pairs, so that elements of $\mathcal{P}$ will take the form $s = (s_1, s_2)$ with $s_1 < s_2$. In what follows, we will need to repeatedly make use of vectors $x \in \mathbb{R}^{|\mathcal{P}|}$ that are indexed by all pairs $s \in \mathcal{P}$. For such vectors, we will assume that the pairs in $\mathcal{P}$ are ordered in a prespecified order and will write $x^{(s)}$ for the entry of the vector x that corresponds to pair s.

Assume that for each pair s the random vector $(X^{(s_1)}, X^{(s_2)})$ satisfies equation (3.1) with scale function $q^{(s)}$ and survival tail function $c^{(s)}$. Following the ideas laid out in Section 3.1, define $\widehat{Q}_n^{(s)}$ as in equation (3.2) but based on the bivariate observations $(X_i^{(s_1)}, X_i^{(s_2)})$, $i = 1, \dots, n$. We now discuss two parametric estimators for the functions $c^{(s)}$.

Assume that we start with a parametric model $\{c_\theta : \theta \in \widetilde{\Theta}\}$, $\widetilde{\Theta} \subseteq \mathbb{R}^{\widetilde{p}}$, for bivariate survival tail functions and that each $c^{(s)}$ falls in this class. This implies that $\widetilde{\Theta}$ can be linked to a spatial parameter space $\Theta \subseteq \mathbb{R}^p$ through the relations $c^{(s)} = c_{h^{(s)}(\vartheta_0)}$, where $h^{(s)} : \Theta \to \widetilde{\Theta}$ for each pair s. To make this idea more concrete, consider the following example, which we will revisit in Sections 5.2 and 6.

EXAMPLE 6. If the process Y is an inverted Brown–Resnick process on $\mathbb{R}^2$ (see Example 5), then X has an inverted Hüsler–Reiss distribution and the bivariate survival tail functions are of the form $c^{(s)}(x, y) = (xy)^{\theta^{(s)}}$, for some $\theta^{(s)} \in (1/2, 1)$. This determines the parametric class $\widetilde{\Theta}$. A more specific model of Brown–Resnick processes corresponds to the subfamily of fractal variograms (Engelke et al. (2015), Kabluchko, Schlather and de Haan (2009)), where

$$\theta^{(s)} = h^{(s)}\big((\alpha, \beta)\big) = \Phi\left(\frac{(\|u_{s_1} - u_{s_2}\|/\beta)^{\alpha/2}}{2}\right), \quad s \in \mathcal{P}, \tag{3.7}$$

where $u_j \in \mathbb{R}^2$ is the coordinate of the location j; see Section 6 for more motivation of this particular parametrization. The global parameter ϑ thus takes the form $\vartheta = (\alpha, \beta)$ and $\Theta = (0, 2] \times (0, \infty)$.

Given the setting above, we can thus compute parametric estimators $\widehat{\theta}_n^{(s)}$, $s \in \mathcal{P}$, by the methods for bivariate estimation discussed in Section 3.2, that is, $(\widehat{\theta}_n^{(s)}, \widehat{\zeta}_n^{(s)})$ is the minimizer of $\|\Psi_n^{*(s)}(\theta, \zeta)\|$, where $\Psi_n^{*(s)}$ is defined as Ψ_n^* in (3.5) with $\widehat{Q}_n^{(s)}$ and an intermediate sequence $k^{(s)}$ replacing $\widehat{Q}_n$ and k. We obtain an estimator of the spatial parameter by least squares minimization,

$$\widehat{\vartheta}_n := \operatorname*{arg\,min}_{\vartheta \in \Theta} \sum_{s \in \mathcal{P}} \|h^{(s)}(\vartheta) - \widehat{\theta}_n^{(s)}\|^2. \tag{3.8}$$

As an alternative, one may use the relations $h^{(s)}$ between the spatial and bivariate parameters and minimize all the objective functions $\Psi_n^{*(s)}$ simultaneously, leading to the estimator

$$(\widetilde{\vartheta}_n, \widetilde{\zeta}_n) := \operatorname*{arg\,min}_{\vartheta \in \Theta, \zeta \in \mathbb{R}_+^{|\mathcal{P}|}} \sum_{s \in \mathcal{P}} \|\Psi_n^{*(s)}\big(h^{(s)}(\vartheta), \zeta^{(s)}\big)\|^2. \tag{3.9}$$

A theoretical analysis of the estimators $\widehat{\vartheta}_n$ and $(\widetilde{\vartheta}_n, \widetilde{\zeta}_n)$ is provided in Theorem 5. We further remark that the computational complexity of the proposed estimators is much lower than that of methods based on full likelihood and it compares favorably to pairwise likelihood. Additional details regarding the latter point can be found in Section S5 of the online supplement.

REMARK 2. At first glance the minimization problem in equation (3.9) seems to be computationally intractable since it contains $|\mathcal{P}| + \dim(\Theta)$ parameters and since $|\mathcal{P}|$ can be very large even for moderate dimension d. However, a closer inspection reveals that for given ϑ, partially minimizing the objective function in (3.9) over $\zeta \in \mathbb{R}_+^{|\mathcal{P}|}$ has the exact solution

$$\widehat{\zeta}_n^{(s)}(\vartheta) = \frac{\sum_{j=1}^q \int g_j(x,y) \widehat{Q}_n^{(s)}(k^{(s)}x/n, k^{(s)}y/n)\,dx\,dy}{\sum_{j=1}^q \int g_j(x,y) c_{h^{(s)}(\vartheta)}(x,y)\,dx\,dy},$$

whenever the right-hand side is positive for all s. This is satisfied if for instance g is positive everywhere and each $\widehat{Q}_n^{(s)}$ is based on at least one data point. Thus only numerical optimization over ϑ, which is usually low dimensional, is required.

4. Theoretical results. We now present our main results on the asymptotic distributions of the various estimators introduced in Section 3. First, functional central limit theorems are stated for $\widehat{c}_n$, followed by our main result on the bivariate M-estimator. Finally, asymptotic normality of the processes $\widehat{c}_n^{(s)}$ and of the two parametric estimators in the spatial setting is established. The proofs of all main results are deferred to Section S1 in the online supplement.

4.1. *The bivariate setting.* All results in this section will be derived under the following fundamental assumption.

CONDITION 1.

(i) Equation (3.1) holds uniformly on $\mathcal{S}^+ = \{(x,y) \in [0,\infty)^2 : x^2 + y^2 = 1\}$ with a function $q_1(t) = O(1/\log(1/t))$ and the function q is such that $\chi := \lim_{t\downarrow 0} q(t)/t \in [0,1]$ exists.
(ii) As $n \to \infty$, $m = m_n := nq(k/n) \to \infty$ and $\sqrt{m}q_1(k/n) \to 0$.

We note that in the proofs, equation (3.1) is required to hold locally uniformly on $[0,\infty)^2$, but by Lemma S2 uniformity on $\mathcal{S}^+$ implies uniformity over compact subsets of $[0,\infty)^2$. Condition 1(ii) is a standard assumption that makes certain bias terms negligible. It is not a model assumption; under Condition 1(i), there always exists a sequence k such that Condition 1(ii) is satisfied, and thus all of the following discussion will focus on Condition 1(i). Notably and in contrast to most of the existing literature involving nonparametric estimation, Condition 1 does not assume any differentiability of the function c. In fact, our proofs show that all the regularity required on c can be derived from equation (3.1). Considering Remark 1, it is possible to use a data-dependent value $\widehat{k}$. In following results, when this is done, we will assume that there is an unknown sequence k that satisfies Condition 1(ii), that m is defined as therein, and that $\widehat{k}$ is chosen so that $n\widehat{Q}_n(\widehat{k}/n, \widehat{k}/n) = m$.

We next discuss this condition in the examples introduced in Section 2.1. Proofs for the claims in the examples below can be found in Sections S3 and S4 of the online supplement.

EXAMPLE 7 (Example 1, continued). Most of the literature on asymptotic analysis of estimators of the stable tail dependence function ℓ or related quantities under domain of attraction conditions makes some version of the following assumption:

$$\frac{1}{t}\mathbb{P}\big(F_1(X) \ge 1 - tx \text{ and } F_2(Y) \ge 1 - ty\big) - R(x,y) = O\big(\widetilde{q}_1(t)\big), \quad x, y \in [0,\infty); \tag{4.1}$$

for a nonvanishing function R on $[0,\infty)^2$ where $\widetilde{q}_1(t) = o(1)$; see, for instance, condition (C2) in Einmahl, Krajina and Segers (2008) or the discussion in Bücher, Volgushev and Zou (2019). A simple computation involving the inclusion–exclusion formula further shows that this is equivalent to assuming that convergence in equation (1.1) takes place with rate $O(\widetilde{q}_1(t))$ and that $\ell(x,y) = x + y - R(x,y)$. Clearly, equation (4.1) implies Condition 1(i) with $q(t) = tR(1,1)$, $c(x,y) = R(x,y)/R(1,1)$ and $q_1(t) = \widetilde{q}_1(t)$.

TABLE 1
Tail expansion of the random scale model in equation (2.4). Here, we set $\mu := x \wedge y$, $\mathcal{M} := x \vee y$ and K_λ is a positive constant given in equation (S4.1) *of the online supplement*

Range of λ	$q(t)$	$c(x, y)$	$q_1(t)$
$(0, 1)$	$K_\lambda t$	$\frac{2-\lambda}{2(1-\lambda)}\mu - \frac{\lambda}{2(1-\lambda)}\mu^{1/\lambda}\mathcal{M}^{1-1/\lambda}$	$t^{1/\lambda-1}$
1	$\frac{K_\lambda t}{\log(1/t)+\log\log(1/t)}$	$\mu(1+\frac{1}{2}\log(\frac{\mathcal{M}}{\mu}))$	$1/\log(1/t)$
$(1, 2)$	$K_\lambda t^\lambda$	$\frac{\lambda}{2(\lambda-1)}\mu\mathcal{M}^{\lambda-1} - \frac{2-\lambda}{2(\lambda-1)}\mu^\lambda$	$t^{(\lambda-1)\wedge(2-\lambda)}$
2	$K_\lambda t^2\log(1/t)$	$\mu\mathcal{M}$	$1/\log(1/t)$
$(2, \infty)$	$K_\lambda t^2$	$\mu\mathcal{M}$	$t^{\lambda-2}$

EXAMPLE 8 (Example 2, continued). Let (X, Y) be a bivariate inverted max-stable distribution and assume that there exists a constant $C < \infty$ such that for all $u, v > 0$,

$$\left|\ell(1+u, 1+v) - \ell(1, 1) - \dot{\ell}_1(1, 1)u - \dot{\ell}_2(1, 1)v\right| \le C(u^2 + v^2),$$

where $\dot{\ell}_j$ represent the directional partial derivatives of ℓ from the right. In particular, it suffices for ℓ to be twice differentiable. Then the random vector (X, Y) satisfies Condition 1(i) with $q(t) = t^{\ell(1,1)}$, $c(x, y) = x^{\dot{\ell}_1(1,1)}y^{\dot{\ell}_2(1,1)}$ and $q_1(t) = 1/\log(1/t)$. Moreover, $\dot{\ell}_j(1, 1) \in (0, 1]$ and $\dot{\ell}_1(1, 1) + \dot{\ell}_2(1, 1) = \ell(1, 1) \in (1, 2]$.

EXAMPLE 9 (Example 3, continued). Let (X, Y) be a random scale construction as defined in equation (2.4) and set $\lambda = \alpha_R/\alpha_W$. Then (X, Y) satisfies Condition 1(i) with functions q, c and q_1 determined by λ as in Table 1.

4.1.1. *Asymptotic theory for nonparametric estimators.* In this section, we consider the estimator $\widehat{c}_n$ from equation (3.3). Since the process convergence results differ under asymptotic dependence and independence, we discuss these settings separately. Our first result deals with asymptotic independence.

THEOREM 1 (Asymptotic normality of $\widehat{c}_n$ under asymptotic independence). *Assume Condition 1. Then under asymptotic independence, that is, when $\chi = 0$,*

$$W_n := \sqrt{m}(\widehat{c}_n - c) \rightsquigarrow W,$$

in $\ell^\infty([0, T]^2)$, for any $T < \infty$. Here, W is a centered Gaussian process with covariance structure given by $\mathbb{E}[W(x, y)W(x', y')] = c(x \wedge x', y \wedge y')$. The same remains true if k is replaced by $\widehat{k}$ as described after Condition 1.

Note that process convergence of the estimator $\widetilde{c}_n$ from equation (3.4) can be obtained from the above result through a straightforward application of the functional delta method. This will not be needed in the theory for M-estimators in the next section and details are omitted for the sake of brevity.

Asymptotic properties of $\widehat{c}_n$ were considered in Draisma et al. (2004). However, the proof of the corresponding result (Lemma 6.1) in the latter reference makes the additional assumption that the partial derivatives of c exist and are continuous on $[0, T]^2$ (cf. Peng ((1999), Theorem 2.2)). In contrast, we are able to show that no condition on existence or continuity of partial derivatives is required. This is a considerable strengthening of the result which further allows to handle many interesting examples that were not covered by the results of Draisma et al. (2004). Indeed, both the popular class of inverted max-stable distributions in Example 2 and the random scale construction in Example 3 lead to functions c that fail to have continuous or even bounded partial derivatives. Before moving on to discussing results under asymptotic dependence, we briefly comment on some of the main ideas of the proof.

REMARK 3 (Main ideas of the proof of Theorem 1). The proof relies on the decomposition

$$\widehat{c}_n(x,y)-c(x,y)=\left\{\frac{Q_n(\frac{ku_n(x)}{n},\frac{kv_n(y)}{n})}{q(k/n)}-c\big(u_n(x),v_n(y)\big)\right\}+\big(c\big(u_n(x),v_n(y)\big)-c(x,y)\big),$$

where

$$u_n(x):=\frac{n}{k}U_{n,\lfloor kx\rfloor} \quad \text{and} \quad v_n(y):=\frac{n}{k}V_{n,\lfloor ky\rfloor},$$

and $U_{n,k}$ and $V_{n,k}$ denote the kth order statistics of $1-F_1(X_1),\dots,1-F_1(X_n)$ and $1-F_2(Y_1),\dots,1-F_2(Y_n)$, respectively with $U_{n,0}=V_{n,0}=0$. The core difficulty is to show that the difference $c(u_n(x),u_n(y))-c(x,y)$ is negligible. Under the assumption of the existence and continuity of partial derivatives of c on $[0,T]^2$ made in Draisma et al. (2004), this is a direct consequence of the fact that under asymptotic independence $\sqrt{m}(u_n(x)-x)=o_P(1)$. Dropping this assumption considerably complicates the theoretical analysis. The proof strategy is to derive bounds on increments of $c(x,y)$ for x, y close to 0 where the partial derivatives of c can become unbounded (see Lemmas S7 and S8) and to combine those bounds with subtle results on weighted weak convergence of $u_n(x)-x$ as a process in x; see Lemma S3 where we essentially leverage the findings of Csörgő and Horváth (1987).

We next turn to the case of asymptotic dependence. Results on convergence of $\widehat{c}_n$ in the space ℓ^∞ are well known under this regime; they are equivalent to similar results about estimated stable tail dependence functions (cf. Huang (1992)). However, they require the existence and continuity of partial derivatives of ℓ or, equivalently, c. As shown in Einmahl, Krajina and Segers (2008), Einmahl, Krajina and Segers (2012), the latter condition is restrictive and in fact not necessary to derive asymptotic normality of M-estimators.

The treatment of M-estimators in Einmahl, Krajina and Segers (2008), Einmahl, Krajina and Segers (2012) involves a direct analysis of certain integrals without using process convergence in $\ell^\infty([0,T]^d)$. While this approach could be transferred to our setting, we will instead follow a strategy put forward in Bücher, Segers and Volgushev (2014) and prove weak convergence of $\widehat{c}_n$ with respect to the hypimetric introduced therein. This approach will turn out to generalize much more easily when we deal with spatial estimation problems. Convergence with respect to this metric holds without any assumptions on the existence of partial derivatives and is sufficiently strong to guarantee convergence of integrals which is needed for the analysis of M-estimators.

Let $\dot{c}_1$ denote the partial derivative of c with respect to x from the left and $\dot{c}_2$ denote its partial derivative with respect to y from the right. Under asymptotic dependence, $c(x,y)\propto x+y-\ell(x,y)$ is concave since ℓ is convex (de Haan and Ferreira ((2006), Proposition 6.1.21)), hence those directional partial derivatives exist everywhere on $(0,\infty)^2$, by Theorem 23.1 of Rockafellar (1970). The definition can be extended to $[0,\infty)^2$ be setting $\dot{c}_1(0,y)$ to be the derivative from the right instead of from the left.

To describe the limiting distribution, recall that $\chi=\lim_{t\to 0}q(t)/t\in[0,1]$ is positive only in the case of asymptotic dependence. For $(x,y),(x',y')\in[0,\infty)^2$, define

$$\Lambda\big((x,y),(x',y')\big)=\begin{bmatrix} c\big(x\wedge x',y\wedge y'\big) & \chi c\big(x\wedge x',y\big) & \chi c\big(x,y\wedge y'\big) \\ \chi c\big(x\wedge x',y'\big) & \chi\big(x\wedge x'\big) & \chi^2 c\big(x,y'\big) \\ \chi c\big(x',y\wedge y'\big) & \chi^2 c\big(x',y\big) & \chi\big(y\wedge y'\big) \end{bmatrix}, \tag{4.2}$$

and let $(W,W^{(1)},W^{(2)})$ be an $\mathbb{R}^3$-valued, zero mean Gaussian process on $[0,\infty)^2$ with covariance function Λ. Note that W is the limiting process in Theorem 1, that $W^{(1)}(x,y)$ is constant in y and that $W^{(2)}(x,y)$ is constant in x.

THEOREM 2 (Asymptotic normality of $\widehat{c}_n$ under asymptotic dependence). *Assume Condition 1. Then under asymptotic dependence, that is, when $\chi > 0$,*

$$W_n \rightsquigarrow B := W - \dot{c}_1 W^{(1)} - \dot{c}_2 W^{(2)}$$

in $(L^\infty([0,T]^2), d_{\mathrm{hypi}})$, for any $T < \infty$. Here, W_n is defined as in Theorem 1. The same remains true if k is replaced by $\widehat{k}$ as described after Condition 1.

Note that weak convergence in the above theorem takes place in $(L^\infty([0,T]^2), d_{\mathrm{hypi}})$ where $L^\infty([0,T]^2)$ corresponds to equivalence classes of functions in $\ell^\infty([0,T]^2)$ with respect to the hypi(semi)metric d_{hypi}; see Bücher, Segers and Volgushev (2014) for additional details.

The proof of Theorem 2 follows by adapting the arguments given in Bücher, Segers and Volgushev (2014) for the function ℓ and builds on the fact that under asymptotic dependence the function c is differentiable almost everywhere. Note however that, in contrast to similar results in Bücher, Segers and Volgushev (2014), our limiting process is stated without appealing to lower semicontinuous extensions. This type of statement is inspired by the representation of certain integrals in Einmahl, Krajina and Segers (2012) and is possible in the bivariate setting due to concavity of c under asymptotic dependence. Additional comments on the representation of the limiting process are given in Remark 4 below.

REMARK 4. In order to obtain asymptotic results for our M-estimator, weak convergence of $\int g W_n \, d\mu_{\mathrm{L}}$ to $\int g B \, d\mu_{\mathrm{L}}$ is sufficient. Under asymptotic dependence, this is seen to follow from Theorem 2 (see the proof of Theorem 3). However, this process convergence result is not necessary. An approach that is used in Einmahl, Krajina and Segers (2012) is to write an expression for the random vector $\int g W_n \, d\mu_{\mathrm{L}}$ and directly work out its weak limit. With this strategy, $\dot{c}_j$ may be defined as left or right derivatives without problem as $\int \dot{c}_j W^{(j)} \, d\mu_{\mathrm{L}}$ will be unchanged. In contrast, proving weak hypiconvergence of W_n to B makes our results more general and more easily generalized to the spatial framework. The cost of doing so is that the directional derivatives $\dot{c}_j$ must be chosen in a specific way; see Lemma S9.

REMARK 5. Recall that under asymptotic independence, process convergence of $\widetilde{c}_n$ could be obtained from Theorem 1 by a simple application of the delta method. This is no longer the case in the general setting of Theorem 2 because weak convergence with respect to the hypimetric does not imply convergence of $W_n(1,1)$, unless the limiting process B has sample paths which are a.s. continuous in $(1,1)$. The latter happens only if the partial derivatives of c exist and are continuous in $(1,1)$. Under this additional assumption convergence of $\widetilde{c}_n$ with respect to the hypimetric can be obtained.

4.1.2. *Asymptotic theory for bivariate M-estimators.* Equipped with the process convergence tools from the previous section, we proceed to analyze the M-estimator introduced in Section 3.2. Consistency is established by standard arguments, and for the sake of brevity we do not state the corresponding results here. In the present section, we focus on the asymptotic distribution. Define the objective function $\Psi : \Theta \times \mathbb{R}_+ \to \Psi(\Theta \times \mathbb{R}_+) \subseteq \mathbb{R}^q$ by

$$\Psi(\theta, \sigma) := \sigma \int g c_\theta \, d\mu_{\mathrm{L}} - \int g c \, d\mu_{\mathrm{L}}. \tag{4.3}$$

Clearly, $\Psi(\theta_0, 1) = 0$. In addition, assume that $(\theta_0, 1)$ is a unique, well separated zero of Ψ and let $J_\Psi(\theta, \sigma)$ denote the Jacobian matrix of Ψ for points $(\theta, \sigma) \in \Theta \times \mathbb{R}_+$ where it exists.

Define $\Gamma((x,y),(x',y')) = c(x \wedge x', y \wedge y')$ under asymptotic independence and otherwise

$$\begin{aligned}\Gamma&\big((x,y),(x',y')\big)\\ &= \big(1, -\dot{c}_1(x,y), -\dot{c}_2(x,y)\big)\Lambda\big((x,y),(x',y')\big)\big(1, -\dot{c}_1(x',y'), -\dot{c}_2(x',y')\big)^\top,\end{aligned}$$

where Λ is defined in equation (4.2). Recall from the previous section that these directional derivatives always exist when $\chi > 0$ since in this case c is concave. In fact, $\Gamma((x,y),(x',y'))$ is the covariance between $W(x,y)$ and $W(x',y')$ (under asymptotic independence) or between $B(x,y)$ and $B(x',y')$ (under asymptotic dependence). Hence in those two regimes,

$$A := \int_{[0,T]^4} g(x,y)g(x',y')^\top \Gamma((x,y),(x',y'))\,dx\,dy\,dx'\,dy' \in \mathbb{R}^{q\times q}$$

is the covariance matrix of the random vector $\int gW\,d\mu_{\mathrm{L}}$ or $\int gB\,d\mu_{\mathrm{L}}$, respectively. We are now ready to state the main result of this section: asymptotic normality of $(\widehat{\theta}_n, \widehat{\zeta}_n)$, which holds under both asymptotic dependence and independence.

THEOREM 3 (Asymptotic normality of $\widehat{\theta}_n$). *Assume that Ψ has a unique, well separated zero at $(\theta_0, 1)$ and is differentiable at that point with Jacobian $J := J_\Psi(\theta_0, 1)$ of full rank $p+1$, $p = \dim(\Theta)$. Further assume Condition 1. Then the estimators $(\widehat{\theta}_n, \widehat{\zeta}_n)$ defined in equation (3.6) satisfy*

$$\sqrt{m}\left(\left(\widehat{\theta}_n, \frac{n\widehat{\zeta}_n}{m}\right) - (\theta_0, 1)\right) \rightsquigarrow N(0, \Sigma),$$

where $\Sigma := (J^\top J)^{-1} J^\top A J (J^\top J)^{-1}$. The same remains true if k is replaced by $\widehat{k}$ as described after Condition 1.

While for simplicity the estimator is defined as an exact minimizer, the same result can be obtained for an approximate minimizer. Precisely, it is obvious from the proof of Theorem 3 that as long as $\Psi_n^*(\widehat{\theta}_n, \widehat{\zeta}_n) = \inf_{\theta,\zeta} \Psi_n^*(\theta,\zeta) + o_P(\sqrt{m}/n)$, the conclusion still holds. Finally, recall that the coefficient of tail dependence η can be recovered from the function c since the latter is homogeneous of order $1/\eta$, and this relation always holds. Therefore, inside the assumed parametric model, η can be represented as a function $\eta(\theta)$. The asymptotic distribution of the resulting estimator can be obtained by a direct application of the delta method and details are omitted for the sake of brevity.

4.2. *The spatial setting.* In this section, we assume the framework of Section 3.3 and establish asymptotic properties of the estimators therein. For each pair $s \in \mathcal{P}$, let $k^{(s)}$ be an intermediate sequence and define

$$\widehat{c}_n^{(s)}(x,y) := \frac{\widehat{Q}_n^{(s)}(k^{(s)}x/n, k^{(s)}y/n)}{q^{(s)}(k^{(s)}/n)}.$$

From Section 4.1.1, the asymptotic distribution of $\widehat{c}_n^{(s)}$ is known under suitable conditions. However, as the spatial estimators $\widehat{\vartheta}_n$ and $\widetilde{\vartheta}_n$ are based on all pairs, a joint convergence statement about all processes $\widehat{c}_n^{(s)}$ is necessary. This will require an additional assumption, which we present and discuss next.

Let $F^{(1)}, \ldots, F^{(d)}$ denote the marginal distribution functions of the random vector X, which itself consists of the spatial process Y evaluated at d different locations. In order to obtain the asymptotic covariance between different processes $\widehat{c}_n^{(s)}$, we need to ensure that certain multivariate tail probabilities converge. Partition the set $\mathcal{P}$ into $\mathcal{P}_I$ and $\mathcal{P}_D$, consisting of the asymptotically independent and asymptotically dependent pairs, respectively. In the formulation of the following assumption, $s = (s_1, s_2)$ and $s^i = (s_1^i, s_2^i)$ are used to denote pairs. For brevity, $x^i = (x_1^i, x_2^i)$ is also used to denote a point in $[0,\infty)^2$.

CONDITION 2. For every $s \in \mathcal{P}$, $(X^{(s_1)}, X^{(s_2)})$ satisfies Condition 1(i) with functions $q^{(s)}$, $q_1^{(s)}$, $c^{(s)}$ and $\chi^{(s)} := \lim_{t\downarrow 0} q^{(s)}(t)/t$ exists. Intermediate sequences $k^{(s)}$ are chosen so that $m^{(s)} := nq^{(s)}(k^{(s)}/n) \to \infty$ and $\sqrt{m^{(s)}}q_1^{(s)}(k^{(s)}/n) \to 0$. For pairs $s^1, s^2 \in \mathcal{P}$, points $x^1, x^2 \in [0,\infty)^2$ and sets J of two-dimensional vectors with entries in $\{1,2\}$, let

$$\Gamma_n(s^1, s^2, x^1, x^2; J) = \frac{n}{\sqrt{m^{(s^1)}m^{(s^2)}}}\mathbb{P}\left(F^{(s_j^i)}(X^{(s_j^i)}) \geq 1 - \frac{k^{(s^i)}x_j^i}{n}, \quad (i,j) \in J\right).$$

We assume that the sequences $k^{(s)}$ are chosen such that the limits

$$\Gamma^{(s^1,s^2)}(x^1,x^2) := \lim_{n\to\infty} \Gamma_n(s^1,s^2,x^1,x^2; \{(1,1),(1,2),(2,1),(2,2)\}), \quad s^1,s^2 \in \mathcal{P},$$

$$\Gamma^{(s^1,s^2,j)}(x^1,x^2) := \chi^{(s^2)} \lim_{n\to\infty} \Gamma_n(s^1,s^2,x^1,x^2; \{(1,1),(1,2),(2,j)\}),$$
$$s^1 \in \mathcal{P}, s^2 \in \mathcal{P}_D,$$

$$\Gamma^{(s^1,j^1,s^2,j^2)}(x^1,x^2) := \chi^{(s^1)}\chi^{(s^2)} \lim_{n\to\infty} \Gamma_n(s^1,s^2,x^1,x^2; \{(1,j^1),(2,j^2)\}), \quad s^1,s^2 \in \mathcal{P}_D,$$

exist for all $j, j^i \in \{1,2\}$, and that the convergence is locally uniform over $x^1, x^2 \in [0,\infty)^2$.

We next discuss the above condition in three special cases of particular interest. The first two are processes in the domain of attraction of max-stable processes and inverted max-stable processes. The third one is a mixture process appearing in Wadsworth and Tawn (2012), which can have asymptotically dependent and independent pairs simultaneously.

EXAMPLE 10 (Example 4, continued). If Y is in the max-domain of attraction of a max-stable process, then X is in the max-domain of attraction of a max-stable distribution G on $\mathbb{R}^d$ with stable tail dependence function

$$\ell(x_1,\dots,x_d) := \lim_{t\downarrow 0} \frac{1}{t}\mathbb{P}(F^{(1)}(X^{(1)}) \geq 1 - tx_1 \text{ or } \dots \text{ or } F^{(d)}(X^{(d)}) \geq 1 - tx_d), \quad x_j \geq 0;$$

see equation (1.1). If moreover the convergence is locally uniform over $(x_1,\dots,x_d) \in [0,\infty)^d$ and if every pair is asymptotically dependent, then Condition 2 holds. Note that this is automatically satisfied if Y itself is max-stable. The sequences $k^{(s)}$ can be chosen all equal to k, say, and for every pair s, $m^{(s)}/k \to \chi^{(s)} > 0$. The sequences $m^{(s)}$ can also be chosen all asymptotically equivalent to m, say, by choosing $k^{(s)} = m/\chi^{(s)}$. The limiting covariance terms can all be deduced from ℓ by straightforward calculations.

EXAMPLE 11 (Example 5, continued). If Y is an inverted max-stable process, then X has an inverted max-stable distribution, and we assume that the associated stable tail dependence function ℓ is componentwise strictly increasing. The latter is trivially satisfied if X has a positive density. Then if all the pairwise functions $\ell^{(s)}$ satisfy the quadratic expansion introduced in Example 8, Condition 2 is satisfied and the sequences $k^{(s)}$ can be chosen so that the $m^{(s)}$ are all equal, that is, for every pair $s \in \mathcal{P}$, $m^{(s)} = m$ for some intermediate sequence m. Here, $\mathcal{P}_D$ is empty so the only required covariance terms are (see Section S3)

$$\Gamma^{(s^1,s^2)}(x^1,x^2) = \begin{cases} c^{(s)}(x_1^1 \wedge x_1^2, x_2^1 \wedge x_2^2) & s^1 = s^2 = s, \\ 0 & s^1 \neq s^2. \end{cases}$$

For instance, any inverted Brown–Resnick process (or rather the implied inverted d-dimensional Hüsler–Reiss distribution corresponding to the d observed locations) satisfies Condition 2 as long as the aforementioned d-variate distribution has a density. The latter can easily be checked (e.g., Engelke and Hitz ((2020), Corollary 2)).

EXAMPLE 12 (Wadsworth and Tawn ((2012), Section 4)). Let Z be a max-stable process and Z' be an inverted max-stable process, both with unit Fréchet margins. Suppose that Z' satisfies the monotonicity condition stated in Example 11, and additionally that none of its pairwise distributions $(Z'(u_1), Z'(u_2))$ is perfectly independent. Let $a \in (0, 1)$ and define the process Y by

$$Y(u) := \max\{aZ(u), (1-a)Z'(u)\}.$$

Then Y also has unit Fréchet margins. If Z becomes independent at a certain spatial distance, the process Y transitions between asymptotic dependence and independence at that distance. An instance of such a max-stable process Z is found in the second example after Theorem 1 of Schlather (2002), assuming that the Radius R of the random disks is bounded (see also Davison, Padoan and Ribatet ((2012), equation (23) and the discussion that precedes)).

The process Y can be shown to satisfy Condition 2 if the sequences $k^{(s)}$ are chosen so that the $m^{(s)}$ are all equal. The terms $\Gamma^{(s^1,s^2)}$, $\Gamma^{(s^1,s^2,j)}$ and $\Gamma^{(s^1,j^1,s^2,j^2)}$ are mostly determined by the process Z, as in Example 10; see Section S3 in the online supplement for details.

4.2.1. *Joint distribution of nonparametric estimators.* The joint limiting behavior of the processes $\widehat{c}_n^{(s)}$ relies on $((W^{(s)})_{s\in\mathcal{P}}, (W^{(s,j)})_{s\in\mathcal{P}_D, j\in\{1,2\}})$, a collection of centered Gaussian processes on $[0,\infty)^2$. The covariance between $W^{(s)}(x, y)$ and $W^{(s')}(x', y')$ is given by $\Gamma^{(s,s')}((x, y), (x', y'))$, the covariance between $W^{(s)}(x, y)$ and $W^{(s',j)}(x', y')$ takes the form $\Gamma^{(s,s',j)}((x, y), (x', y'))$, and the covariance between $W^{(s,j)}(x, y)$ and $W^{(s',j')}(x', y')$ is equal to $\Gamma^{(s,j,s',j')}((x, y), (x', y'))$. For $s \in \mathcal{P}_I$, let $B^{(s)} = W^{(s)}$ and for $s \in \mathcal{P}_D$, let

$$B^{(s)} = W^{(s)} - \dot{c}_1^{(s)} W^{(s,1)} - \dot{c}_2^{(s)} W^{(s,2)},$$

where $\dot{c}_j^{(s)}$ are defined similar to $\dot{c}_j$ in Section 4.1.1.

THEOREM 4 (Asymptotic normality of $\widehat{c}_n^{(s)}$). *Assume Condition 2. Then*

$$(W_n^{(s)})_{s\in\mathcal{P}} := \left(\sqrt{m^{(s)}}\left(\widehat{c}_n^{(s)} - c^{(s)}\right)\right)_{s\in\mathcal{P}} \rightsquigarrow (B^{(s)})_{s\in\mathcal{P}}$$

in the product space $(L^\infty([0, T]^2), d_{\mathrm{hypi}})^{|\mathcal{P}|}$, *for any* $T < \infty$. *The same remains true if each* $k^{(s)}$ *is replaced by the data-dependent sequence* $\widehat{k}^{(s)}$ *as described after Condition 1.*

The preceding result can be applied in all generality as long as the four-dimensional tails of the spatial process of interest are sufficiently smooth. The admissible settings include, but are far from limited to, Examples 10 to 12.

According to Bücher, Segers and Volgushev (2014), convergence in the hypimetric is equivalent to uniform convergence when the limit is a continuous function. The process $B^{(s)}$ clearly has almost surely continuous sample paths under asymptotic independence, as well as under asymptotic dependence if the partial derivatives of c exist everywhere and are continuous. It follows that in those cases $W_n^{(s)}$ converges in $(\ell^\infty([0, T]^2), \|\cdot\|_\infty)$. In fact, one may replace the product space in the result above by $\bigotimes_{s\in\mathcal{P}} \mathbb{D}^{(s)}$, where $\mathbb{D}^{(s)}$ represents either $\ell^\infty([0, T]^2)$ equipped with the supremum distance (if $s \in \mathcal{P}_I$ or c has continuous partial derivatives) or $L^\infty([0, T]^2)$ equipped with the hypimetric (otherwise). In particular, for processes where every pair is asymptotically independent such as inverted max-stable processes, the hypimetric can be replaced by the supremum distance everywhere.

4.2.2. *Asymptotics for parametric estimators.* We now show how Theorem 4 leads to asymptotic results for the parametric estimators $\widehat{\vartheta}_n$ and $\widetilde{\vartheta}_n$ introduced in equations (3.8) and (3.9). Recall the setting of Section 3.3, and in particular the functions $h^{(s)} : \Theta \to \widetilde{\Theta}$ and the relation $c^{(s)} = c_{h^{(s)}(\vartheta_0)}$. Similar to the bivariate setting, define

$$\Psi^{(s)} : \widetilde{\Theta} \times \mathbb{R}_+ \to \mathbb{R}^q, \qquad \Psi^{(s)}(\theta, \sigma) = \sigma \int g c_\theta \, d\mu_{\mathrm{L}} - \int g c^{(s)} \, d\mu_{\mathrm{L}}.$$

In the bivariate setting, we required Ψ to be differentiable and have a unique well-separated zero. In the spatial setting, we need a comparable assumption.

CONDITION 3. For every pair $s \in \mathcal{P}$, the functions $\Psi^{(s)}$ and $h^{(s)}$ are continuously differentiable at the points $(h^{(s)}(\vartheta_0), 1)$ and ϑ_0, respectively, with Jacobian matrices $J_{\Psi^{(s)}}(h^{(s)}(\vartheta_0), 1)$ and $J_{h^{(s)}}(\vartheta_0)$ of full ranks $\widetilde{p} + 1$ and p. Additionally, (i) or (ii) holds.

(i) The functions $\Psi^{(s)}$ and $\vartheta \mapsto (h^{(s)}(\vartheta) - h^{(s)}(\vartheta_0))_{s \in \mathcal{P}}$ have a unique, well-separated zero at the points $(h^{(s)}(\vartheta_0), 1)$ and ϑ_0, respectively.

(ii) The function $(\vartheta, \sigma) \mapsto (\Psi^{(s)}(h^{(s)}(\vartheta), \sigma^{(s)}))_{s \in \mathcal{P}}$ as a function on $\Theta \times \mathbb{R}_+^{|\mathcal{P}|}$ has a unique, well-separated zero at the point $(\vartheta_0, 1, \ldots, 1)$.

Assuming both parts of Condition 3, we now introduce the notation that is needed to define the limiting covariance matrices of the two estimators. In the following, elements of a vector $x \in \mathbb{R}^{q|\mathcal{P}|}$ are ordered by pair $s \in \mathcal{P}$ first, and then by dimension $j \in \{1, \ldots, q\}$. The same convention is used when ordering the rows or columns of a matrix.

Letting $B^{(s)}$ denote the limiting Gaussian processes appearing in Theorem 4, consider the matrix $A \in \mathbb{R}^{q|\mathcal{P}| \times q|\mathcal{P}|}$ with blocks of the form

$$A^{(s,s')} := \int_{[0,T]^4} g(x, y) g(x', y')^\top \mathbb{C}\mathrm{ov}(B^{(s)}(x, y); B^{(s')}(x', y')) \, dx \, dy \, dx' \, dy'.$$

Let $\mathcal{D} \in \mathbb{R}^{\widetilde{p}|\mathcal{P}| \times q|\mathcal{P}|}$ be a block-diagonal matrix with blocks given by

$$\mathcal{D}^{(s)} := \left[\left(J_{\Psi^{(s)}}(h^{(s)}(\vartheta_0), 1)^\top J_{\Psi^{(s)}}(h^{(s)}(\vartheta_0), 1) \right)^{-1} J_{\Psi^{(s)}}(h^{(s)}(\vartheta_0), 1)^\top \right]_{1:\widetilde{p}, 1:q} \in \mathbb{R}^{\widetilde{p} \times q}, \tag{4.4}$$

where $s \in \mathcal{P}$ and $[M]_{1:\widetilde{p},1:q}$ indicates the submatrix consisting of rows 1 to $\widetilde{p}$ and columns 1 to q of the matrix M. Define $J_1 \in \mathbb{R}^{\widetilde{p}|\mathcal{P}| \times p}$ by stacking the matrices $J_{h^{(s)}}(\vartheta_0)$, $s \in \mathcal{P}$, on top of each other. Denote by $(e^{(s)})^\top$ the unit vector in $\mathbb{R}^{|\mathcal{P}|}$ with a one in the position corresponding to the pair s and let $J_2 \in \mathbb{R}^{q|\mathcal{P}| \times (p+|\mathcal{P}|)}$ be obtained by stacking the matrices

$$J_{\Psi^{(s)}}(h^{(s)}(\vartheta_0), 1) \begin{bmatrix} J_{h^{(s)}}(\vartheta_0) & 0 \\ 0 & e^{(s)} \end{bmatrix} \in \mathbb{R}^{q \times (p+|\mathcal{P}|)}, \quad s \in \mathcal{P},$$

on top of each other. Finally, define

$$\Sigma_1 = (J_1^\top J_1)^{-1} J_1^\top \mathcal{D} A \mathcal{D}^\top J_1 (J_1^\top J_1)^{-1}, \qquad \Sigma_2 = (J_2^\top J_2)^{-1} J_2^\top A J_2 (J_2^\top J_2)^{-1}.$$

THEOREM 5 (Asymptotic normality of the estimators of ϑ). *Assume Condition 2 and suppose that the sequences $m^{(s)}$ are all asymptotically equivalent to m, say. Then under Condition 3(i), the estimator defined in equation (3.8) satisfies*

$$\sqrt{m}(\widehat{\vartheta}_n - \vartheta_0) \rightsquigarrow N(0, \Sigma_1)$$

and under Condition 3(ii), the estimators defined in equation (3.9) satisfy

$$\sqrt{m}\left(\left(\widetilde{\vartheta}_n, \frac{n \widetilde{\zeta}_n}{m} \right) - (\vartheta_0, 1, \ldots, 1) \right) \rightsquigarrow N(0, \Sigma_2),$$

where Σ_1 and Σ_2 are as above. The same remains true if each $k^{(s)}$ is replaced by the data-dependent sequence $\widehat{k}^{(s)}$, based on the same sequence m, as described after Condition 1.

The assumption of asymptotic equivalence of all $m^{(s)}$ can be substantially relaxed. Otherwise, a simple way to satisfy it is to select one m and use data-driven sequences $\widehat{k}^{(s)}$.

5. Simulations.

5.1. *Bivariate distributions.* In this section, we study the finite sample behavior of the estimator introduced in the paper. We simulate samples from the bivariate vector $(X+X', Y+Y')$, where (X, Y) is the signal and (X', Y') is and independent noise vector. We consider three different models for the bivariate distributions (X, Y).

(M1) The inverted Hüsler–Reiss model from Example 2(i) with unit Fréchet margins, whose corresponding class of functions c takes the form $c_\theta(x, y) = (xy)^\theta$ where $\theta \in (1/2, 1]$.

(M2) The inverted asymmetric logistic model from Example 2(ii) with fixed $r = 2$ and unit Fréchet margins. We fit the full parametric model $\{c_\theta(x, y) = x^{\theta_1} y^{\theta_2} : \theta \in \Theta\}$, where $\Theta := \{(\theta_1, \theta_2) \in (0, 1]^2 : \theta_1 + \theta_2 > 1\}$, even though due to our choice of r the only attainable parameters are approximately the square $[0.7, 1]^2$; see Figure 4.

(M3) The random scale construction from Example 3 where we fix $\alpha_W = 1$ and vary α_R. The collection of possible functions $c = c_\lambda$, $\lambda \in (0, 2)$ is given in Table 1.

Figures S1 to S3 in the online supplement show realizations of models M1–M3 corresponding to different parameter values and rescaled to unit exponential margins for illustration.

As a noise vector we simulate samples of (X', Y'), where X' and Y' are independent with Pareto distribution function $1 - 1/x^4$, $x \geq 1$. Note that this tail is lighter than that of the marginal distributions in all three models; it can be shown that this additive noise does not affect the functions q and c of (X, Y).

All of the results that follow are based on 1000 simulation repetitions and samples of size $n = 5000$. In all the simulations, we use the same weight function (represented by g in equation (3.5)), which we now describe. Consider the following rectangles: $I_1 := [0, 1]^2$, $I_2 := [0, 2]^2$, $I_3 := [1/2, 3/2]^2$, $I_4 := [0, 1] \times [0, 3]$ and $I_5 := [0, 3] \times [0, 1]$. The function $g : \mathbb{R}^2 \to \mathbb{R}^5$ is given by

$$g(x, y) := \left(\mathbb{1}\{(x, y) \in I_1\}/a_{1,\theta_{\mathrm{REF}}}, \dots, \mathbb{1}\{(x, y) \in I_5\}/a_{5,\theta_{\mathrm{REF}}}\right)^\top, \tag{5.1}$$

where $a_{j,\theta_{\mathrm{REF}}} := \int_{I_j} c_{\theta_{\mathrm{REF}}}\, d\mu_{\mathrm{L}}$ and θ_{REF} is simply a reference point in the parameter space that ensures that all components of g have comparable magnitude. In the three models above, the reference points are 0.6, (0.6, 0.6) and 1, respectively. The rectangles are chosen in order to capture various aspects of the function c: I_3 contains information about the unknown scale ζ (recall that we scale c so that $c(1, 1) = 1$). The rectangles I_1, I_2 are geared toward determining homogeneity properties of c since $I_2 = 2I_1$ and are especially useful for estimating η. The rectangles I_4, I_5 are informative about asymmetry of the function c with respect to its arguments. Different choices of the weight function would be possible, and the best choice will be different for each model under consideration and even for each specific parameter value within a given model class. Nevertheless, the aforementioned choice seems close to optimal for all the models considered here. In Section S6 of the online supplement, a sensitivity analysis is carried out where we repeat the simulation study with different weight functions that are constructed by considering only some of the rectangles $I_1, \dots, I_5$ instead of all five. See also Einmahl, Krajina and Segers (2008), Einmahl, Krajina and Segers (2012) for a related discussion in the estimation of stable tail dependence functions.

5.1.1. *The inverted Hüsler–Reiss model* ($M1$). Figure 1 shows the effect of k on the estimation performance of $\widehat{\theta}_n$ from equation (3.6) in terms of absolute bias and root MSE for the three parameter values $\theta = 0.6$, 0.75 and 0.9. We observe that for larger values of θ (or

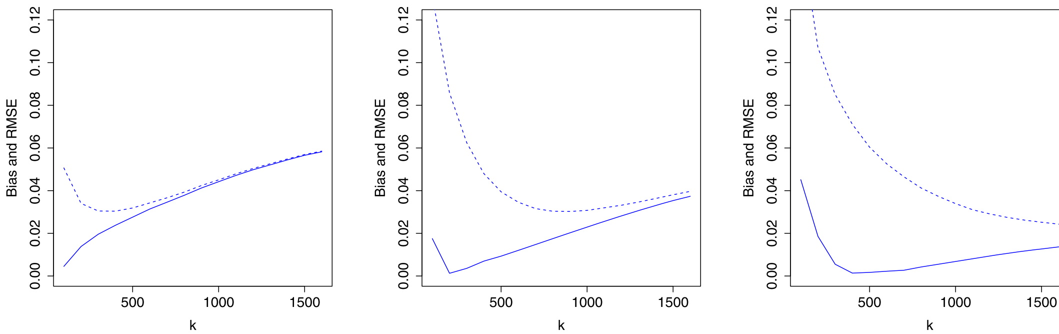

FIG. 1. *Absolute bias* (*solid lines*) *and RMSE* (*dashed lines*) *of the M-estimator of* θ *as a function of* k, *based on* 1000 *samples of size* 5000 *from model M*1 *with parameter values* 0.6, 0.75 *and* 0.9, *from left to right*.

smaller values of η, corresponding to more independence in the extremes) larger values of k lead to the best RMSE. This is in line with our theory as, for fixed k, smaller η corresponds to smaller values of m, and hence larger asymptotic variance.

An analysis of $\widehat{\theta}_n$ for a finer range of parameter values is provided in Figure 2. Motivated by the findings in Figure 1, we fix $k = 800$; this choice leads to reasonable performance across all parameter values. Overall, the results are satisfactory, with a more pronounced negative bias for smaller values of θ and more variance for increasing θ.

5.1.2. *The inverted asymmetric logistic model* (*M*2). Figure 3 shows the impact of k on estimated parameter values for three different choices of θ. Since here the parameter is two-dimensional, we consider (and estimate) the Euclidean bias and RMSE of the estimator $\widehat{\theta}_n$, defined as $\|\mathbb{E}[\widehat{\theta}_n - \theta]\|$ and $(\mathbb{E}\|\widehat{\theta}_n - \theta\|^2)^{1/2}$, respectively.

Similar to the pattern observed in Figure 1, we see that smaller values of η necessitate larger values of k in order to achieve a good balance between bias and variance.

Figure 4 shows the performance of the proposed M-estimator for a range of different parameters (θ_1, θ_2) with Euclidean bias in the left panel and RMSE in the right panel; the value $k = 800$ is fixed throughout. Since the relation $(\nu, \phi) \mapsto (\theta_1, \theta_2)$ is not easily invertible, we selected a grid of values of $(\nu, \phi) \in [0, 1]^2$, calculated all the corresponding points θ and kept the values for which $\theta_j \leq 0.95$, $j = 1, 2$.

We observe that the estimators perform better for parameter values close to the diagonal, with larger bias and variance for more asymmetric parameter values. The overall estimation accuracy is reasonably good, with worst case RMSE values around 0.07.

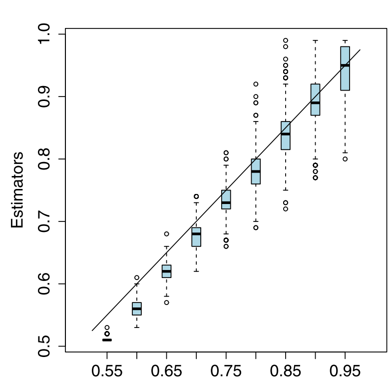

FIG. 2. *Box plots of the M-Estimators of* θ *based on* 1000 *samples of size* 5000 *for each parameter value*.

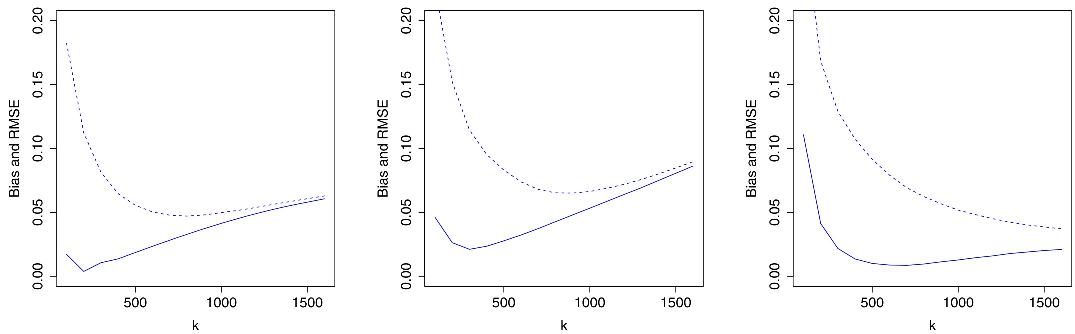

FIG. 3. *Absolute bias (solid lines) and RMSE (dashed lines) of the M-estimator of* θ *as a function of* k, *based on* 1000 *samples of size* 5000 *from model M2 with parameter* θ *equal to* $(0.72, 0.72)$, $(0.75, 0.91)$ *and* $(0.91, 0.91)$, *from left to right. In the original parametrization, the corresponding values of* (ν, ϕ) *are* $(0.94, 0.94)$, $(0.44, 0.94)$ *and* $(0.31, 0.31)$, *respectively.*

5.1.3. *The Pareto random scale model* ($M3$). Figure 5 shows the effect of k on the performance of our M-estimator $\widehat{\lambda}_n$ in terms of absolute bias and root MSE for the three parameter values $\lambda = 0.4$, 1 and 1.6. We notice that the estimator is considerably more biased at $\lambda = 1$ than at other parameter values. This is expected as, according to Table 1, the bias function q_1 vanishes only at a logarithmic rate when $\lambda = 1$, compared to a polynomial rate elsewhere. Moreover, like in the other models, we observe that for more independent data (characterized by larger λ), larger values of k are required to drive down the variance of the estimator.

An analysis of $\widehat{\lambda}_n$ for a finer range of parameter values is provided in Figure 6. Motivated by Figure 5, we fix $k = 400$, which approximately minimizes the maximal RMSE. Overall the estimator is very precise for small values of λ, but incurs a bias around $\lambda = 0.8$ where it struggles to distinguish between values slightly smaller and slightly larger than 1. This phenomenon is not completely unexpected; a close look at Table 1 reveals that c_λ has almost (but not quite) a symmetry around the point $\lambda = 1$, for example, $c_{0.8}$ is very similar in shape to $c_{1.2}$. This point also corresponds to the transition between asymptotic dependence and independence, which makes estimation challenging.

5.2. *Spatial models.* In this section, we illustrate the performance of the proposed methodology for spatial data. The candidate class for c_θ results from inverted Brown–Resnick

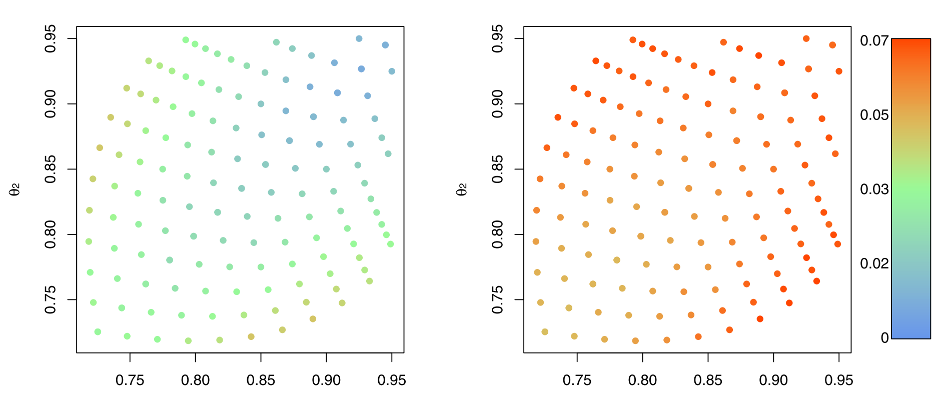

FIG. 4. *Absolute bias (left) and RMSE (right) of the M-estimator of* $\theta = (\theta_1, \theta_2)$ *as a function of* θ, *based on* 1000 *samples of size* 5000 *from model M2.*

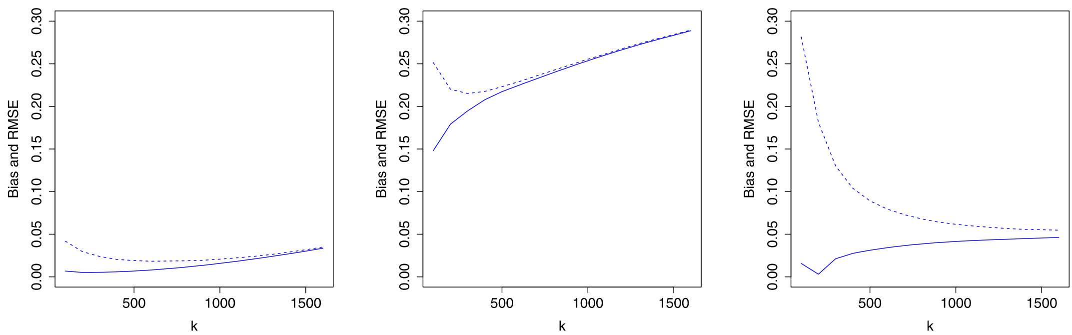

FIG. 5. *Absolute bias (solid lines) and RMSE (dashed lines) of the M-estimator of λ as a function of k, based on* 1000 *samples of size* 5000 *from model M*3 *with parameter values* 0.4, 1 *and* 1.6, *from left to right.*

processes with fractal variograms (see Example 6) and takes the form

$$c_{\vartheta}^{(s)}(x,y) = (xy)^{\theta^{(s)}}, \qquad \theta^{(s)} = \theta(\Delta^{(s)};\vartheta) := \Phi\left(\frac{1}{2}(\Delta^{(s)}/\beta)^{\alpha/2}\right), \quad s \in \mathcal{P}, \tag{5.2}$$

where $\vartheta = (\alpha, \beta) \in (0, 2] \times \mathbb{R}_+$ and $\Delta^{(s)}$ is the Euclidean distance between the two locations in pair s (measured in units of latitude). Motivated by the data application in the following section, the true parameter values are set as $\vartheta_0 = (1, 3)$ and the values for $\Delta^{(s)}$ are obtained from 40 randomly sampled pairs of locations in that data set; see Figure S5 in the online supplement for a histogram of the distances in this sample.

To evaluate the performance of our estimators, we simulate 1000 independent data sets, each of size 5000, of an inverted Brown–Resnick process with unit Fréchet margins and fractal variogram from equation (3.7) with $\alpha = 1$, $\beta = 3$. Following the bivariate simulations, to each of the 40 components of the data, we add an independent random variable with Pareto distribution function $1 - 1/x^4$, $x \geq 1$. Using the same weight function g as in the bivariate simulations (see equation (5.1)), we compute the two estimators introduced in equations (3.8) and (3.9). Since the performance of both estimators turns out to be very similar, we only report results for the least squares estimator from equation (3.8) here and defer all simulations for the estimator (3.9) to Section S6 in the online supplement.

Following the discussion in Remark 1, we fix a value m and select each $k^{(s)}$ such that $\widehat{Q}_n^{(s)}(k^{(s)}/n, k^{(s)}/n) = m$. The first two panels of Figure 7 show the absolute bias and RMSE of the estimators $\widehat{\alpha}$ and $\widehat{\beta}$, respectively, as functions of $m \in \{75, 100, \ldots, 500\}$. We observe that the RMSE for both estimators is relatively large across all values of m. Interestingly, this does not result in a bad performance in estimating the function $\theta(\cdot;\vartheta)$. Indeed, the last

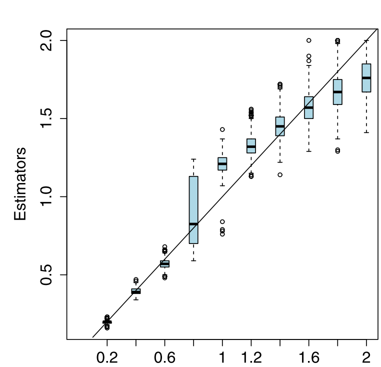

FIG. 6. *Box plots of the M-Estimators of λ based on* 1000 *samples of size* 5000 *for each parameter value.*

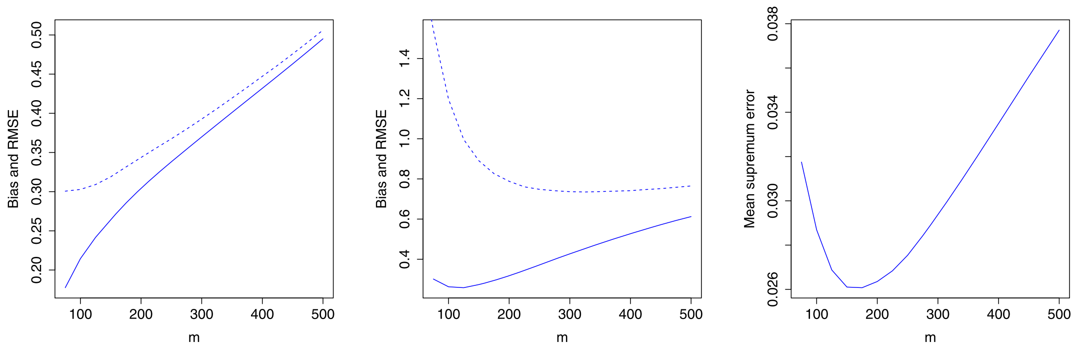

FIG. 7. *Left and middle columns*: *Bias* (*solid line*) *and RMSE* (*dotted line*) *of the estimators of the two spatial parameters* α (*left*) *and* β (*middle*) *as a function of* m. *Right*: *Mean of the supremum error* $\sup_{0\le\Delta\le3}|\theta(\Delta;\widehat{\alpha},\widehat{\beta})-\theta(\Delta;\alpha,\beta)|$ *as a function of* m.

panel of Figure 7 shows averaged (over simulation runs) values for $\sup_{0\le\Delta\le3}|\theta(\Delta;\widehat{\vartheta})-\theta(\Delta;\vartheta)|$ and indicates a good overall performance; note that the observed values of Δ are all smaller than 3 (see Figure S5 in the online supplement). This can be explained by the fact that different values of (α,β) can lead to somewhat similar curves in the range of interest. This is further illustrated in the left panel of Figure 8 where a random sample of 50 estimated functions $\theta(\Delta;\widehat{\vartheta})$ is displayed.

We conclude this section by fixing $m=150$ and comparing the performance of estimators for $\theta^{(s)}$ based on a bivariate sample at a given distance and the spatial estimator discussed above. Boxplots corresponding to five pairs of stations with distances $\Delta^{(s)}\approx 0.5, 1, \dots, 2.5$ are shown in the left panel of Figure 8. As expected from the theory, using the spatial estimator is advantageous as it allows to combine information from different distances and leads to a reduced variance.

6. Application to rainfall data. In a data set introduced in Le et al. (2018), rainfall was measured daily from 1960 to 2009 at a set of 92 different locations in the state of Victoria, southeastern Australia, for a total of $n=18{,}263$ measurements. The conclusions in that paper are that an asymptotically independent model is suitable. A subset of 40 locations, for a total of 780 pairs, was randomly sampled; see the right panel of Figure 9. To the data at those selected locations, we fit the same tail model as in Section 5.2, given in equation (5.2). The

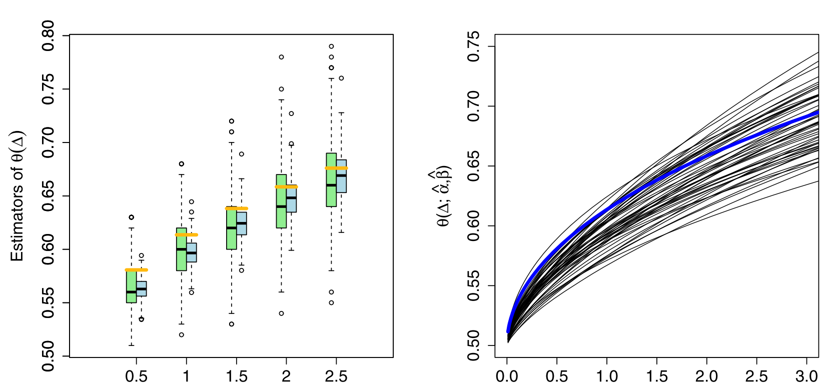

FIG. 8. *Left panel*: *Estimators of* $\theta(\Delta)$ *for* 5 *different distances*. *For each distance*, *bivariate M-estimator* $\widehat{\theta}_n^{(s)}$ (*green*) *and spatial estimator* $\theta(\Delta^{(s)};\widehat{\alpha},\widehat{\beta})$ (*blue*) *based on the* $d=40$ *locations*. *Right panel*: 50 *sampled curves* $\theta(\cdot;\widehat{\alpha},\widehat{\beta})$. *Blue represents the true curve* $\theta(\cdot;\alpha,\beta)$.

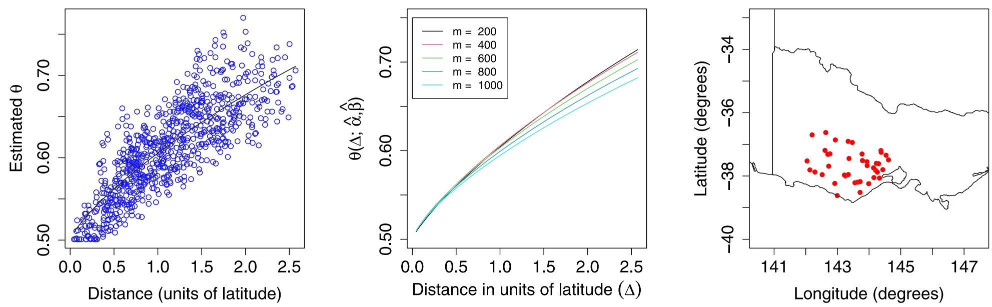

FIG. 9. *Left*: *Estimated parameters* $\widehat{\theta}_n^{(s)}$ *against the distances* $\Delta^{(s)}$. *The black line represents the estimated curve* $\theta(\cdot; 1.55, 2.24)$. *Middle*: *Estimated curve* $\theta(\cdot; \widehat{\alpha}, \widehat{\beta})$ *for the least squares estimator with different values of* m. *Right*: *The* 40 *sampled locations in the state of Victoria*, *southeastern Australia*.

weight function g that we use is the same as before and as in Section 5.2, we make use of Remark 1 by fixing a value m and choosing each $k^{(s)}$ accordingly.

We set $m = 400$. The left panel of Figure 9 shows the 780 pairwise estimators $\widehat{\theta}_n^{(s)}$ plotted against the distances $\Delta^{(s)}$. Despite some estimates at the boundary of the parameter space, the results do not provide much evidence for asymptotic dependence, whereas all estimates are away from the boundary for distances of at least 0.3 units of latitude, strongly suggesting asymptotic independence at these distances. Our two estimators (3.8) and (3.9) of (α, β) yield estimates $(\widehat{\alpha}, \widehat{\beta})$ of (1.55, 2.24) and (1.56, 2.24), respectively. They are extremely similar, as hinted by the simulation study from Section 5.2. The curve $\theta(\cdot; \widehat{\alpha}, \widehat{\beta})$ corresponding to the least squares estimator is also shown in the left panel of Figure 9. The middle panel of Figure 9 displays similar curves for the least squares estimator when m varies from 200 to 1000. It shows that the estimated curve is robust with respect to the choice of m.

Acknowledgments. We are grateful to two reviewers for their valuable input. Their constructive comments and suggestions resulted in a substantial improvement of this paper.

Funding. Michaël Lalancette was supported by the Fonds de recherche du Québec—Nature et technologies and by an Ontario Graduate Scholarship. Sebastian Engelke was supported by the Swiss National Science Foundation and the Fields Institute for Research in Mathematical Sciences. Stanislav Volgushev was supported in part by a discovery grant from NSERC of Canada and a Connaught new researcher award.

SUPPLEMENTARY MATERIAL

Supplement to "Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes" (DOI: 10.1214/20-AOS2046SUPP; .pdf). The Supplementary Material (Lalancette, Engelke and Volgushev (2021)) is divided into six sections. Section S1 contains the proofs of all main results, with a number of necessary technical results deferred to Section S2. Sections S3 and S4 present proofs of several claims from different examples in the paper. A brief discussion of computational complexity in spatial estimation is given in Section S5 and additional simulation results appear in Section S6.

REFERENCES

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050. MR3456363 https://doi.org/10.1214/15-AOAS863

BROWN, B. M. and RESNICK, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probab.* **14** 732–739. MR0517438 https://doi.org/10.2307/3213346

BÜCHER, A. and DETTE, H. (2013). Multiplier bootstrap of tail copulas with applications. *Bernoulli* **19** 1655–1687. MR3129029 https://doi.org/10.3150/12-BEJ425

BÜCHER, A., SEGERS, J. and VOLGUSHEV, S. (2014). When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi- and hypographs. *Ann. Statist.* **42** 1598–1634. MR3262462 https://doi.org/10.1214/14-AOS1237

BÜCHER, A., VOLGUSHEV, S. and ZOU, N. (2019). On second order conditions in the multivariate block maxima and peak over threshold method. *J. Multivariate Anal.* **173** 604–619. MR3959186 https://doi.org/10.1016/j.jmva.2019.04.011

COLES, S., HEFFERNAN, J. and TAWN, J. (1999). Dependence measures for extreme value analyses. *Extremes* **2** 339–365.

CSÖRGŐ, M. and HORVÁTH, L. (1987). Approximation of intermediate quantile processes. *J. Multivariate Anal.* **21** 250–262. MR0884099 https://doi.org/10.1016/0047-259X(87)90004-2

DAVISON, A. C. and GHOLAMREZAEE, M. M. (2012). Geostatistics of extremes. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **468** 581–608. MR2874052 https://doi.org/10.1098/rspa.2011.0412

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980 https://doi.org/10.1214/11-STS376

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

DE HAAN, L. and LIN, T. (2001). On convergence toward an extreme value distribution in $C[0, 1]$. *Ann. Probab.* **29** 467–483. MR1825160 https://doi.org/10.1214/aop/1008956340

DE HAAN, L., NEVES, C. and PENG, L. (2008). Parametric tail copula estimation and model testing. *J. Multivariate Anal.* **99** 1260–1275. MR2419346 https://doi.org/10.1016/j.jmva.2007.08.003

DOMBRY, C., ENGELKE, S. and OESTING, M. (2017). Asymptotic properties of the maximum likelihood estimator for multivariate extreme value distributions. Available from https://arxiv.org/abs/1612.05178.

DRAISMA, G., DREES, H., FERREIRA, A. and DE HAAN, L. (2004). Bivariate tail estimation: Dependence in asymptotic independence. *Bernoulli* **10** 251–280. MR2046774 https://doi.org/10.3150/bj/1082380219

DREES, H. and HUANG, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64** 25–47. MR1619974 https://doi.org/10.1006/jmva.1997.1708

EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2008). A method of moments estimator of tail dependence. *Bernoulli* **14** 1003–1026. MR2543584 https://doi.org/10.3150/08-BEJ130

EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793. MR3015043 https://doi.org/10.1214/12-AOS1023

EINMAHL, J. H. J. and SEGERS, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.* **37** 2953–2989. MR2541452 https://doi.org/10.1214/08-AOS677

EINMAHL, J. H. J., KIRILIOUK, A., KRAJINA, A. and SEGERS, J. (2016). An M-estimator of spatial tail dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 275–298. MR3453656 https://doi.org/10.1111/rssb.12114

ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 871–932. MR4136498

ENGELKE, S. and IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.* **8**. https://doi.org/10.1146/annurev-statistics-040620-041554

ENGELKE, S., OPITZ, T. and WADSWORTH, J. (2019). Extremal dependence of random scale constructions. *Extremes* **22** 623–666. MR4031852 https://doi.org/10.1007/s10687-019-00353-3

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 239–265. MR3299407 https://doi.org/10.1111/rssb.12074

GUILLOTTE, S., PERRON, F. and SEGERS, J. (2011). Non-parametric Bayesian inference on bivariate extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 377–406. MR2815781 https://doi.org/10.1111/j.1467-9868.2010.00770.x

GUMBEL, É. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Stat. Univ. Paris* **9** 171–173. MR0119279

HEFFERNAN, J. E. and RESNICK, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.* **17** 537–571. MR2308335 https://doi.org/10.1214/105051606000000835

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 497–546. With discussions and reply by the authors. MR2088289 https://doi.org/10.1111/j.1467-9868.2004.02050.x

HUANG, X. (1992). Statistics of bivariate extreme value theory. Ph.D. thesis, Erasmus Univ. Rotterdam.

HUSER, R., OPITZ, T. and THIBAUD, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spat. Stat.* **21** 166–186. MR3692183 https://doi.org/10.1016/j.spasta.2017.06.004

HUSER, R. and WADSWORTH, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *J. Amer. Statist. Assoc.* **114** 434–444. MR3941266 https://doi.org/10.1080/01621459.2017.1411813

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR0980699 https://doi.org/10.1016/0167-7152(89)90106-5

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440 https://doi.org/10.1214/09-AOP455

KEEF, C., TAWN, J. and SVENSSON, C. (2009). Spatial risk assessment for extreme river flows. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 601–618. MR2750258 https://doi.org/10.1111/j.1467-9876.2009.00672.x

LALANCETTE, M., ENGELKE, S. and VOLGUSHEV, S. (2021). Supplement to "Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes." https://doi.org/10.1214/20-AOS2046SUPP

LE, P. D., DAVISON, A. C., ENGELKE, S., LEONARD, M. and WESTRA, S. (2018). Dependence properties of spatial rainfall extremes and areal reduction factors. *J. Hydrol.* **565** 711–719.

LEDFORD, A. W. and TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187. MR1399163 https://doi.org/10.1093/biomet/83.1.169

LEDFORD, A. W. and TAWN, J. A. (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B* **59** 475–499. MR1440592 https://doi.org/10.1111/1467-9868.00080

PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. MR2757202 https://doi.org/10.1198/jasa.2009.tm08577

PENG, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statist. Probab. Lett.* **43** 399–409. MR1707950 https://doi.org/10.1016/S0167-7152(98)00280-6

PENG, L. and QI, Y. (2008). Bootstrap approximation of tail dependence function. *J. Multivariate Anal.* **99** 1807–1824. MR2444820 https://doi.org/10.1016/j.jmva.2008.01.018

POON, S.-H., ROCKINGER, M. and TAWN, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Rev. Financ. Stud.* **17** 581–610.

RAMOS, A. and LEDFORD, A. (2009). A new class of models for bivariate joint tails. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 219–241. MR2655531 https://doi.org/10.1111/j.1467-9868.2008.00684.x

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes. Applied Probability. A Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810 https://doi.org/10.1007/978-0-387-75953-1

RESNICK, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5** 303–336. MR2002121 https://doi.org/10.1023/A:1025148622954

ROCKAFELLAR, R. T. (1970). *Convex Analysis. Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, N.J. MR0274683

SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44. MR1947786 https://doi.org/10.1023/A:1020977924878

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

TAWN, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* **75** 397–415. MR0967580 https://doi.org/10.1093/biomet/75.3.397

WADSWORTH, J. L. and TAWN, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika* **99** 253–272. MR2931252 https://doi.org/10.1093/biomet/asr080

WADSWORTH, J. L., TAWN, J. A., DAVISON, A. C. and ELTON, D. M. (2017). Modelling across extremal dependence classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 149–175. MR3597968 https://doi.org/10.1111/rssb.12157

WELLER, G. B. and COOLEY, D. (2014). A sum characterization of hidden regular variation with likelihood inference via expectation–maximization. *Biometrika* **101** 17–36. MR3180655 https://doi.org/10.1093/biomet/ast046

WESTRA, S. and SISSON, S. A. (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *J. Hydrol.* **406** 119–128.

ZHOU, C. (2010). Dependence structure of risk factors and diversification effects. *Insurance Math. Econom.* **46** 531–540. MR2642530 https://doi.org/10.1016/j.insmatheco.2010.01.010

ZSCHEISCHLER, J. and SENEVIRATNE, S. I. (2017). Dependence of drivers affects risks associated with compound events. *Sci. Adv.* **3** e1700263. https://doi.org/10.1126/sciadv.1700263