



Bayesian inference for multivariate extreme value distributions

Clément Dombry, Sebastian Engelke, Marco Oesting

► To cite this version:

Clément Dombry, Sebastian Engelke, Marco Oesting. Bayesian inference for multivariate extreme value distributions. *Electronic Journal of Statistics* , 2017, 11, pp.4813 - 4844. 10.1214/17-EJS1367 . hal-01771128

HAL Id: hal-01771128

<https://hal.science/hal-01771128>

Submitted on 19 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian inference for multivariate extreme value distributions

Clément Dombry

*Univ. Bourgogne Franche-Comté
Laboratoire de Mathématiques de Besançon
UMR CNRS 6623, 16 Route de Gray
25030 Besançon cedex, France
e-mail: clement.dombry@univ-fcomte.fr*

Sebastian Engelke

*École Polytechnique Fédérale de Lausanne
EPFL-FSB-MATHAA-STAT, Station 8
1015 Lausanne, Switzerland
e-mail: sebastian.engelke@epfl.ch*

Marco Oesting

*Universität Siegen
Department Mathematik
Walter-Flex-Str. 3
57068 Siegen, Germany
e-mail: oesting@mathematik.uni-siegen.de*

Abstract: Statistical modeling of multivariate and spatial extreme events has attracted broad attention in various areas of science. Max-stable distributions and processes are the natural class of models for this purpose, and many parametric families have been developed and successfully applied. Due to complicated likelihoods, the efficient statistical inference is still an active area of research, and usually composite likelihood methods based on bivariate densities only are used. Thibaud et al. (2016) use a Bayesian approach to fit a Brown–Resnick process to extreme temperatures. In this paper, we extend this idea to a methodology that is applicable to general max-stable distributions and that uses full likelihoods. We further provide simple conditions for the asymptotic normality of the median of the posterior distribution and verify them for the commonly used models in multivariate and spatial extreme value statistics. A simulation study shows that this point estimator is considerably more efficient than the composite likelihood estimator in a frequentist framework. From a Bayesian perspective, our approach opens the way for new techniques such as Bayesian model comparison in multivariate and spatial extremes.

MSC 2010 subject classifications: Primary 62F15; secondary 60G70, 62F12.

Keywords and phrases: Asymptotic normality, Bayesian statistics, efficient inference, full likelihood, Markov chain Monte Carlo, max-stability, multivariate extremes.

Received August 2017.

1. Introduction

Extremes and the impacts of rare events have been brought into public focus in the context of climate change or financial crises. The temporal or spatial concurrence of several such events has often shown to be most catastrophic. Arising naturally as limits of rescaled componentwise maxima of random vectors, max-stable distributions are frequently used to describe this joint behavior of extremes. The generalization to continuous domains gives rise to max-stable processes that have become popular models in spatial extreme value statistics (e.g., Davison and Gholamrezaee, 2012), and are applied in various fields such as meteorology (Buishand et al., 2008; Engelke et al., 2015; Einmahl et al., 2016) and hydrology (Asadi et al., 2015).

For a k -dimensional max-stable random vector $Z = (Z_1, \dots, Z_k)$ with unit Fréchet margins, there exists an exponent function V describing the dependence between the components of Z such that $\mathbb{P}[Z \leq z] = \exp\{-V(z)\}$, $z \in (0, \infty)^k$. Many parametric models $\{F_\theta, \theta \in \Theta\}$ for the distribution function of Z have been proposed (cf. Schlather, 2002; Boldi and Davison, 2007; Kabluchko et al., 2009; Opitz, 2013), but likelihood-based inference remains challenging. The main reason is the lack of simple forms of the likelihood $L(z; \theta)$ in these models, which, by Faà di Bruno's formula, is given by

$$L(z; \theta) = \sum_{\tau \in \mathcal{P}_k} L(z, \tau; \theta) = \sum_{\tau \in \mathcal{P}_k} \exp\{-V(z)\} \prod_{j=1}^{|\tau|} \{-\partial_{\tau_j} V(z)\}, \quad (1)$$

where \mathcal{P}_k is the set of all partitions $\tau = \{\tau_1, \dots, \tau_{|\tau|}\}$ of $\{1, \dots, k\}$ and $\partial_{\tau_j} V(\cdot; \theta)$ denotes the partial derivative of the exponent function $V = V_\theta$ of F_θ with respect to the variables z_i , $i \in \tau_j$. The fact that the cardinality of \mathcal{P}_k is the k th Bell number that grows super-exponentially in the dimension k inhibits the use of the maximum likelihood methods based on $L(z; \theta)$ in (1).

The most common way to avoid this problem is to maximize the composite pairwise likelihood that relies only on the information in bivariate sub-vectors of Z (Padoan et al., 2010). Apart from the fact that this likelihood is misspecified, there might also be considerable losses in efficiency by using the composition of bivariate likelihoods instead of the full likelihood $L(z; \theta)$. To reduce this efficiency loss, higher order composite likelihood has been considered (Genton et al., 2011; Huser and Davison, 2013; Castruccio et al., 2016).

In practice, to obtain observations from the random variable Z , the data, typically a multivariate time series, is split into disjoint blocks and a max-stable distribution is fitted to the componentwise maxima within each block. To increase the efficiency, not only the block maxima but additional information from the time series can be exploited. The componentwise occurrence times of the maxima within each block lead to a partition τ of $\{1, \dots, k\}$ with indices belonging to the same subset if and only if the maxima in this component occurred at the same time. The knowledge of this partition makes inference much easier, as a single summand $L(z, \tau; \theta)$ in the full likelihood $L(z; \theta)$ given in (1) corresponds to the likelihood contribution of the specific partition τ . This

joint likelihood $L(z, \tau; \theta)$ was introduced in Stephenson and Tawn (2005b) and is tractable for many extreme value models and, consequently, can be used for inference if occurrence times are available.

In real data applications, however, the distribution of the block maxima is only approximated by a max-stable distribution and the distribution of the observed partitions of occurrence times are only approximations to the limiting distribution (as the block size tends to infinity) given by the likelihood $L(z, \tau; \theta)$. This approximation introduces a significant bias in the Stephenson–Tawn estimator and a bias correction has been proposed in Wadsworth (2015).

In many cases, only observations $z^{(1)}, \dots, z^{(N)} \in \mathbb{R}^k$ of the random max-stable vector Z are available, but there is no information about the corresponding partitions $\tau^{(1)}, \dots, \tau^{(N)}$. In this case, the Stephenson–Tawn likelihood cannot be used since the partition information is missing. In the context of conditional simulation of max-stable processes, Dombry et al. (2013) proposed a Gibbs sampler to obtain conditional samples of $\tau^{(l)}$ given the observation $z^{(l)}$, $l = 1, \dots, N$. Thibaud et al. (2016) use this approach to treat the missing partitions as latent variables in a Bayesian framework to estimate the parameters of a Brown–Resnick model (cf., Kabluchko et al., 2009) for extreme temperature. They obtain samples from the posterior distribution

$$L\left(\theta, \{\tau^{(l)}\}_{l=1}^N \mid \{z^{(l)}\}_{l=1}^N\right) \propto \pi_\theta(\theta) \prod_{l=1}^N L(z^{(l)}, \tau^{(l)}; \theta), \quad (2)$$

via a Markov chain Monte Carlo algorithm, where π_θ is the prior distribution on Θ .

In this paper we extend the Bayesian approach to general max-stable distributions and provide various examples of parametric models F_θ where it can be applied. The first focus is to study the statistical efficiency of the point estimators obtained as the median of the posterior distribution (2). This frequentist perspective allows to compare the efficiency of the Bayesian estimator that uses the full likelihoods to other frequentist estimators. A simulation study shows a substantial improvement of the estimation error when using full likelihoods rather than the commonly used pairwise likelihood estimator of Padoan et al. (2010).

From the Bayesian perspective, this approach opens up many new possibilities for Bayesian techniques in multivariate extreme value statistics. Besides readily available credible intervals, we discuss how Bayesian model comparison can be implemented. Thanks to the full, well-specified likelihoods in our approach, no adjustment of the posterior distribution as in the composite pairwise likelihood methods (Ribatet et al., 2012) is required.

Finally, we note that Dombry et al. (2017b) follow a complementary approach to ours where they apply an expectation-maximization algorithm to use full likelihoods $L(z; \theta)$ in the frequentist framework. The large sample asymptotic behavior of the frequentist and Bayesian estimators are the same (see section 3 below) but the Monte-Carlo Markov Chain computation of the Bayesian estimator offers better convergence guarantees than the expectation-maximization

computation of the maximum likelihood estimator. Moreover, alternatively to the perspective of max-stability and block maxima, inference can be based on threshold exceedances (Engelke et al., 2014; Wadsworth and Tawn, 2014; Thibaud and Opitz, 2015) and the corresponding multivariate Pareto distributions (Rootzén and Tajvidi, 2006; Rootzén et al., 2017).

The paper is organized as follows. In Section 2 we provide some background on max-stable distributions and their likelihoods, and we present the general methodology for the Bayesian full-likelihood approach. Section 3 develops an asymptotic theory for the resulting estimator. We show in Section 4 that our method and the asymptotic theory are applicable for the popular models in multivariate and spatial extremes, including the Brown–Resnick and extremal- t processes. The simulation studies in Section 5 quantify the finite-sample efficiency gains of the Bayesian approach when used as a frequentist point estimator of the extremal dependence parameters. Interestingly, this advantage persists when the dependence is a nuisance parameter and one is only interested in estimating marginal parameters (Section 5.3), at least in the case of a well-specified model. The posterior distribution and genuinely Bayesian techniques are studied in Section 6, with a focus on Bayesian model comparison. Section 7 concludes the paper with a discussion on computational aspects.

2. Methodology

In Section 2.1 we review some facts on max-stable distributions and their likelihoods. We describe the general setup of our approach and review the Markov chain Monte Carlo algorithm from Thibaud et al. (2016) and the Gibbs sampler from Dombry et al. (2013) in Section 2.2.

2.1. Max-stable distributions, partitions and joint likelihoods

Let us assume from now on that the max-stable vector Z belongs to a parametric family $\{F_\theta, \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^p$ is the parameter space, and that it admits a density f_θ . The exponent function of F_θ is $V_\theta(z) = -\log F_\theta(z)$. If there is no confusion we might omit the dependence on θ for simplicity.

Recall that if Z has standard Fréchet margins, it can be represented as the componentwise maximum

$$Z_i = \max_{j \in \mathbb{N}} \psi_i^{(j)}, \quad i = 1, \dots, k, \quad (3)$$

where $\{\psi^{(j)} : j \in \mathbb{N}\}$ is a Poisson point process on $E = [0, \infty)^k \setminus \{0\}$ with intensity measure Λ such that $\Lambda(E \setminus [0, z]) = V(z)$. For more details and an exact simulation method of Z via this representation, we refer to Dombry et al. (2016).

Analogously to the occurrence times in case of block maxima, the Poisson point process induces a random limit partition T of the index set $\{1, \dots, k\}$, where two indices $i_1 \neq i_2$ belong to the same subset if and only if $Z_{i_1} = \psi_{i_1}^{(j)}$

and $Z_{i_2} = \psi_{i_2}^{(j)}$ for the same $j \in \mathbb{N}$ (Dombry and Éyi-Minko, 2013). The joint likelihood of the max-stable vector Z and the limit partition T under the model F_θ satisfies

$$L(z, \tau; \theta) = \exp\{-V(z)\} \prod_{j=1}^{|\tau|} \{-\partial_{\tau_j} V(z)\}, \quad z \in (0, \infty)^k, \quad \tau \in \mathcal{P}_k, \quad (4)$$

and it equals the likelihood introduced in Stephenson and Tawn (2005b). This fact provides another interpretation of Equation (1), namely that the likelihood of Z is the integrated joint likelihood of Z and T .

In Dombry et al. (2017a) it has been shown that the existence of a density for the simple max-stable random vector Z with exponent measure Λ is equivalent to the existence of a density λ_I for the restrictions of Λ to the different faces $E_I \subset E$ defined by

$$E_I = \{z \in E; z_i > 0 \text{ for } i \in I \text{ and } z_i = 0 \text{ for } i \notin I\}, \quad \emptyset \neq I \subset \{1, \dots, k\},$$

that is,

$$\Lambda(A) = \sum_{\emptyset \neq I \subset \{1, \dots, k\}} \int_{\{z_I: z \in A \cap E_I\}} \lambda_I(z_I) \mu_I(dz_I),$$

Thus, the Stephenson–Tawn likelihood $L(z, \tau; \theta)$ can be rewritten as

$$L(z, \tau; \theta) = \exp\{-V(z)\} \prod_{j=1}^{\ell} \omega(\tau_j, z), \quad (5)$$

where

$$\omega(\tau_j, z) = \sum_{\tau_j \subset I \subset \{1, \dots, k\}} \int_{(0, z_{\tau_j^c \cap I})} \lambda_I(z_{\tau_j}, u_j) du_j, \quad (6)$$

and τ_1, \dots, τ_ℓ denote the $\ell = |\tau|$ different blocks of the partition τ , and z_{τ_j} and $z_{\tau_j^c}$ are the restrictions of z to τ_j and $\tau_j^c = \{1, \dots, k\} \setminus \tau_j$, respectively.

Equation (5) provides a formula for the joint likelihood of max-stable distributions with unit Fréchet margins and its partition. From this we can deduce a formula for the joint likelihood of a general max-stable distribution that admits a density. More precisely, let \bar{Z} be a k -dimensional max-stable random vector whose i th component, $i = 1, \dots, k$, has a generalized extreme value distribution with parameters $(\mu_i, \sigma_i, \xi_i) \in \mathbb{R} \times (0, \infty) \times \mathbb{R}$, that is,

$$\mathbb{P}(\bar{Z}_i \leq z_i) = \exp \left\{ - \left(1 + \xi_i \frac{z_i - \mu_i}{\sigma_i} \right)_+^{-1/\xi_i} \right\}, \quad z_i \in \mathbb{R}.$$

Then, $U_i(\bar{Z}_i)$ has unit Fréchet distribution where U_i denotes the marginal transformation

$$U_i(x) = \left(1 + \xi_i \frac{x - \mu_i}{\sigma_i} \right)_+^{1/\xi_i}, \quad 1 + \xi_i \frac{x - \mu_i}{\sigma_i} > 0.$$

For a vector $z_I = (z_i)_{i \in I}$ with $I \subset \{1, \dots, k\}$, we define $U(z_I) = (U_i(z_i))_{i \in I}$, such that $\mathbb{P}(\bar{Z} \leq z) = \exp\{-V[U(z)]\}$, where V is the exponent measure of the normalized max-stable distribution $U(\bar{Z})$. Consequently, the joint density of the general max-stable vector \bar{Z} and the limit partition T is

$$L(z, \tau; \theta) = \exp\{-V[U(z)]\} \cdot \left(\prod_{j=1}^{\ell} \omega(\tau_j, U(z)) \right) \cdot \left(\prod_{i=1}^k \frac{1}{\sigma_i} U_i(z_i)^{1-\xi_i} \right), \quad (7)$$

for $z \in (0, \infty)^k$ such that $1 + \xi_i(z_i - \mu_i)/\sigma_i > 0$, $i = 1, \dots, k$ and $\tau = \{\tau_1, \dots, \tau_\ell\} \in \mathcal{P}_k$.

2.2. Bayesian inference and Markov chain Monte Carlo

Extreme value statistics is concerned with the estimation and uncertainty quantification of the parameter vector $\theta \in \Theta$. Here, θ might include both marginal and dependence parameters of the max-stable model. In a Bayesian setup we introduce a prior $\pi_\theta(\theta)$ on the parameter space Θ . Given independent data $z^{(1)}, \dots, z^{(N)} \in \mathbb{R}^k$ from the max-stable distribution $Z \sim F_\theta$, we are interested in the posterior distribution of the parameter θ conditional on the data. As explained in Section 1, the complex structure of the full likelihood $L(\{z^{(l)}\}; \theta) = \prod_{l=1}^N L(z^{(l)}; \theta)$ prevents a direct assessment of the posterior distribution, which is proportional to the product of $L(\{z^{(l)}\}; \theta)$ and the prior density $\pi_\theta(\theta)$. Instead, we introduce the corresponding limit partitions $T^{(1)}, \dots, T^{(N)}$ as latent variables and sample from the joint distribution of $(\theta, T^{(1)}, \dots, T^{(N)})$ conditional on the data $z^{(1)}, \dots, z^{(N)}$, which is given in Equation (2).

It is customary to use Monte Carlo Markov Chain methods to sample from a target distribution which is known up to a multiplicative constant only. The aim is to construct a Markov chain which possesses the target distribution as stationary distribution and has good mixing properties. To this end, in each step of the Markov chain, the parameter vector θ and the partitions $T^{(1)}, \dots, T^{(N)}$ are updated separately by the Metropolis–Hastings algorithm and a Gibbs sampler, respectively (cf., Thibaud et al., 2016).

For fixed partitions $T^{(l)} = \tau^{(l)}$, $l = 1, \dots, N$, and the current state θ for the parameter vector, we propose a new state θ^* according to a probability density $q(\theta, \cdot)$ which satisfies $q(\theta_1, \theta_2) > 0$ if and only if $q(\theta_2, \theta_1) > 0$ for $\theta_1, \theta_2 \in \Theta$. The proposal is accepted, that is, θ is updated to θ^* , with probability

$$a(\theta, \theta^*) = \min \left\{ \frac{\prod_{l=1}^N L(z^{(l)}, \tau^{(l)}; \theta^*) \pi_\theta(\theta^*) q(\theta^*, \theta)}{\prod_{l=1}^N L(z^{(l)}, \tau^{(l)}; \theta) \pi_\theta(\theta) q(\theta, \theta^*)}, 1 \right\} \quad (8)$$

where $L(z, \tau; \theta)$ is given by (7). In general, there are various ways of choosing an appropriate proposal density q . For instance, it might be advisable to update the vector θ component by component. It has to be ensured that any state θ_2 with positive posterior density can be reached from any other state θ_1 with positive posterior density in a finite number of steps, that is, that the Markov chain is

irreducible. The convergence of the Markov chain to its stationary distribution (2) is then guaranteed. Note that the framework described above enables estimation of marginal and dependence parameters simultaneously. In particular, it allows for response surface methodology such as (log-)linear models for the marginal parameters.

For a fixed parameter vector $\theta \in \Theta$ we use the Gibbs sampler in Dombry et al. (2013) to update the current states of the partitions $\tau^{(1)}, \dots, \tau^{(N)} \in \mathcal{P}_k$ conditional on the data $z^{(1)}, \dots, z^{(N)}$. Thanks to independence, for each $l = 1, \dots, N$, we can update $\tau = \tau^{(l)}$ conditional on $z = z^{(l)}$ separately, where the conditional distribution is

$$L(\tau \mid z; \theta) = \frac{L(z, \tau; \theta)}{\sum_{\tau' \in \mathcal{P}_k} L(z, \tau'; \theta)} = \frac{1}{C_z} \prod_{j=1}^{\ell} \omega\{\tau_j, U(z)\}, \quad (9)$$

with C_z the normalization constant

$$C_z = \sum_{\tau \in \mathcal{P}_k} \prod_{j=1}^{\ell} \omega\{\tau_j, U(z)\}.$$

For $i \in \{1, \dots, k\}$, let τ_{-i} be the restriction of τ to the set $\{1, \dots, k\} \setminus \{i\}$. As usual with Gibbs samplers, our goal is to simulate from

$$\mathbb{P}_{\theta}(T = \cdot \mid T_{-i} = \tau_{-i}, Z = z), \quad (10)$$

where τ is the current state of the Markov chain and \mathbb{P}_{θ} denotes the probability under the assumption that Z follows the law F_{θ} . It is easy to see that the number of possible updates according to (10) is always less than k , so that a combinatorial explosion is avoided. Indeed, the index i can be reallocated to any of the components of τ_{-i} or to a new component with a single point: the number of possible updates $\tau^* \in \mathcal{P}_k$ such that $\tau_{-i}^* = \tau_{-i}$ equals ℓ if $\{i\}$ is a partitioning set of τ , and $\ell + 1$ otherwise.

The distribution (10) has nice properties. From (9), we obtain that

$$\mathbb{P}_{\theta}(T = \tau^* \mid T_{-i} = \tau_{-i}, Z = z) = \frac{L(z, \tau^*)}{\sum_{\tau' \in \mathcal{P}_k} L(z, \tau') 1_{\{\tau'_{-i} = \tau_{-i}\}}} \propto \frac{\prod_{j=1}^{|\tau^*|} w\{\tau_j^*, U(z)\}}{\prod_{j=1}^{|\tau|} w\{\tau_j, U(z)\}}. \quad (11)$$

for all $\tau^* \in \mathcal{P}_k$ with $\tau_{-i}^* = \tau_{-i}$. Since τ and τ^* share many components, all the factors in the right-hand side of (11) cancel out except at most four of them. This makes the Gibbs sampler particularly convenient.

We suggest a random scan implementation of the Gibbs sampler, meaning that one iteration of the Gibbs sampler selects randomly an element $i \in \{1, \dots, k\}$ and then updates the current state τ according to the proposal distribution (10). For the sake of simplicity, we use the uniform random scan, i.e., i is selected according to the uniform distribution on $\{1, \dots, k\}$.

3. Asymptotic results

In the previous section, we presented a procedure that allows to sample from the posterior distribution of the parameter θ of a parametric model $\{f_\theta, \theta \in \Theta\}$ given a sample of N observations. In this section, we will discuss the asymptotic properties of the posterior distribution as the sample size N tends to ∞ .

The asymptotic analysis of Bayes procedures usually relies on the Bernstein–von Mises theorem which allows for an asymptotic normal approximation of the posterior distribution of $\sqrt{N}(\theta - \theta_0)$, given the observations $z^{(1)}, \dots, z^{(N)}$ from the parametric model f_{θ_0} . The theorem then implies the asymptotic normality and efficiency of Bayesian point estimators such as the posterior mean or posterior median with the same asymptotic variance as the maximum likelihood estimator.

A key assumption is that for every $\varepsilon > 0$ there exists a sequence of uniformly consistent tests $\phi_N = \phi_N(z^{(1)}, \dots, z^{(N)}) \in \{0, 1\}$ for testing the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \|\theta - \theta_0\|_\infty \geq \varepsilon$, where H_0 is rejected if and only if $\phi_N = 1$. The uniformity means that

$$\mathbb{P}_{\theta_0}(\phi_N = 1) \rightarrow 0 \quad \text{and} \quad \sup_{\|\theta - \theta_0\|_\infty \geq \varepsilon} \mathbb{P}_\theta(\phi_N = 0) \rightarrow 0 \quad \text{as } N \rightarrow \infty. \quad (12)$$

where \mathbb{P}_θ denotes the probability measure induced by N independent copies of $Z \sim f_\theta$.

Theorem 1 (Bernstein-von Mises, Theorems 10.1 and 10.8 in van der Vaart (1998)). *Let the parametric model $\{f_\theta, \theta \in \Theta\}$ be differentiable in quadratic mean at θ_0 with non-singular Fisher information matrix I_{θ_0} , and assume that the mapping $\theta \mapsto \sqrt{f_\theta(z)}$ is differentiable at θ_0 for f_{θ_0} -almost every z . For every $\varepsilon > 0$, suppose there exists a sequence of uniformly consistent tests ϕ_N as in (12). Suppose further that the prior distribution $\pi_{\text{prior}}(d\theta)$ is absolutely continuous in a neighborhood of θ_0 with a continuous positive density at θ_0 . Then, under the distribution f_{θ_0} , the posterior distribution satisfies*

$$\left\| \pi_{\text{post}}(d\theta \mid z^{(1)}, \dots, z^{(N)}) - \mathcal{N}\left(\theta_0 + N^{-1/2} \Delta_{N, \theta_0}, N^{-1} I_{\theta_0}^{-1}\right) \right\|_{TV} \xrightarrow{d} 0 \quad \text{as } N \rightarrow \infty,$$

where $\Delta_{N, \theta_0} = N^{-1/2} \sum_{i=1}^N I_{\theta_0}^{-1} \partial_\theta \log f_{\theta_0}(z^{(i)})$ and $\|\cdot\|_{TV}$ is the total variation distance.

As a consequence, if the prior distribution $\pi_{\text{prior}}(d\theta)$ has a finite mean, the posterior median $\hat{\theta}_n^{\text{Bayes}}$ is asymptotically normal and efficient, that is, it satisfies

$$\sqrt{N}(\hat{\theta}_n^{\text{Bayes}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I_{\theta_0}^{-1}), \quad \text{as } N \rightarrow \infty.$$

In order to apply this theorem to max-stable distributions, two main assumptions are required: the differentiability in quadratic mean of the statistical model and the existence of uniformly consistent tests satisfying (12). Differentiability

in quadratic mean is a technical condition that imposes a certain regularity on the likelihood f_{θ_0} . For the case of multivariate max-stable models this property has been considered in detail in Dombry et al. (2017a), where equivalent conditions on the exponent function and the spectral density are given.

We now discuss the existence of uniformly consistent tests and propose a criterion based on pairwise extremal coefficients. This criterion turns out to be simple and general enough since it applies for most of the standard models in extreme value theory. Indeed, in many cases, pairwise extremal coefficients can be explicitly computed and allow for identifying the parameter θ .

For a max-stable vector Z with unit Fréchet margins, the pairwise extremal coefficient $\eta_{i_1, i_2} \in [1, 2]$ between margins $1 \leq i_1 < i_2 \leq k$ is defined by

$$\mathbb{P}(Z_{i_1} \leq z, Z_{i_2} \leq z) = \exp\{-\eta_{i_1, i_2}/z\}, \quad z > 0.$$

It is the scale exponent of the unit Fréchet variable $Z_{i_1} \vee Z_{i_2}$ and hence satisfies

$$\eta_{i_1, i_2} = \left(\mathbb{E} \left[\frac{1}{Z_{i_1} \vee Z_{i_2}} \right] \right)^{-1}.$$

In the case that Z follows the distribution f_{θ} , we write $\eta_{i_1, i_2}(\theta)$ for the associated pairwise extremal coefficient.

Proposition 1. *Let $\theta_0 \in \Theta$ and $\varepsilon > 0$. Assume that*

$$\inf_{\|\theta - \theta_0\|_{\infty} \geq \varepsilon} \max_{1 \leq i_1 < i_2 \leq k} |\eta_{i_1, i_2}(\theta) - \eta_{i_1, i_2}(\theta_0)| > 0. \quad (13)$$

Then there exists a uniformly consistent sequence of tests ϕ_N satisfying (12).

Remark 1. *The identifiability of the model parameters $\theta \in \Theta$ through the pairwise extremal coefficients $\eta_{i_1, i_2}(\theta)$, $1 \leq i_1 < i_2 \leq k$ is a direct consequence of Equation (13).*

Remark 2. *If $\theta = (\theta_1, \dots, \theta_p) \in \Theta$, and for any $1 \leq j \leq p$ there exists $1 \leq i_1 < i_2 \leq k$, such that $\eta_{i_1, i_2}(\theta)$ depends only on θ_j and it is strictly monotone with respect to this component, then Equation (13) is satisfied.*

Proof of Proposition 1. For a random vector Z with distribution f_{θ} and $1 \leq i_1 < i_2 \leq k$, the random variable $1/(Z_{i_1} \vee Z_{i_2})$ follows an exponential distribution with parameter $\eta_{i_1, i_2}(\theta) \in [1, 2]$ and variance $\eta_{i_1, i_2}^{-2}(\theta) \in [1/4, 1]$. Hence,

$$T_{i_1, i_2}^{-1} = \frac{1}{N} \sum_{i=1}^N \frac{1}{Z_{i_1}^{(i)} \vee Z_{i_2}^{(i)}}$$

is an unbiased estimator of $\eta_{i_1, i_2}^{-1}(\theta)$ with variance less than or equal to $1/N$. Chebychev's inequality entails

$$\mathbb{P}_{\theta}(|T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta)| > \delta) \leq \frac{1}{N\delta^2}, \quad \text{for all } \delta > 0. \quad (14)$$

Define the test ϕ_N by

$$\phi_N = \begin{cases} 0 & \text{if } \max_{1 \leq i_1 < i_2 \leq k} |T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta_0)| \leq \delta, \\ 1 & \text{otherwise.} \end{cases}$$

We prove below that, for $\delta > 0$ small enough, the sequence ϕ_N satisfies (12). For $\theta = \theta_0$, the union bound together with Eq. (14) yield

$$\mathbb{P}_{\theta_0}(\phi_N = 1) \leq \sum_{1 \leq i_1 < i_2 \leq k} \mathbb{P}_{\theta_0}(|T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta_0)| > \delta) \leq \frac{k(k-1)}{2N\delta^2} \rightarrow 0,$$

as $N \rightarrow \infty$. On the other hand, Eq. (13) implies that there is some $\gamma > 0$ such that

$$\max_{1 \leq i_1 < i_2 \leq k} |\eta_{i_1, i_2}^{-1}(\theta) - \eta_{i_1, i_2}^{-1}(\theta_0)| \geq \gamma, \quad \text{for all } \|\theta - \theta_0\|_\infty \geq \varepsilon.$$

Let $\theta \in \Theta$ be such that $\|\theta - \theta_0\|_\infty \geq \varepsilon$ and consider $1 \leq i_1 < i_2 \leq k$ realizing the maximum in the above equation. By the triangle inequality,

$$\begin{aligned} |T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta)| &\geq |\eta_{i_1, i_2}^{-1}(\theta) - \eta_{i_1, i_2}^{-1}(\theta_0)| - |T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta_0)| \\ &\geq \gamma - |T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta_0)|, \end{aligned}$$

so that, on the event $\{\phi_N = 0\} \subset \{|T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta_0)| \leq \delta\}$, we have

$$|T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta)| \geq \gamma - \delta.$$

Applying Eq. (14) again, we deduce, for $\delta \in (0, \gamma)$,

$$\mathbb{P}_\theta(\phi_N = 0) \leq \mathbb{P}_\theta(|T_{i_1, i_2}^{-1} - \eta_{i_1, i_2}^{-1}(\theta)| \geq \gamma - \delta) \leq \frac{1}{N(\gamma - \delta)^2}.$$

Since the upper bound goes to 0 uniformly in θ with $\|\theta - \theta_0\|_\infty \geq \varepsilon$, as $N \rightarrow \infty$, this proves Eq. (12). \square

4. Examples

In the previous sections we discussed the practical implementation of a Markov chain Monte Carlo algorithm to obtain samples from the posterior distribution $L(\theta, \{\tau^{(l)}\}_{l=1}^N | \{z^{(l)}\}_{l=1}^N)$ and its asymptotic behavior as $N \rightarrow \infty$. The only model-specific quantity needed to run the algorithm are the weights $\omega(\tau_j, z)$ in (6). In this section, we provide explicit formulas for these weights for several classes of popular max-stable models and prove that the models satisfy the assumptions of the Bernstein–von Mises theorem; see Theorem 1. It follows that the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient for these models.

For the calculation of the weights $\omega(\tau_j, z)$, we first note that all of the examples in this section admit densities as a simple consequence of Prop. 2.1

in Dombry et al. (2017a). We further note that, for the models considered in Subsections 4.1-4.4, we have $\lambda_I \equiv 0$ for all $I \subsetneq \{1, \dots, k\}$, i.e.

$$\Lambda(A) = \int_A \lambda(z) \mu(dz), \quad A \subset [0, \infty)^k \setminus \{0\},$$

and, consequently, Equation (6) simplifies to

$$\omega(\tau_j, z) = \int_{(0, z_{\tau_j^c})} \lambda(z_{\tau_j}, u_j) du_j.$$

For the posterior median $\hat{\theta}_n^{Bayes}$, in the sequel, we will always assume that the prior distribution is absolutely continuous with strictly positive density in a neighborhood of θ_0 , and that it has finite mean. Given the differentiability in quadratic mean of the model, it suffices to verify condition (13) in Prop. 1. This implies the existence of a uniformly consistent sequence of tests and, by Remark 1, the identifiability of the model. Theorem 1 then ensures asymptotic normality of the posterior median.

Analogously to the notation $z_I = (z_i)_{i \in I}$ for a vector $z \in \mathbb{R}^k$ and an index set $\emptyset \neq I \subset \{1, \dots, k\}$, we write $A_{I,J} = (A_{ij})_{i \in I, j \in J}$ for a matrix $A = (A_{ij})_{1 \leq i, j \leq k}$ and index sets $\emptyset \neq I, J \subset \{1, \dots, k\}$. Proofs for the results presented in this section can be found in Appendix A.

4.1. The logistic model

One of the simplest multivariate extreme value distributions is the logistic model where

$$V(z) = \left(z_1^{-1/\theta} + \dots + z_k^{-1/\theta} \right)^\theta, \quad \theta \in (0, 1). \quad (15)$$

The logistic model is symmetric in its variables and interpolates between independence as $\theta \uparrow 1$ and complete dependence as $\theta \downarrow 0$.

Proposition 2. *Let $\tau = (\tau_1, \dots, \tau_\ell) \in \mathcal{P}_k$ and $z \in E$. The weights $\omega(\tau_j, z)$ in (6) for the logistic model with exponent measure (15) are*

$$\omega(\tau_j, z) = \theta^{-|\tau_j|+1} \frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1 - \theta)} \left(\sum_{i=1}^k z_i^{-1/\theta} \right)^{\theta - |\tau_j|} \prod_{i \in \tau_j} z_i^{-1-1/\theta}. \quad (16)$$

Remark 3. *From (16), it can be seen that we can also write*

$$L(\tau, z) = \exp(-V(z)) \left(\prod_{i=1}^k z_i^{-1-1/\theta} \right) \left(\sum_{i=1}^k z_i^{-1/\theta} \right)^{-k} \theta^{-k} \prod_{j=1}^\ell \tilde{\omega}(\tau_j, z)$$

with

$$\tilde{\omega}(\tau_j, z) = \theta \frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1 - \theta)} \left(\sum_{i=1}^k z_i^{-1/\theta} \right)^\theta.$$

This suggests to use the simplified weights $\tilde{\omega}$ for the Gibbs sampler.

Proposition 3. *For the logistic model with $\theta_0 \in (0, 1)$, the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient as $N \rightarrow \infty$.*

4.2. The Dirichlet model

The Dirichlet model (Coles and Tawn, 1991) is defined by its spectral density h on the simplex $S^{k-1} = \{w \in [0, \infty)^k : w_1 + \dots + w_k = 1\}$. For parameters $\alpha_1, \dots, \alpha_k > 0$, it is given by

$$h(w) = \frac{1}{k} \frac{\Gamma(1 + \sum_{i=1}^k \alpha_i)}{(\sum_{i=1}^k \alpha_i w_i)^{k+1}} \prod_{i=1}^k \frac{\alpha_i}{\Gamma(\alpha_i)} \left(\frac{\alpha_i w_i}{\sum_{j=1}^k \alpha_j w_j} \right)^{\alpha_i - 1}, \quad w \in S^{k-1}, \quad (17)$$

and it has no mass on lower-dimensional faces of S^{k-1} (Coles and Tawn, 1991). Equivalently, the exponent function of the Dirichlet model is given by

$$V(z) = k \mathbb{E} \left[\max_{i=1, \dots, k} \frac{W_i}{z_i} \right],$$

where W is a random vector with density $h(w)$.

Proposition 4. *Let $\tau = (\tau_1, \dots, \tau_\ell) \in \mathcal{P}_k$ and $z \in E$. The weights $\omega(\tau_j, z)$ in (6) for the Dirichlet model with spectral density (17) are*

$$\omega(\tau_j, z) = \prod_{i \in \tau_j} \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \int_0^\infty e^{-\frac{1}{r} \sum_{i \in \tau_j} \alpha_i z_i} \left(\prod_{i \in \tau_j^c} F_{\alpha_i}(\alpha_i z_i / r) \right) r^{-2 - \sum_{i=1}^k \alpha_i} dr, \quad (18)$$

where

$$F_\alpha(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt$$

is the distribution function of a Gamma variable with shape $\alpha > 0$.

Proposition 5. *Consider the Dirichlet model with $\theta_0 = (\alpha_1, \dots, \alpha_k) \in \Theta = (0, \infty)^k$. For $k \geq 3$ and almost every $\theta_0 \in \Theta$, the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient as $N \rightarrow \infty$.*

Remark 4. *We believe that the result for the posterior median holds true even for every $\theta_0 \in \Theta$. In the proof of Proposition 5, we need the partial derivatives of $(\alpha_1, \alpha_2) \mapsto \eta(\alpha_1, \alpha_2)$ to be negative, but this can only be concluded almost everywhere.*

4.3. The extremal- t model and the Schlather process

The extremal- t model (Nikoloulopoulos et al., 2009; Opitz, 2013) is given by an exponent measure of the form

$$V(z) = c_\nu \mathbb{E} \left[\max_{i=1, \dots, k} \frac{\max\{0, W_i\}^\nu}{z_i} \right], \quad (19)$$

where $(W_1, \dots, W_k)^\top$ is a standardized Gaussian vector with correlation matrix Σ , $c_\nu = \sqrt{\pi} 2^{-(\nu-2)/2} \Gamma\{(\nu+1)/2\}^{-1}$ and $\nu > 0$.

Proposition 6 ((Thibaud and Opitz, 2015)). *Let $\tau = (\tau_1, \dots, \tau_\ell) \in \mathcal{P}_k$ and $z \in E$. The weights $\omega(\tau_j, z)$ in (6) for the extremal- t model with exponent function (19) are*

$$\begin{aligned} \omega(\tau_j, z) = & T_{|\tau_j|+\nu} \left(z_{\tau_j^c}^{1/\nu} - \tilde{\mu}, \tilde{\Sigma} \right) \cdot \nu^{1-|\tau_j|} \cdot \pi^{(1-|\tau_j|)/2} \cdot \det(\Sigma_{\tau_j, \tau_j})^{-1/2} \\ & \cdot \frac{\Gamma\{(\nu + |\tau_j|)/2\}}{\Gamma\{(\nu + 1)/2\}} \cdot \prod_{i \in \tau_j} |z_i|^{1/\nu-1} \cdot \left\{ \left(z_{\tau_j}^{1/\nu} \right)^\top \Sigma_{\tau_j, \tau_j}^{-1} z_{\tau_j}^{1/\nu} \right\}^{-(\nu+|\tau_j|)/2} \end{aligned} \quad (20)$$

where $\tilde{\mu} = \Sigma_{\tau_j^c, \tau_j} \Sigma_{\tau_j}^{-1} z_{\tau_j}^{1/\nu}$,

$$\tilde{\Sigma} = (|\tau_j| + \nu)^{-1} \left(z_{\tau_j}^{1/\nu} \right)^\top \Sigma_{\tau_j}^{-1} z_{\tau_j} \left(\Sigma_{\tau_j^c} - \Sigma_{\tau_j^c, \tau_j} \Sigma_{\tau_j}^{-1} \Sigma_{\tau_j, \tau_j^c} \right)$$

and $T_k(\cdot; \Sigma)$ denotes a multivariate Student distribution function with k degrees of freedom and scale matrix Σ .

Proposition 7. *Consider the extremal- t model with $\theta_0 = (\Sigma, \nu)$ where Σ is a positive definite correlation matrix and $\nu > 0$. Then, for fixed $\nu > 0$ the posterior median $\hat{\theta}_N^{\text{Bayes}}$ is asymptotically normal and efficient as $N \rightarrow \infty$.*

Remark 5. *If ν is not fixed, then the parameter $\theta = (\Sigma, \nu)$ cannot be identified from the pairwise extremal coefficients and Equation (13) is not satisfied. The identifiability can still be shown by considering the behavior of the bivariate angular measure at the origin (Engelke and Ivanovs, 2017, Section A.3.3).*

A popular model in spatial extremes is the extremal- t process (Opitz, 2013), a max-stable process $\{Z(x), x \in \mathbb{R}^d\}$ whose finite-dimensional distributions $(Z(x_1), \dots, Z(x_k))^\top, x_1, \dots, x_k \in \mathbb{R}^d$ have an exponent function of the form (19) where the Gaussian vector is replaced by a standardized stationary Gaussian process $\{W(x), x \in \mathbb{R}^d\}$ evaluated at x_1, \dots, x_k . The correlation matrix Σ then has the form

$$\Sigma = \{\rho(x_i - x_j)\}_{1 \leq i, j \leq k},$$

where $\rho : \mathbb{R}^d \rightarrow [-1, 1]$ is the correlation function of the Gaussian process W . The special case $\nu = 1$ corresponds to the extremal Gaussian process (Schlather, 2002), also called Schlather process.

Corollary 1. *Let Z be the Schlather process on \mathbb{R}^d with correlation function ρ coming from the parametric family*

$$\rho(h) = \exp(-\|h\|_2^\alpha/s), \quad (s, \alpha) \in \Theta = (0, \infty) \times (0, 2].$$

Suppose that Z is observed at pairwise distinct locations $t_1, \dots, t_k \in \mathbb{R}^d$ such that not all pairs of locations have the same Euclidean distance. Then, the posterior median of $\theta = (s, \alpha)$ is asymptotically normal.

4.4. The Hüsler-Reiss model and the Brown-Resnick model

The Hüsler-Reiss distribution (cf., Hüsler and Reiss, 1989; Kabluchko et al., 2009) can be characterized by its exponent function

$$V(z) = \mathbb{E} \left[\max_{i=1, \dots, k} \frac{\exp \{W_i - \Sigma_{ii}/2\}}{z_i} \right], \quad (21)$$

where $W = (W_1, \dots, W_k)^\top$ is a Gaussian vector with expectation 0 and covariance matrix Σ . It can be shown that the exponent function can be parameterized by the matrix

$$\Lambda = \{\lambda_{i,j}^2\}_{1 \leq i, j \leq k} = \left\{ \frac{1}{4} \mathbb{E}(W_i - W_j)^2 \right\}_{1 \leq i, j \leq k}$$

as we have the equality

$$V(z) = \sum_{p=1}^k z_p^{-1} \Phi_{k-1} \left(2\lambda_{p,-p}^2 + \log(z_{-p}/z_p); \Sigma^{(p)} \right), \quad z \in (0, \infty)^k, \quad (22)$$

(cf. Nikoloulopoulos et al., 2009), where for $p = 1, \dots, k$, the matrix $\Sigma^{(p)}$ has (i, j) th entry $2(\lambda_{p,i}^2 + \lambda_{p,j}^2 - \lambda_{i,j}^2)$, $i, j \neq p$ and $\Phi_{k-1}(\cdot, \Sigma^{(p)})$ denotes the $(k-1)$ -dimensional normal distribution function with covariance matrix $\Sigma^{(p)}$.

Note that the positive definiteness of the matrices $\Sigma^{(p)}$, $p = 1, \dots, k$, follows from the fact that Λ is conditionally negative definite, i.e.

$$\sum_{1 \leq i, j \leq k} a_i a_j \lambda_{i,j}^2 \leq 0 \quad (23)$$

for all $a_1, \dots, a_k \in \mathbb{R}$ summing up to 0 (cf. Berg et al., 1984, Lem. 3.2.1). In the following, we will assume that Λ is even strictly positive definite, i.e. equality in (23) holds true if and only if $a_1 = \dots = a_k = 0$. Then, all the matrices $\Sigma_{I,I}^{(p)}$ with $p \in \{1, \dots, k\}$ and $\emptyset \neq I \subset \{1, \dots, k\}$ are strictly positive definite.

Proposition 8 ((Wadsworth and Tawn, 2014), (Asadi et al., 2015)). *Let $\tau = (\tau_1, \dots, \tau_\ell) \in \mathcal{P}_k$ and $z \in E$. For $j \in \{1, \dots, k\}$, choose any $p \in \tau_j$ and let $\tilde{\tau} = \tau_j \setminus \{p\}$, $\tilde{\tau}^c = \{1, \dots, k\} \setminus \tau_j$. The weights $\omega(\tau_j, z)$ in (6) for the Hüsler-Reiss distribution with exponent function (22) are*

$$\omega(\tau_j, z) = \frac{1}{z_p^2 \prod_{i \in \tilde{\tau}} z_i} \varphi_{|\tilde{\tau}|} \left\{ z_{\tilde{\tau}}^*; \Sigma_{\tilde{\tau}, \tilde{\tau}}^{(p)} \right\} \Phi_{|\tilde{\tau}^c|} \left\{ z_{\tilde{\tau}^c}^* - \Sigma_{\tilde{\tau}^c, \tilde{\tau}}^{(p)} (\Sigma_{\tilde{\tau}, \tilde{\tau}}^{(p)})^{-1} z_{\tilde{\tau}}^*; \hat{\Sigma}^{(p)} \right\}, \quad (24)$$

where

$$z^* = \left\{ \log \left(\frac{z_i}{z_p} \right) + \frac{\Gamma(x_i, x_p)}{2} \right\}_{i=1, \dots, k} \quad \text{and} \quad \hat{\Sigma}^{(p)} = \Sigma_{\tilde{\tau}^c, \tilde{\tau}^c}^{(p)} - \Sigma_{\tilde{\tau}^c, \tilde{\tau}}^{(p)} (\Sigma_{\tilde{\tau}, \tilde{\tau}}^{(p)})^{-1} \Sigma_{\tilde{\tau}, \tilde{\tau}^c}^{(p)}.$$

Here $\Phi_k(\cdot; \Sigma)$ denotes a k -dimensional Gaussian distribution function with mean 0 and covariance matrix Σ , and $\varphi_k(\cdot; \Sigma)$ its density. The functions Φ_0 and φ_0 are set to be constant 1.

Proposition 9. *Consider the Hüsler–Reiss model with $\theta_0 = \Lambda$ being a strictly conditionally negative definite matrix. Then, the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient as $N \rightarrow \infty$.*

Hüsler–Reiss distributions are the finite dimensional distributions of the max-stable Brown–Resnick process, a popular class in spatial extreme value statistics. Here, the Gaussian vectors $(W_1, \dots, W_k)^\top$ in (21) are the finite-dimensional distributions of a centered Gaussian process $\{W(x), x \in \mathbb{R}^d\}$ which is parameterized via a conditionally negative definite variogram $\gamma : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty)$, $\gamma(x_1, x_2) = \mathbb{E}(W(x_1) - W(x_2))^2$. If W has stationary increments, we have that $\gamma(x_1, x_2) = \gamma(x_1 - x_2, 0) =: \gamma(x_1 - x_2)$ and the resulting Brown–Resnick process is stationary (Brown and Resnick, 1977; Kabluchko et al., 2009). The most common parametric class of variograms belonging to Gaussian processes with stationary increments is the class of fractional variograms, which we consider in the following corollary.

Corollary 2. *Consider a Brown–Resnick process on \mathbb{R}^d with variogram coming from the parametric family*

$$\gamma(h) = \|h\|_2^\alpha / s, \quad (s, \alpha) \in \Theta = (0, \infty) \times (0, 2).$$

Suppose that the process is observed on a finite set of locations $t_1, \dots, t_m \in \mathbb{R}^d$ such that the pairwise Euclidean distances are not all equal. Then the posterior median of $\theta = (s, \alpha)$ is asymptotically normal.

5. Simulation study

Let $z^{(l)} = (z_1^{(l)}, \dots, z_k^{(l)})$, $l = 1, \dots, N$, be N realizations of a k -dimensional max-stable vector Z whose distribution belongs to some parametric family $\{F_\theta, \theta \in \Theta\}$. As described in Section 2, including the partition $\tau^{(l)}$ associated to a realization $z^{(l)}$ in a Bayesian framework allows to obtain samples from the posterior distribution $L(\theta \mid z^{(1)}, \dots, z^{(N)})$ of θ given the data. This procedure uses the full dependence information of the multivariate distribution Z . This is in contrast to frequentist maximum likelihood estimation for the max-stable vector Z , where even in moderate dimensions the likelihoods are too complicated for practical applications. Instead, at the price of likelihood misspecification, it is common practice to use only pairwise likelihoods which are assumed to be mutually independent. The maximum pairwise likelihood estimator (Padoan et al., 2010) is then

$$\hat{\theta}_{PL} = \arg \max_{\theta \in \Theta} \sum_{l=1}^N \sum_{1 \leq i < j \leq k} \log f_{\theta; i, j}(z_i^{(l)}, z_j^{(l)}), \quad (25)$$

where $f_{\theta; i, j}$ denotes the joint density of the i th and j th component of Z under the model F_θ . Using only bivariate information on the dependence results in efficiency losses.

In this section, we analyze the performance of our proposed Bayesian estimator and compare it to $\hat{\theta}_{\text{PL}}$ and other existing methods. Since the latter are all frequentist approaches, for a Markov chain whose stationary distribution is the posterior, we obtain a point estimator $\hat{\theta}_{\text{Bayes}}$ of θ as the posterior median, i.e.,

$$\hat{\theta}_{\text{Bayes}} = \text{median} \left\{ L(\theta | z^{(1)}, \dots, z^{(N)}) \right\}.$$

As the parametric model we choose the logistic distribution introduced in Subsection 4.1 with parameter space $\Theta = (0, 1)$ and uniform prior. This choice covers a range of situations from strong to very weak dependence. Other choices of parametric models will result in different efficiency gains but the general observations in the next sections should remain the same.

We note that other functionals of the posterior distribution can be used to obtain point estimators. Simulations based on the posterior mean, for instance, gave very similar results, and we therefore restrict to the posterior median in the sequel. Similarly, changing the prior distributions does not have a strong effect on the posterior distribution for the sample sizes we consider; see also Section 6.1.

5.1. Max-stable data

We first take the marginal parameters to be fixed and known and quantify the efficiency gains of $\hat{\theta}_{\text{Bayes}}$ compared to $\hat{\theta}_{\text{PL}}$. We simulate $N = 100$ samples $z^{(1)}, \dots, z^{(N)}$ from the logistic distribution for different dimensions $k \in \{6, 10, 50\}$ and different dependence parameters $\theta = 0.1 \times i$, $i = 1, \dots, 9$. For each combination of dimension k and parameter θ we then run a Markov chain with length 1500, where we discard the first 500 steps as the burn-in time. The empirical median of the remaining 1000 elements gives $\hat{\theta}_{\text{Bayes}}$. The chain is sufficiently long to reliably estimate the posterior median; see also the mixing properties in Section 6.1. The maximum pairwise likelihood estimator $\hat{\theta}_{\text{PL}}$ is obtained according to (25). The whole procedure is repeated 1500 times to compute the corresponding root mean squared errors and biases shown in Figure 1.

As expected, the use of full dependence information substantially decreases the root mean squared errors and thus increases the efficiency of the estimates. In extreme value statistics, where typically only small data sets are available, this allows to reduce uncertainty due to parameter estimation. The advantage of this additional information becomes stronger for both higher dimensions and weaker dependence, analogously to the observations in Huser et al. (2015). This behavior can to some extent be understood by the results in Shi (1995) on the Fisher information of the logistic distribution for different dimensions and dependence parameters. When $\theta \downarrow 0$, pairwise likelihood performs just as well as full likelihood, which is sensible since, up to a multiplicative constant, the pairwise likelihood equals the full likelihood when $\theta = 0$.

It is interesting to note that the estimates $\hat{\theta}_{\text{Bayes}}$ appear to be unbiased in

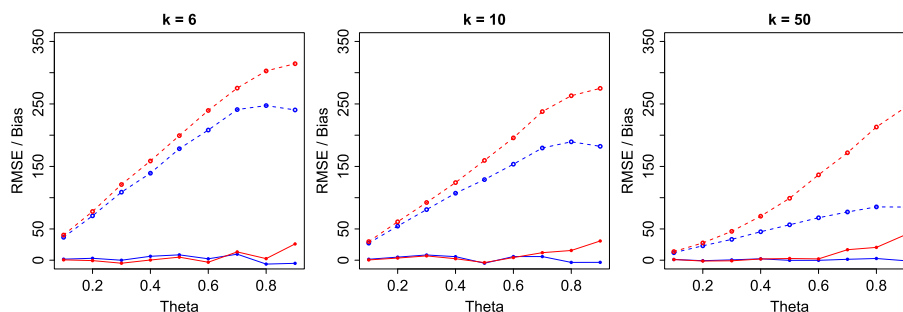


FIG 1. Root mean squared errors (dashed) and biases (solid) of $\hat{\theta}_{\text{Bayes}}$ (blue) and $\hat{\theta}_{\text{PL}}$ (red) for different dimensions k and different parameters θ . Values have been multiplied by 10000.

almost all cases, whereas the pairwise estimator has a finite sample bias for θ close to 1.

5.2. Data in the max-domain of attraction

In applications, the max-stable distribution Z might not be observed exactly but only as an approximation by componentwise block maxima of data vectors $X^{(1)}, \dots, X^{(b)}$ in its max-domain of attraction with standard Fréchet margins, where $b \in \mathbb{N}$ is the block size. Indeed, the random vector

$$\tilde{Z} = \frac{1}{b} \left(\max_{l=1, \dots, b} X_1^{(l)}, \dots, \max_{l=1, \dots, b} X_k^{(l)} \right),$$

approximates the distribution of Z , where the approximation improves for increasing b . In this situation we can associate to \tilde{Z} the partition of occurrence times of the maxima, say $\tilde{\tau}$. Stephenson and Tawn (2005b) proposed to use this information on the partition to simplify the likelihood of the max-stable distribution. For N observations $\tilde{z}^{(1)}, \dots, \tilde{z}^{(N)}$ of \tilde{Z} with partitions $\tilde{\tau}^{(1)}, \dots, \tilde{\tau}^{(N)}$ they defined the estimator

$$\hat{\theta}_{\text{ST}} = \arg \max_{\theta \in \Theta} \sum_{l=1}^N \log L(\tilde{z}^{(l)}, \tilde{\tau}^{(l)}; \theta).$$

This estimator suffers from two kinds of misspecification biases. Firstly, the $\tilde{z}^{(l)}$ are only approximately Z distributed and, secondly, the partitions $\tilde{\tau}^{(l)}$ are only finite sample approximations to the true distribution of the limit partition T . For the latter, Wadsworth (2015) proposed a bias reduction method for moderate dimensions and showed in a simulation study that it significantly decreases the bias of the Stephenson–Tawn estimator in the case where the $X^{(k)}$ and thus also \tilde{Z} follow exactly a max-stable logistic distribution. However, if the $X^{(k)}$ are samples from the outer power Clayton copula (cf. Hofert and Mächler, 2011)

k	$\theta_0 = 0.1$		$\theta_0 = 0.4$		$\theta_0 = 0.7$		$\theta_0 = 0.9$	
	6	10	6	10	6	10	6	10
RMSE(θ_{Bayes})	36	29	144	111	241	191	262	220
RMSE(θ_{PL})	40	32	159	127	279	235	311	286
RMSE(θ_{ST})	38	29	148	126	352	401	647	840
RMSE(θ_{W})	38	29	134	108	230	228	313	434
Bias(θ_{Bayes})	-9	-10	-44	-35	-57	-70	-96	-114
Bias(θ_{PL})	-10	-10	-47	-32	-44	-42	-46	-42
Bias(θ_{ST})	-12	-11	-90	-88	-315	-385	-634	-835
Bias(θ_{W})	-11	-11	-61	-58	-158	-194	-277	-422

TABLE 1

Root mean squared errors (top four rows) and biases (bottom four rows) of $\hat{\theta}_{\text{Bayes}}$, $\hat{\theta}_{\text{PL}}$, $\hat{\theta}_{\text{ST}}$ and $\hat{\theta}_{\text{W}}$, estimated from 1500 estimates; figures have been multiplied by 10000.

and thus only in the max-domain of attraction of the logistic distribution, then even the bias reduced estimator suffers from significant bias (cf., Wadsworth, 2015, Table 3).

We repeat the simulation study from Section 5.1 with the only difference that, instead of sampling from Z , we simulate $N = 100$ samples $\tilde{z}^{(1)}, \dots, \tilde{z}^{(N)}$ of \tilde{Z} , which is the rescaled maximum of $b = 50$ samples from the outer power Clayton copula for different parameters. Based on these data in the max-domain of attraction of the logistic distribution we estimate the dependence parameter θ using our Bayes estimator and compare it to the pairwise likelihood estimator. Both approaches ignore the additional information on the partitions $\tilde{\tau}^{(l)}$ that we have in this setup. On the other hand, we can also compute the Stephenson–Tawn estimator and its bias reduced version by Wadsworth (2015), which explicitly include the partition information.

Table 1 shows the root mean squared errors and biases of the four estimators. For all of them the bias plays a significant role for the overall estimation error and that is due to the model misspecification for only approximately max-stable data. This bias is however much stronger for $\hat{\theta}_{\text{ST}}$ and $\hat{\theta}_{\text{W}}$, which use the again misspecified partitions. In this case, the Bayes estimator that treats the partitions as unknown and samples from them automatically seems to be more robust and does not need a bias correction. At the same time it has a small variance and thus in many cases the smallest root mean squared error. Especially in higher dimensions (≥ 20) where the bias reduction of Wadsworth (2015) can no longer be used, the Bayes estimator still provides a robust and efficient method of inference. As one would expect, the pairwise likelihood estimator has the smallest bias since it is less sensitive to model misspecification, but it still a higher root mean squared error due to its higher variance.

5.3. Estimation of marginal extreme value parameters

In spatial settings, the marginal extreme value parameters are often estimated by using the independence likelihood (Chandler and Bate, 2007), where all locations

$\xi = -0.2$	$\theta_0 = 0.1$			$\theta_0 = 0.4$			$\theta_0 = 0.7$			$\theta_0 = 0.9$		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
Bayes full	105	75	22	97	58	21	71	37	20	51	28	19
Pairwise	106	73	31	98	57	28	72	39	24	52	30	21
Independence	111	75	67	101	58	52	73	39	35	52	30	24
$\xi = 0.4$	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
Bayes full	102	99	41	96	89	39	71	65	34	51	45	32
Pairwise	106	98	57	97	89	54	71	67	45	51	47	38
Independence	112	100	96	100	89	79	72	67	56	51	48	42
$\xi = 1$	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
Bayes full	109	155	85	100	144	76	77	111	59	53	75	46
Pairwise	106	146	94	96	135	90	74	108	72	52	75	53
Independence	110	146	127	98	135	107	74	109	81	52	76	57

TABLE 2

Root mean squared errors of (μ, σ, ξ) estimates with different values of ξ for the Bayesian approach, pairwise likelihoods and independence likelihoods, respectively, where θ is an unknown nuisance parameter; figures have been multiplied by 1000.

are assumed independent. This avoids to specify a dependence structure but can result in efficiency losses, even if only the marginal parameters are of interest.

We perform a simulation study to assess how using the full likelihoods in a Bayesian framework improves estimation of the marginal parameters. We fix the dimension $k = 10$ and set the marginal parameters to $\mu = 1$, $\sigma = 1$ and $\xi \in \{-0.2, 0.4, 1\}$, equal for all k margins. The dependence is logistic with unknown nuisance parameter $\theta_0 \in \{0.1, 0.4, 0.7, 0.9\}$.

Based on $N = 100$ independent samples from this model, we compare three different estimation procedures. The first one is our Bayesian approach using the full joint likelihood of the marginal parameters and the dependence parameter. We use a uniform prior for θ , and independent normal priors for μ , $\log \sigma$ and ξ with large standard deviations. For the univariate case, more sophisticated choices for the prior distributions are possible, including dependencies between the three extreme value parameters (e.g., Stephenson and Tawn, 2005a; Northrop and Attalides, 2016).

The second procedure is the maximum pairwise likelihood estimator that only uses bivariate dependence, and the third is the maximum independence likelihood estimator that completely ignores dependence between different components. Each simulation and estimation is repeated 1500 times.

Table 2 contains the root mean squared errors of the marginal parameters for the three approaches. Interestingly, for the location and scale parameter we see only little difference between the three methods, meaning that they can be efficiently estimated without taking into account dependencies. For the shape parameter, however, there are substantial improvements in the estimation error by including the unknown dependence structure in the model and estimating it simultaneously. Since estimation of the shape is both the most difficult and the most important of the three extreme value parameters, the Bayesian approach is promising also for marginal tail estimation. Finally, we observe that there is

already an efficiency gain for the shape parameter when only the pairwise dependence is considered, but it is even more remarkable in the Bayesian setting with full likelihoods. Table 2 also shows that these observations hold across different ranges for the shape parameter ξ . It should be noted that we considered the case of a well-specified model, where the class of dependence structures is known. An additional model uncertainty might render the independence likelihood more favorable.

6. Applications in a Bayesian framework

In the previous sections we discussed the efficiency gains of the Bayesian full likelihood approach in the frequentist framework of point estimates. The Markov chain from Section 2.2 however produces not only a point estimate but an estimate of the entire posterior distribution. For instance, this can directly be used to produce credible intervals for the parameter of interest. As a further application of our approach in the Bayesian framework, we will present Bayesian model comparison in this section.

6.1. The posterior distribution and credible intervals

As an illustration of the methodology we simulate a sample of $N = 15$ data $z^{(l)} = (z_1^{(l)}, \dots, z_k^{(l)})$, $l = 1, \dots, N$, from a k -dimensional max-stable vector Z whose distribution belongs to the parametric family of logistic distributions introduced in Subsection 4.1 with parameter space $\Theta = (0, 1)$. We run the Markov chain from Subsection 2.2. The left panel of Figure 2 shows the Markov chain for the parameter θ with simulated data from the logistic distribution in dimension $k = 10$ with $\theta_0 = 0.8$. The prior distribution is uniform, that is, $\pi_\theta = \text{Unif}(0, 1)$. The chain seems to have converged to its stationary distribution, namely the posterior distribution

$$L(\theta \mid \{z^{(l)}\}_{l=1}^N) \propto \pi_\theta(\theta) \prod_{l=1}^N L(z^{(l)}; \theta), \quad (26)$$

after a burn-in period of about 200 steps. The auto correlation of the Markov chain in Figure 3 suggests that there is serial dependence up to a lag of 30 steps. The parallel chain that updates the partitions is difficult to plot. The right panel of Figure 2 therefore shows in each step as a summary the mean number m of sets in the partitions $\tau^{(1)}, \dots, \tau^{(N)}$, that is, $m = 1/N \sum_{l=1}^N |\tau^{(l)}|$. For complete independence ($\theta_0 = 1$) we must have $m = k = 10$, whereas for complete dependence ($\theta_0 = 0$) we have $m = 1$.

The left panel of Figure 4 shows a histogram and an approximated smooth version of the posterior distribution, together with the uniform prior. In order to assess the impact of the prior distribution on the posterior, the two other panels contain the corresponding plots for the same data set but for different priors, namely the beta distributions $\pi_\theta = \text{Beta}(0.5, 0.5)$ (center) and $\pi_\theta = \text{Beta}(4, 4)$

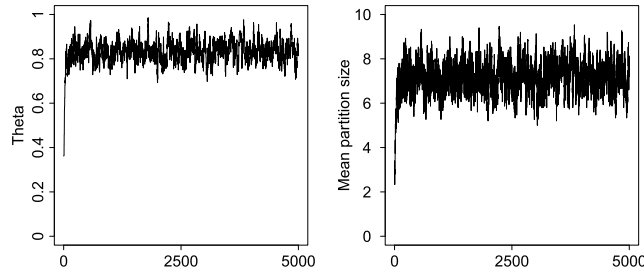


FIG 2. Markov chains for θ (left) and the mean partition size (right) with uniform prior.

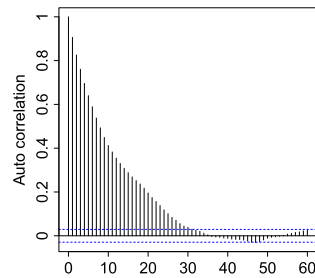


FIG 3. Auto correlation function of the Markov chain for the parameter θ .

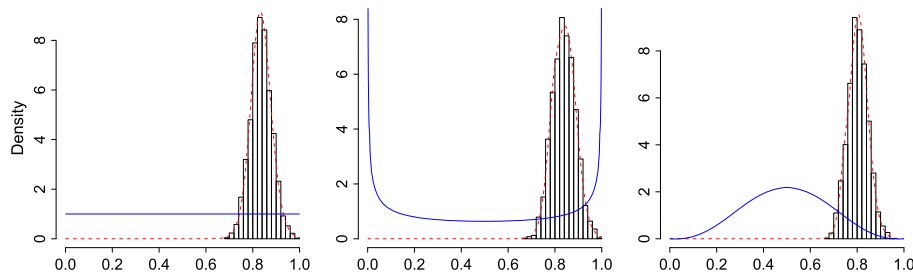


FIG 4. Histogram and smooth approximation of the posterior distribution (dotted red) for different priors (solid blue): $Unif(0,1)$ (left), $Beta(0.5,0.5)$ (center) and $Beta(4,4)$ (right).

(right). Even for a relatively small amount of $N = 15$ data points, the influence of the prior is not very strong.

The Bayesian setup provides us with a whole distribution for the parameter instead of a point estimate only. From this we can readily deduce credible intervals for the parameter θ . This is an advantage compared to frequentist composite likelihood methods since the Fisher information matrix has a “sandwich” form adjusting for the misspecified likelihood, and confidence intervals are thus not easily computed (Padoan et al., 2010). When using composite likelihoods in

k	$\theta_0 = 0.1$			$\theta_0 = 0.4$			$\theta_0 = 0.7$			$\theta_0 = 0.9$		
	6	10	50	6	10	50	6	10	50	6	10	50
Coverage (in %)	94	93	90	95	94	94	94	94	94	94	94	90

TABLE 3
Empirical coverage rates of 95% credible intervals obtained from the posterior distributions using full likelihood.

a Bayesian setup, the posterior distributions are much too concentrated and the empirical coverage rates are very small. Adjustments are necessary to obtain appropriate inference (Ribatet et al., 2012). Since our approach uses the full, correct likelihood, no adjustment is needed to obtain accurate empirical coverage rates. Indeed, in Table 3 we provide the coverage rates of the 95% credible intervals obtained in the simulation study in Section 5.1 for some values of θ_0 .

6.2. Bayesian model comparison

Starting from data z from a family of max-stable distributions $\{F_\theta, \theta \in \Theta\}$, we consider two sub-models $M_1 : \theta \in \Theta_1$ and $M_2 : \theta \in \Theta_2$ for disjoint sets $\Theta_1, \Theta_2 \subset \Theta$. In Bayesian statistics, comparison of such models is often based on the Bayes factor $B_{1,2}$, which translates the prior odds into the posterior odds (e.g., Kass and Raftery, 1995), that is,

$$\frac{\pi_{\text{posterior}}(\Theta_1)}{\pi_{\text{posterior}}(\Theta_2)} = B_{1,2} \times \frac{\pi_{\text{prior}}(\Theta_1)}{\pi_{\text{prior}}(\Theta_2)}. \quad (27)$$

The Bayes factor can also be written as $B_{1,2} = L(z | M_1)/L(z | M_2)$, where

$$L(z | M_i) = \int_{\Theta} L(z; \theta) \pi(\theta | M_i) d\theta, \quad i = 1, 2, \quad (28)$$

are the so-called marginal probabilities of the data and $\pi(\cdot | M_i)$ is the prior density of the parameter θ under the model M_i . Since the max-stable likelihood cannot be computed, the integral in (28) is computationally infeasible. However, we can use the estimation of the posterior probability (26) discussed in the previous subsection and estimate

$$B_{1,2} = \frac{\pi_{\theta}(\Theta_2)}{\pi_{\theta}(\Theta_1)} \times \frac{\int_{\Theta_1} L(\theta | \{z^{(l)}\}_{l=1}^N) d\theta}{\int_{\Theta_2} L(\theta | \{z^{(l)}\}_{l=1}^N) d\theta}. \quad (29)$$

As an example, we consider a simple regression model

$$\xi_i = \alpha + i\beta, \quad i = 1, \dots, k, \quad (30)$$

for the marginal shape parameters ξ_1, \dots, ξ_k of the k -dimensional max-stable distribution in dimension k . One might be interested in testing if there is a linear trend in the shape parameters, and, thus, in comparing the models $M_1 : \{\beta = 0\}$ and $M_2 : \{\beta \neq 0\}$. In order to compute the Bayes factor as the ratio of the

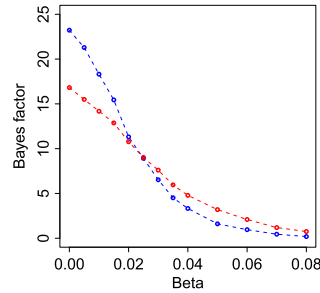


FIG 5. Bayes factors for different value of β for full likelihood (blue) and independence likelihood (red).

posterior probabilities of the two models according to (29), the prior distribution π_β of β must be a mixture $p_{0,\pi} \times \delta_{\{0\}} + (1 - p_{0,\pi}) \times \pi_\beta^c$ of a Dirac point mass $\delta_{\{0\}}$ on 0 and an appropriate continuous distribution π_β^c on \mathbb{R} with mixture weight $p_{0,\pi} \in (0, 1)$. This ensures that we have a positive posterior probability on both sets $\{\beta = 0\}$ and $\{\beta \neq 0\}$ and the Bayes factor is well-defined. Here, it is important to note that the choice of the mixture weight $p_{0,\pi}$ does not have an effect on the Bayes factor $B_{1,2}$ as Equations (27)–(29) show.

Similarly as in Section 5.3, we simulate $N = 15$ data from a max-stable logistic distribution with dimension $k = 10$ and dependence parameter $\theta_0 = 0.5$, with marginal parameters $\mu_i = 1$, $\sigma_i = 1$ and ξ_i as in (30) with $\alpha = 1$ and different values for β , $i = 1, \dots, k$. The prior distributions for the dependence, location and scale parameters are chosen as in Section 5.3. The prior for α is standard normal and for the prior for β is a mixture of $0.5 \times \delta_{\{0\}} + 0.5 \times \pi_\beta^c$ of a point mass and a centered normal with standard deviation 0.5 as the continuous component π_β^c .

A Markov chain whose stationary distribution is the posterior distribution of the parameters given the data can be constructed analogously to Section 2.2. However, given the current state β of the Markov chain, the proposal β^* is not drawn from a continuous distribution with density q , but from a mixture

$$p_0(\beta)\delta_{\{0\}}(\cdot) + (1 - p_0(\beta))q^c(\beta, \cdot)$$

of a Dirac point mass on $\{0\}$ and a continuous distribution with density $q^c(\beta, \cdot)$, with mixture weight $p_0(\beta) \in (0, 1)$. To ensure convergence of the Markov Chain, the densities $q^c(\beta, \cdot)$ should be chosen such that $q^c(\beta, \beta^*) > 0$ if and only if $q^c(\beta^*, \beta) > 0$.

Figure 5 shows the Bayes factors $B_{1,2}$ that compare the model without trend $M_1 : \{\beta = 0\}$ and the model with trend $M_2 : \{\beta \neq 0\}$ for the simulated data described above. The true trend varies from $\beta = 0$, in which case M_1 would be correct, over positive values up to $\beta = 0.08$ where M_2 is the correct model. As comparison, we implemented a Bayesian approach based on the independence likelihood (Chandler and Bate, 2007), which is the product of the marginal densities and ignores the dependence structure. The results show that using the full

likelihood that takes the dependence into account and treats it as a nuisance parameter significantly facilitates the distinction between the two different models. The Bayes factors for the full likelihood show stronger support for M_1 if $\beta = 0$, and decrease more rapidly to 0 if $\beta > 0$ than the Bayes factors for independence likelihood.

Finally, we note that a similar approach has been proposed in the univariate setting for estimation of the shape parameter in Stephenson and Tawn (2005a) in order to allow the Gumbel case $\xi = 0$ with positive probability.

7. Discussion

We present an approach that allows for inference of max-stable distributions based on full likelihoods by perceiving the underlying random partition of the data as latent variables in a Bayesian framework. The formulas for $\omega(\tau_j, z)$ provided in Section 4 allow in principle to perform Bayesian inference based on full likelihoods for many popular max-stable distributions in any dimension. However, computational challenges arise for both the extremal- t and the Brown–Resnick model in higher dimensions since the corresponding $\omega(\tau_j, z)$ require the evaluation of a multivariate Student and Gaussian distribution functions, respectively, which have to be approximated numerically; see also Thibaud et al. (2016). The recent work de Fondeville and Davison (2017) on efficient computation of Gaussian distribution functions allows for even higher dimensions.

Making use of the weights $\omega(\tau_j, z)$, the posterior distribution of the parameters becomes numerically available by samples based on Markov chain Monte Carlo techniques. As the results in Section 6.1 indicate, the posterior distribution does not show strong influence of the prior distribution even in case of a rather small amount of data; cf., Figure 4. In most of the examples presented here, the proposal distributions for the model parameters in the Metropolis–Hastings algorithms were chosen to be centered around the current state of the Markov chain with an appropriate standard deviation, resulting in chains with satisfactory convergence and mixing properties; cf., Figures 2 and 3, for instance. Further improvements of these properties might be possible, e.g., by implementing an adaptive design of the Markov chain Monte Carlo algorithms.

In the frequentist framework, we propose to use the posterior median as a point estimator for the model parameters. As the simulation studies in Section 5 show, the use of full likelihoods considerably improves the estimation errors compared to the commonly used composite likelihood method even in the case of a rather small sample size. This complements our theoretical results on the asymptotic efficiency of the posterior median. Besides the point estimator in the frequentist setting, we can also make use of the posterior distribution in a Bayesian framework. In Section 6, we discuss the use of credible intervals and Bayesian model comparison for max-stable distributions. Further applications such as Bayesian prediction are possible.

Appendix A: Proofs postponed from Section 4

A.1. Proofs from Subsection 4.1

Proof of Prop. 2. Taking partial derivative of the exponent function (15) we obtain

$$-\partial_{\tau_j} V_\theta(z) = \prod_{i=1}^{|\tau_j|-1} \left(\frac{i}{\theta} - 1 \right) \left(\sum_{i=1}^k z_i^{-1/\theta} \right)^{\theta-|\tau_j|} \prod_{i \in \tau_j} z_i^{-1/\theta-1}.$$

We note that

$$\frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1 - \theta)} = \prod_{i=1}^{|\tau_j|-1} (i - \theta).$$

Using this, Equation (5) becomes for the logistic model

$$L(\tau, z; \theta) = \exp\{-V(z)\} \prod_{j=1}^{\ell} \omega(\tau_j, z)$$

with

$$\omega(\tau_j, z) = \theta^{-1|\tau_j|+1} \frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1 - \theta)} \left(\sum_{i=1}^k z_i^{-1/\theta} \right)^{\theta-|\tau_j|} \prod_{i \in \tau_j} z_i^{-1-1/\theta}. \quad \square$$

Proof of Prop. 3. From Prop. 4.1 in Dombry et al. (2017a) it follows that the model is differentiable in quadratic mean. For any $1 \leq i_1 < i_2 \leq k$, the pairwise extremal coefficient of the logistic model with parameter $\theta \in (0, 1)$ is $\eta_{i_1, i_2}(\theta) = 2^\theta$, a strictly increasing function in θ . The assertion of the proposition follows by Remark 2. \square

A.2. Proofs from Subsection 4.2

Both the proof of Prop. 4 and Prop. 5 rely on the following lemma.

Lemma 1. *Let $Y(\alpha_1), \dots, Y(\alpha_k)$ be independent random variables such that $Y(\alpha)$ has a Gamma distribution with shape parameter $\alpha > 0$ and scale 1.*

- (i) *Let $U_1 > U_2 > \dots$ be the points of a Poisson point process on $(0, \infty)$ with intensity $u^{-2}du$ and $\tilde{Y}^{(1)}, \tilde{Y}^{(2)}, \dots$ independent copies of the random vector $\tilde{Y} = (Y(\alpha_i)/\alpha_i)_{1 \leq i \leq k}$. Then the simple max-stable random vector $Z = \bigvee_{i \geq 1} U_i \tilde{Y}^{(i)}$ has angular density (17).*
- (ii) *In the Dirichlet max-stable model (17), the pair extremal coefficient η_{i_1, i_2} , $1 \leq i_1 < i_2 \leq k$, is given by*

$$\eta_{i_1, i_2} = \eta(\alpha_{i_1}, \alpha_{i_2}) = \mathbb{E} \left[\frac{Y(\alpha_{i_1})}{\alpha_{i_1}} \vee \frac{Y(\alpha_{i_2})}{\alpha_{i_2}} \right].$$

Furthermore, $\eta : (0, \infty)^2 \rightarrow [1, 2]$ is continuously differentiable and strictly decreasing in both components.

Proof. For the proof of the first part, we note that the intensity of the spectral measure is given by

$$\lambda(z) = \int_0^\infty f_{\tilde{Y}}(z/u) u^{-k-2} du, \quad z \in (0, \infty)^k,$$

where

$$f_{\tilde{Y}}(\tilde{y}) = \prod_{i=1}^k \frac{\alpha_i^{\alpha_i}}{\Gamma(\alpha_i)} \tilde{y}_i^{\alpha_i-1} e^{-\alpha_i \tilde{y}_i}, \quad \tilde{y} \in (0, \infty)^k,$$

is the density of the random vector \tilde{Y} . A direct computation yields

$$\lambda(z) = \frac{\Gamma(1 + \sum_{i=1}^d \alpha_i)}{(\sum_{i=1}^d \alpha_i z_i)^{1 + \sum_{i=1}^d \alpha_i}} \prod_{i=1}^d \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i-1}}{\Gamma(\alpha_i)}.$$

We see that the restriction of λ to the simplex S^{k-1} is equal to h which proves the claim.

The first statement of the second part, is a direct consequence of the first part since

$$\eta(\alpha_1, \alpha_2) = -\log \mathbb{P}(Z_1 \leq 1, Z_2 \leq 1) = \mathbb{E} \left[\frac{Y(\alpha_1)}{\alpha_1} \vee \frac{Y(\alpha_2)}{\alpha_2} \right].$$

The proof of the strict monotonicity relies on the notion of convex order, see Chapter 3.4 in Denuit et al. (2005). For two real-valued random variables X_1, X_2 we say that X_1 is lower than X_2 in convex order if $\mathbb{E}[\varphi(X_1)] \leq \mathbb{E}[\varphi(X_2)]$ for all convex functions $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ such that the expectations exist. It is known that the family of random variables $(Y(\alpha)/\alpha)_{\alpha>0}$ is non-increasing in convex order (Ramos et al., 2000, Section 4.3), and, in this case, the Lorenz order is equivalent to the convex order (Denuit et al., 2005, Property 3.4.41). We show below that this implies that $\eta(\alpha_1, \alpha_2)$ is strictly decreasing in its arguments.

Let $\alpha'_1 > \alpha_1 > 0$ and $\alpha_2 > 0$ and let us prove that $\eta(\alpha'_1, \alpha_2) < \eta(\alpha_1, \alpha_2)$. For independent random variables $Y(\alpha_1)$, $Y(\alpha'_1)$ and $Y(\alpha_2)$, we have

$$\eta(\alpha'_1, \alpha_2) = \mathbb{E} \left[\frac{Y(\alpha'_1)}{\alpha'_1} \vee \frac{Y(\alpha_2)}{\alpha_2} \right] \quad \text{and} \quad \eta(\alpha_1, \alpha_2) = \mathbb{E} \left[\frac{Y(\alpha_1)}{\alpha_1} \vee \frac{Y(\alpha_2)}{\alpha_2} \right].$$

Using that $Y(\alpha'_1)/\alpha'_1$ is lower than $Y(\alpha_1)/\alpha_1$ in convex order, we obtain

$$\mathbb{E} \left[\frac{Y(\alpha'_1)}{\alpha'_1} \vee \frac{y_2}{\alpha_2} \right] \leq \mathbb{E} \left[\frac{Y(\alpha_1)}{\alpha_1} \vee \frac{y_2}{\alpha_2} \right] \quad \text{for all } y_2 > 0, \quad (31)$$

because the map $u \mapsto u \vee (y_2/\alpha_2)$ is convex. Replacing y_2 by $Y(\alpha_2)$ and integrating, we get $\eta(\alpha'_1, \alpha_2) \leq \eta(\alpha_1, \alpha_2)$. The equality $\eta(\alpha'_1, \alpha_2) = \eta(\alpha_1, \alpha_2)$ would imply that for almost every $y_2 > 0$ the equality holds in (31) which is true if and only if $Y(\alpha_1)/\alpha_1$ and $Y(\alpha'_1)/\alpha'_1$ have the same distribution. Since this is not the case, $\eta(\alpha'_1, \alpha_2) < \eta(\alpha_1, \alpha_2)$ and η is strictly decreasing in α_1 . By symmetry, η is also strictly decreasing in α_2 .

Finally, the fact that $(\alpha_1, \alpha_2) \mapsto \eta(\alpha_1, \alpha_2)$ is continuously differentiable follows from the integral representation

$$\eta(\alpha_1, \alpha_2) = \int_0^\infty \int_0^\infty \frac{y_1}{\alpha_1} \vee \frac{y_2}{\alpha_2} \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1} y_2^{\alpha_2-1} e^{-y_1} e^{-y_2} dy_1 dy_2, \quad (32)$$

for $\alpha_1, \alpha_2 > 0$, and standard theorems for integrals depending on a parameter. \square

Proof of Prop. 4. From the construction given in the first part of Lemma 1, we obtain

$$\begin{aligned} \lambda(z) &= \int_0^\infty \left(\prod_{i=1}^k \frac{\alpha_i^{\alpha_i}}{\Gamma(\alpha_i)} \left(\frac{z_i}{r} \right)^{\alpha_i-1} e^{-(\alpha_i z_i/r)} \right) r^{-2-k} dr \\ &= \prod_{i=1}^k \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i-1}}{\Gamma(\alpha_i)} \int_0^\infty e^{-\frac{1}{r} \sum_{i=1}^k \alpha_i z_i} r^{-2-\sum_{i=1}^k \alpha_i} dr \end{aligned}$$

and, consequently,

$$\begin{aligned} &\int_{u_j < z_{\tau_j^c}} \lambda(z_{\tau_j}, u_j) du_j \\ &= \prod_{i \in \tau_j} \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i-1}}{\Gamma(\alpha_i)} \int_0^\infty \int_{u_j < z_{\tau_j^c}} \prod_{i \in \tau_j^c} \left(\frac{\alpha_i^{\alpha_i} z_i^{\alpha_i-1}}{\Gamma(\alpha_i)} e^{-(\alpha_i z_i/r)} \right) du_j \\ &\quad \cdot e^{-\frac{1}{r} \sum_{i \in \tau_j} \alpha_i z_i} r^{-2-\sum_{i=1}^k \alpha_i} dr \\ &= \prod_{i \in \tau_j} \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i-1}}{\Gamma(\alpha_i)} \int_0^\infty e^{-\frac{1}{r} \sum_{i \in \tau_j} \alpha_i z_i} \left(\prod_{i \in \tau_j^c} F_{\alpha_i}(\alpha_i z_i/r) \right) r^{-2-\sum_{i \in \tau_j} \alpha_i} dr, \end{aligned}$$

where

$$F_\alpha(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt,$$

is the distribution function of a Gamma variable with shape $\alpha > 0$. \square

Proof of Prop. 5. Prop. 4.2 in Dombry et al. (2017a) implies that the model is differentiable in quadratic mean. In order to verify Eq. (13) for the Dirichlet model, we consider the mapping

$$\Psi : (0, \infty)^k \rightarrow [1, 2]^k, \quad \theta = (\alpha_1, \dots, \alpha_k) \mapsto (\eta_{1,2}, \eta_{2,3}, \eta_{1,3}, \eta_{1,4}, \dots, \eta_{1,k}).$$

We first show that Ψ is injective. To this end, let $\theta^{(1)} \neq \theta^{(2)} \in \Theta$ where $\psi_i = (\alpha_1^{(i)}, \dots, \alpha_k^{(i)})$, $i = 1, 2$. We distinguish between two cases. First, we assume that $\theta^{(1)}$ and $\theta^{(2)}$ share at least one common component. Then, there is a pair $(i, j) \in \{(1, 2), (2, 3), (1, 3), (1, 4), \dots, (1, k)\}$ such that $(\alpha_i^{(1)}, \alpha_j^{(1)})$ and $(\alpha_i^{(2)}, \alpha_j^{(2)})$ differ in exactly one component. As $\eta_{i,j} = \eta(\alpha_i, \alpha_j)$ is strictly decreasing both in

α_i and α_j , by Lemma 1, we have that $\eta(\alpha_i^{(1)}, \alpha_j^{(1)}) \neq \eta(\alpha_i^{(2)}, \alpha_j^{(2)})$. Secondly, we consider the case that $\theta^{(1)}$ and $\theta^{(2)}$ do not share any common component. Then, there is a pair $(i, j) \in \{(1, 2), (2, 3), (1, 3)\}$ such that both components $(\alpha_i^{(1)} - \alpha_i^{(2)}, \alpha_j^{(1)} - \alpha_j^{(2)})$ have the same sign and, again, by the strict monotonicity of $\eta(\alpha_i, \alpha_j)$, it follows that $\eta(\alpha_i^{(1)}, \alpha_j^{(1)}) \neq \eta(\alpha_i^{(2)}, \alpha_j^{(2)})$. Hence, in both cases, $\Psi(\theta^{(1)}) \neq \Psi(\theta^{(2)})$, that is, Ψ is injective and there exists a unique inverse function $\Psi^{-1} : \Psi((0, \infty)^k) \rightarrow (0, \infty)^k$.

Consider the set

$$\Theta' = \{(\alpha_1, \dots, \alpha_k) \in (0, \infty)^k : \partial_{\alpha_i} \eta(\alpha_i, \alpha_j) < 0 \\ \text{and } \partial_{\alpha_j} \eta(\alpha_i, \alpha_j) < 0 \text{ for all } 1 \leq i < j \leq k\}.$$

Note that since η is continuously differentiable and strictly decreasing in its argument, $\Theta \setminus \Theta'$ has Lebesgue measure 0. For all $\theta_0 = (\alpha_1, \dots, \alpha_k) \in \Theta'$, the Jacobian $D\Psi$ satisfies

$$\det\{D\Psi(\theta_0)\} = \{\partial_{\alpha_1} \eta(\alpha_1, \alpha_2) \cdot \partial_{\alpha_2} \eta(\alpha_2, \alpha_3) \cdot \partial_{\alpha_3} \eta(\alpha_1, \alpha_3) \\ + \partial_{\alpha_2} \eta(\alpha_1, \alpha_2) \cdot \partial_{\alpha_3} \eta(\alpha_2, \alpha_3) \cdot \partial_{\alpha_1} \eta(\alpha_1, \alpha_3)\} \cdot \prod_{j=4}^k \partial_{\alpha_j} \eta(\alpha_1, \alpha_j) \\ \neq 0.$$

The inverse function theorem then implies that Ψ^{-1} is continuously differentiable at $\Psi(\theta_0)$, that is, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\|\Psi(\theta_0) - \Psi(\theta)\|_\infty < \delta$ implies $\|\theta_0 - \theta\|_\infty < \varepsilon$. In particular, we obtain

$$\inf_{\|\theta_0 - \theta\|_\infty > \varepsilon} \|\Psi(\theta_0) - \Psi(\theta)\|_\infty \geq \delta,$$

that is, Eq. (13), and the asymptotic normality and efficiency of the posterior median for $\theta_0 \in \Theta'$ follow from Prop. 1. Finally, we note that each extremal coefficient η is continuously differentiable and strictly decreasing with respect to both components by Lemma 1. Thus, $\partial_{\alpha_1} \eta(\alpha_1, \alpha_2) < 0$ and $\partial_{\alpha_2} \eta(\alpha_1, \alpha_2) < 0$ for almost every $\theta \in \Theta$. \square

A.3. Proofs from Subsection 4.3

Proof of Prop. 7. By Prop. 4.3 in Dombry et al. (2017a), the model is differentiable in quadratic mean (even if $\nu > 0$ is not fixed). For any $1 \leq i_1 < i_2 \leq k$, and fixed $\nu > 0$, the pairwise extremal coefficient of the extremal- t model with parameter matrix $\Sigma = \{\rho_{ij}\}_{1 \leq i, j \leq k}$ is

$$\eta_{i_1, i_2}(\Sigma) = 2T_{\nu+1} \left(\sqrt{(\nu+1) \frac{1 - \rho_{i_1 i_2}}{1 + \rho_{i_1 i_2}}} \right), \quad (33)$$

where $T_{\nu+1}$ denotes the distribution function of a t -distribution with $\nu + 1$ degrees of freedom. Therefore, $\eta_{i_1, i_2}(\Sigma)$ as a function of $\rho_{i_1 i_2} \in [-1, 1]$ is strictly decreasing and the claim follows by Remark 2 together with Prop. 1. \square

Proof of Cor. 1. Analogously to the proof of Cor. 4.4 in Dombry et al. (2017a), it can be shown that the model is differentiable in quadratic mean. Suppose that $\|t_1 - t_2\|_2 \neq \|t_2 - t_3\|_2$ and observe that the mapping $\Psi : \Theta \rightarrow \Psi(\Theta)$, $\theta = (s, \alpha) \mapsto \{\rho_{ij}\}_{1 \leq i, j \leq k} = \{\exp(-\|t_i - t_j\|_2^\alpha / s)\}_{1 \leq i, j \leq k}$ is continuously differentiable. Since

$$\alpha = \frac{\log\{\log \rho_{12}\} - \log\{\log \rho_{23}\}}{\log \|t_1 - t_2\|_2 - \log \|t_2 - t_3\|_2}, \quad s = -\frac{\|t_1 - t_2\|_2^\alpha}{\log \rho_{12}},$$

the same holds true for the inverse mapping Ψ^{-1} .

Further, from the continuity of Ψ^{-1} at $\Psi(\theta_0)$ at any $\theta_0 \in \Theta$, we obtain that for every $\varepsilon > 0$ there is some $\delta > 0$ such that for all $\Psi(\theta) \in \Psi(\Theta)$ with $\|\Psi(\theta_0) - \Psi(\theta)\|_\infty < \delta$ we have $\|\theta_0 - \theta\|_\infty < \varepsilon$. Consequently,

$$\inf_{\|\theta_0 - \theta\|_\infty > \varepsilon} \|\Psi(\theta_0) - \Psi(\theta)\|_\infty \geq \delta.$$

From the proof of Prop. 7 and with the notation in (33), we obtain that $\|\Psi(\theta_0) - \Psi(\theta)\|_\infty \geq \delta$ implies $\max_{1 \leq i_1 < i_2 \leq k} |\eta_{i_1, i_2}\{\Psi(\theta_0)\} - \eta_{i_1, i_2}\{\Psi(\theta)\}| > \delta'$ for some $\delta' > 0$, that is, Equation (13) holds. The assertion follows then from Prop. 1. \square

A.4. Proofs from Subsection 4.4

Proof of Prop. 9. From Prop. 4.5 in Dombry et al. (2017a), it follows that the model is differentiable in quadratic mean. For any $1 \leq i_1 < i_2 \leq k$, the pairwise extremal coefficient of the Hüsler–Reiss model with parameter matrix $\Lambda = \{\lambda_{i,j}^2\}_{1 \leq i, j \leq k}$ is

$$\eta_{i_1, i_2}(\Lambda) = 2\Phi_1 \left\{ \sqrt{\lambda_{i_1, i_2}^2} \right\},$$

which is a strictly increasing function in $\lambda_{i_1, i_2}^2 > 0$, and the claim follows by Remark 2 together with Prop. 1. \square

Proof of Cor. 2. Analogously to Cor. 4.6 in Dombry et al. (2017a), the model can be shown to be differentiable in quadratic mean. Suppose that $\|t_1 - t_2\|_2 \neq \|t_2 - t_3\|_2$. As the mapping $\Psi : \Theta \rightarrow \Psi(\Theta)$, $\theta = (\lambda, \alpha) \mapsto \{\lambda_{ij}^2\}_{1 \leq i, j \leq k} = \left\{ \frac{\|t_i - t_j\|_2^\alpha}{4s} \right\}_{1 \leq i, j \leq k}$ is continuously differentiable and

$$\alpha = \frac{\log \gamma_{12} - \log \gamma_{23}}{\log \|t_1 - t_2\|_2 - \log \|t_2 - t_3\|_2}, \quad s = \frac{\|t_i - t_j\|_2^\alpha}{4\lambda_{ij}^2},$$

the inverse mapping Ψ^{-1} is continuously differentiable, as well. The same arguments as in the proof of Cor. 1 together with the proof of Prop. 9 yield the assertion. \square

Acknowledgments

We thank the editorial team and two referees for helpful comments. Financial support by the Swiss National Science Foundation (S. Engelke) and by the Bourgogne–Franche–Comté region (C. Dombry, grant OPE-2017-0068) is gratefully acknowledged.

References

- P. Asadi, A. Davison, and S. Engelke. Extremes on river networks. *Ann. Appl. Stat.*, 9:2023–2050, 2015. [MR3456363](#)
- C. Berg, J. P. R. Christensen, and P. Ressel. *Harmonic Analysis on Semigroups*. Springer-Verlag, New York, 1984. [MR0747302](#)
- M.-O. Boldi and A. C. Davison. A mixture model for multivariate extremes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69:217–229, 2007. [MR2325273](#)
- B. M. Brown and S. I. Resnick. Extreme values of independent stochastic processes. *J. Appl. Probab.*, 14:732–739, 1977. [MR0517438](#)
- T. A. Buishand, L. de Haan, and C. Zhou. On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.*, 2:624–642, 2008. [MR2524349](#)
- S. Castruccio, R. Huser, and M. G. Genton. High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Statist.*, 25:1212–1229, 2016. [MR3572037](#)
- R. E. Chandler and S. Bate. Inference for clustered data using the independence loglikelihood. *Biometrika*, 94:167–183, 2007. [MR2367830](#)
- S. G. Coles and J. A. Tawn. Modelling extreme multivariate events. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 53:377–392, 1991. [MR1108334](#)
- A. C. Davison and M. M. Gholamrezaee. Geostatistics of extremes. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 468:581–608, 2012. [MR2874052](#)
- R. de Fondeville and A. Davison. High-dimensional peaks-over-threshold inference. Available from <https://arxiv.org/abs/1605.08558>, 2017.
- M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley, 2005.
- C. Dombry and F. Éyi-Minko. Regular conditional distributions of continuous max-infinitely divisible random fields. *Electron. J. Probab.*, 18:1–21, 2013. [MR3024101](#)
- C. Dombry, F. Éyi-Minko, and M. Ribatet. Conditional simulation of max-stable processes. *Biometrika*, 100:111–124, 2013. [MR3034327](#)
- C. Dombry, S. Engelke, and M. Oesting. Exact simulation of max-stable processes. *Biometrika*, 103:303–317, 2016. [MR3509888](#)
- C. Dombry, S. Engelke, and M. Oesting. Asymptotic properties of the maximum likelihood estimator for multivariate extreme value distributions. Available from <https://arxiv.org/abs/1612.05178v2>, 2017a.
- C. Dombry, M. Genton, R. Huser, and M. Ribatet. Full likelihood inference for max-stable data. Available from <https://arxiv.org/abs/1703.08665>, 2017b.
- J. H. J. Einmahl, A. Kiriliouk, A. Krajina, and J. Segers. An M-estimator of spatial tail dependence. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 78:275–298, 2016. [MR3453656](#)
- S. Engelke and J. Ivanovs. Robust bounds in multivariate extremes. *The Annals of Applied Probability*, 2017. To appear.
- S. Engelke, A. Malinowski, M. Oesting, and M. Schlather. Statistical inference for max-stable processes by conditioning on extreme events. *Adv. Appl. Probab.*, 46:478–495, 2014. [MR3215542](#)

- S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 77:239–265, 2015. [MR3299407](#)
- M. G. Genton, Y. Ma, and H. Sang. On the likelihood function of gaussian max-stable processes. *Biometrika*, 98:481–488, 2011. [MR2806443](#)
- M. Hofert and M. Mächler. Nested Archimedean copulas meet R: The nacopula package. *J. Stat. Softw.*, 39:1–20, 2011.
- R. Huser and A. Davison. Composite likelihood estimation for the brown-resnick process. *Biometrika*, 100:511–518, 2013. [MR3068451](#)
- R. Huser, A. C. Davison, and M. G. Genton. Likelihood estimators for multivariate extremes. *Extremes*, 19:79–103, 2015. [MR3454032](#)
- J. Hüsler and R.-D. Reiss. Maxima of normal random vectors: between independence and complete dependence. *Statist. Probab. Lett.*, 7:283–286, 1989. [MR0980699](#)
- Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, 37:2042–2065, 2009. [MR2561440](#)
- R. E. Kass and A. E. Raftery. Bayes factors. *J. Am. Stat. Assoc.*, 90:773–795, 1995. [MR3363402](#)
- A. K. Nikoloulopoulos, H. Joe, and H. Li. Extreme value properties of multivariate t copulas. *Extremes*, 12:129–148, 2009. [MR2515644](#)
- P. J. Northrop and N. Attalides. Posterior propriety in Bayesian extreme value analyses using reference priors. *Statist. Sinica*, 26:721–743, 2016. [MR3497768](#)
- T. Opitz. Extremal t processes: Elliptical domain of attraction and a spectral representation. *J. Multivariate Anal.*, 122:409–413, 2013. [MR3189331](#)
- S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood-based inference for max-stable processes. *J. Am. Stat. Assoc.*, 105:263–277, 2010. [MR2757202](#)
- H. M. Ramos, J. Ollero, and S. M. A. A sufficient condition for generalized lorenz order. *Journal of Economic Theory*, 90:286–292, 2000. [MR1742135](#)
- M. Ribatet, D. Cooley, and A. C. Davison. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22:813–845, 2012. [MR2954363](#)
- H. Rootzén and N. Tajvidi. Multivariate generalized Pareto distributions. *Bernoulli*, 12:917–930, 2006. [MR2265668](#)
- H. Rootzén, J. Segers, and J. L. Wadsworth. Multivariate peaks over thresholds models. *Extremes*, 2017.
- M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5:33–44, 2002. [MR1947786](#)
- D. Shi. Fisher information for a multivariate extreme value distribution. *Biometrika*, 82:644–649, 1995. [MR1366288](#)
- A. Stephenson and J. Tawn. Bayesian inference for extremes: Accounting for the three extremal types. *Extremes*, 7:291–307, 2005a. [MR2212389](#)
- A. Stephenson and J. A. Tawn. Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92:213–227, 2005b. [MR2158621](#)
- E. Thibaud and T. Opitz. Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102:855–870, 2015. [MR3431558](#)

- E. Thibaud, J. Aalto, D. S. Cooley, A. C. Davison, and J. Heikkinen. Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *Ann. Appl. Stat.*, 10:2303–2324, 2016. [MR3592058](#)
- A. W. van der Vaart. *Asymptotic statistics*. Cambridge University Press, Cambridge, 1998. [MR1652247](#)
- J. L. Wadsworth. On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika*, 102:705–711, 2015. [MR3394285](#)
- J. L. Wadsworth and J. A. Tawn. Efficient inference for spatial extreme value processes associated to log-gaussian random functions. *Biometrika*, 101:1–15, 2014. [MR3180654](#)