



Semiparametric bivariate modelling with flexible extremal dependence

Manuele Leonelli¹ · Dani Gamerman²

Received: 8 May 2018 / Accepted: 10 March 2019 / Published online: 28 May 2019
© The Author(s) 2019

Abstract

Inference over multivariate tails often requires a number of assumptions which may affect the assessment of the extreme dependence structure. Models are usually constructed in such a way that extreme components can either be asymptotically dependent or be independent of each other. Recently, there has been an increasing interest on modelling multivariate extremes more flexibly, by allowing models to bridge both asymptotic dependence regimes. Here we propose a novel semiparametric approach which allows for a variety of dependence patterns, be them extremal or not, by using in a model-based fashion the full dataset. We build on previous work for inference on marginal exceedances over a high, unknown threshold, by combining it with flexible, semiparametric copula specifications to investigate extreme dependence, thus separately modelling marginals and dependence structure. Because of the generality of our approach, bivariate problems are investigated here due to computational challenges, but multivariate extensions are readily available. Empirical results suggest that our approach can provide sound uncertainty statements about the possibility of asymptotic independence, and we propose a criterion to quantify the presence of either extreme regime which performs well in our applications when compared to others available. Estimation of functions of interest for extremes is performed via MCMC algorithms. Attention is also devoted to the prediction of new extreme observations. Our approach is evaluated through simulations, applied to real data and assessed against competing approaches. Evidence demonstrates that the bulk of the data do not bias and improve the inferential process for extremal dependence in our applications.

Keywords Asymptotic dependence · Copulae · GPD distribution · High quantiles · Prediction · Threshold estimation

1 Introduction

Precise knowledge of the tail behaviour of a distribution as well as predicting capabilities about the occurrence of extremes are fundamental in many applications, as for instance environmental sciences and finance. Evidence points to an increasing trend of such extreme events in environmental applications with associated economic and insurance losses growing dramatically (Salvatori et al.

2007). In most cases, the analysis of such extreme events is inherently multivariate. Interest is then on the concomitant observation of extremes on a number of variables. For instance, the effects on the human respiratory system are particularly dramatic after exposure to high concentrations of both ozone O_3 and nitrogen dioxide NO_2 .

Since standard statistical methods do not guarantee precise extrapolation towards the tail of the distribution, a variety of methods tailored to inference about tails have been introduced under the general name of *extreme-value theory*. Whilst univariate models can be faithfully applied in most applications, since their underlying assumptions are flexible enough to be met in practice, the application of multivariate methods often requires a number of assumptions which may impact the inferential process. First, a large number of models assume that variables are asymptotically dependent, meaning that there exist a dependence between the marginal extreme events (e.g. Boldi and Davison 2007; Sabourin and Naveau 2014). Such models thus exclude the possibility that,

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11222-019-09878-w>) contains supplementary material, which is available to authorized users.

✉ Manuele Leonelli
manuele.leonelli@glasgow.ac.uk

¹ School of Mathematics and Statistics, University of Glasgow, Glasgow, UK

² Instituto de Matematica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil

although dependence between random variables exists at finite levels, their extremes will be independent in the limit, a condition usually referred to as *asymptotic independence*. Second, the application of such methods requires the arbitrary selection of data points considered “extreme”, usually selected as those that exceed a fixed threshold. However, this choice can greatly affect the inferential process (Scarrott and MacDonald 2012; Wan and Davis 2019). Finally, data is usually transformed via the empirical cumulative function so that each marginal has the same distribution. However, this transformation is known to bias inference on extremes (Einmahl and Segers 2009).

A new sound, flexible approach is proposed here to model multivariate extremes that requires little, if no, assumptions and that has the capability of bridging the two possible cases of asymptotic dependence and independence (there has been an increasing interest in the development of models with this property, e.g. Huser et al. 2017; Huser and Wadsworth 2019; Wadsworth and Tawn 2012; Wadsworth et al. 2017). Our approach formally uses in a model-based fashion the full dataset, thus not requiring the arbitrary selection of an extreme region of points (similarly to Aulbach et al. 2012; Vrac et al. 2007). Our applications show that the bulk of the data does not bias the inferential ascertainment of the asymptotic dependence structure. The approach proposed here splits modelling into two tasks: modelling marginals using some recently developed methodology for univariate extremes justified by the asymptotic theory for tails, and modelling dependence using a flexible semiparametric copula structure which does not require any assumption about the tail dependence decay. Although the use of copula functions with specific marginals to build multivariate models is not new, we show below that our semiparametric specification has critical implications for inference on extremes.

Inference is carried out within the *Bayesian* paradigm using the *MCMC* machinery (Gelman and Lopes 2006), enabling us to straightforwardly deliver a wide variety of estimates and predictions of quantities of interest, e.g. high quantiles. The flexibility of our approach makes it computationally intensive but easily handled. Here we thus restrict our attention to bivariate problems (for some recent references on bivariate extremes see e.g. Camilo and de Carvalho 2017; Engelke et al. 2018; Guillothe et al. 2011; Wadsworth et al. 2017). However, multivariate extensions of the approach are readily available and discussed in Sect. 8.

The code for the implementation of our approach is written in OX (Doornik 1996)¹. Multi-purpose software for the application of the methods is currently being developed.

The paper is structured as follows. Section 2 reviews univariate models for exceedances and extreme-value mixture

models. Section 3 deals with multivariate extremes and measures of extreme dependence. Having described in Sect. 2 models for marginals, we introduce copulae and mixtures of these in Sect. 4. This section further provides an overview of the inferential results for mixtures we use. Our approach and inferential routines are described in Sect. 5. Section 6 presents a simulation study to both investigate their performance and address the issue of model choice. In Sect. 7 our methodology is applied to two real-world applications: river flows in Puerto Rico and NO₂/O₃ concentrations in the city of Leeds. We conclude with a discussion.

2 Modelling of univariate extremes

2.1 Peaks over threshold approach

A common approach to model extremes, often referred to as *peaks over threshold* (POT), studies the exceedances over a threshold. A key result to apply this methodology is due to Pickands (1975) which states that, under general regularity conditions, the only possible non-degenerate limiting distribution of properly rescaled exceedances over a threshold u is the *generalized Pareto distribution* (GPD). Its cumulative distribution function (cdf) P is defined as

$$P(x|\xi, \sigma, u) = \begin{cases} 1 - (1 + \xi \frac{x-u}{\sigma})^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp(-\frac{x-u}{\sigma}), & \text{if } \xi = 0, \end{cases}$$

for $u, \xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$, where the support is $x \geq u$ if $\xi \geq 0$ and $u \leq x \leq u - \sigma/\xi$ if $\xi < 0$. Therefore, the GPD is bounded if $\xi < 0$ and unbounded from above if $\xi \geq 0$. The application of this result in practice entails first the selection of a threshold u beyond which the GPD approximation appears to be tenable and then the fit of a GPD over data points that exceed the threshold.

The POT approach has two serious drawbacks. First, only a small subset of the data points, those beyond the threshold, are formally retained in a model-based approach during the inferential process. Thus, parameter estimates may not be reliable when the number of data points is small. Second, the choice of the threshold over which to fit a GPD is arbitrary. Although tools to guide this choice exist (e.g. Davison and Smith 1990), inference can greatly vary for different thresholds (Einmahl et al. 2009; Scarrott and MacDonald 2012).

2.2 The MGPD model

A variety of models called *extreme-value mixture models* (Scarrott and MacDonald 2012) have been recently defined to formally take into account the full dataset and not require a fixed threshold. These combine a flexible model

¹ The code implementing the models for the applications of Sect. 7 is freely available at https://github.com/manueleleonelli/Biv_ext_exc.

for the bulk of the data points, those below the threshold, a formally justifiable model for the tail and uncertainty measures for the threshold. A building block of our approach is the extreme-value mixture model of Nascimento et al. (2012), which we henceforth refer to as mixture of gamma Pareto distribution (MGPD). This consists of a finite mixture of gamma distributions for the bulk coupled with a GPD for the tail. The parametrization of the gamma suggested in Wiper et al. (2001) in terms of shape, η , and mean, μ , parameters is used. Its density, g , is $g(x|\mu, \eta) = \Gamma(\eta)^{-1} (\eta/\mu)^\eta x^{\eta-1} \exp(-\eta x/\mu)$, and its cdf is denoted by G .

A finite mixture of these distributions is defined next. For $n \in \mathbb{N}$, let $[n] = \{1, \dots, n\}$. The density h and the cdf H of a finite mixture of n gammas are formally defined as

$$\begin{aligned}
 h(x|\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w}) &= \sum_{i \in [n]} w_i g(x|\mu_i, \eta_i), \\
 H(x|\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w}) &= \sum_{i \in [n]} w_i G(x|\mu_i, \eta_i),
 \end{aligned}
 \tag{1}$$

where $\boldsymbol{\mu} = (\mu_i)_{i \in [n]}$, $\boldsymbol{\eta} = (\eta_i)_{i \in [n]}$, $\boldsymbol{w} = (w_i)_{i \in [n]}$ and \boldsymbol{w} is such that $w_i \geq 0$ and $\sum_{i \in [n]} w_i = 1$.

The density f of an MGPD then consists of a mixture of gamma densities h for the bulk and a GPD density p for the right tail. Formally,

$$f(x|\Theta) = \begin{cases} h(x|\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w}), & \text{if } x \leq u, \\ (1 - H(u|\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w})) p(x|\xi, \sigma, u), & \text{if } x > u, \end{cases}$$

where $\Theta = \{\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w}, \xi, \sigma, u\}$. An example of an MGPD density fitting simulated data is presented in Fig. 1, where it is clearly discernible that the bulk of the distribution consists of a mixture of 2 gammas, whilst beyond the threshold the density has GPD decay.

The cdf of an MGPD F is similarly defined in a piecewise fashion. Whilst below the threshold u , this is the cdf of the mixture of gammas H , over the threshold, i.e. for $x > u$, it can be written as $F(x|\Theta) = H(u|\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w}) + (1 - H(u|\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w})) P(x|\xi, \sigma, u)$.

A great advantage of the MGPD model is that high quantiles beyond the threshold, i.e. q values such that $\mathbb{P}(X >$

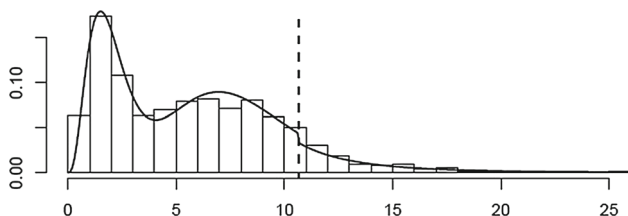


Fig. 1 Example of a MGPD density fit consisting of a mixture of 2 gammas for the bulk: solid line—MGPD density; dashed line—threshold

$q|\Theta) = 1 - p$ for $p > F(u|\Theta)$, have a closed-form expression. Specifically, this is a function q of both the probability p and the parameter Θ defined as

$$q(p|\Theta) = u + \frac{\sigma}{\xi} \left(\left(1 - \frac{p - H(u|\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w})}{1 - H(u|\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{w})} \right)^{-\xi} - 1 \right).$$

Nascimento et al. (2012) showed that the MGPD can outperform standard POT methods in situations where determination of the threshold is difficult. So nothing is lost using this approach instead of considering only the extreme points as in the standard POT approach. The number of mixture components for the bulk can be safely estimated as shown by the extensive studies of Nascimento et al. (2011, 2012, 2016). More details on the estimation of mixtures are discussed in Sect. 4.

2.3 Priors for the MGPD model

The MGPD model definition is completed by an appropriate prior distribution for Θ , as given in Nascimento et al. (2012). A gamma prior with shape c_i and mean d_i is assigned to each η_i , where these parameters may be chosen to achieve a large prior variance. The parameter space of $\boldsymbol{\mu}$ is restricted to

$$C(\boldsymbol{\mu}) = \{\boldsymbol{\mu} : 0 < \mu_i < \dots < \mu_n\},
 \tag{2}$$

to ensure the parameters’ identifiability (see Sect. 4 for details). To each μ_i is assigned an inverse gamma prior with shape a_i and mean b_i , where again these parameters may be chosen to achieve a large variance. The prior for $\boldsymbol{\mu}$ is thus $\pi(\boldsymbol{\mu}) = K \prod_{i \in [n]} f_{IG}(\mu_i|a_i, b_i) \mathbb{1}_{C(\boldsymbol{\mu})}(\boldsymbol{\mu})$, where $\mathbb{1}_A(x) = 1$ if $x \in A$ and zero otherwise, f_{IG} is the inverse gamma density and $K^{-1} = \int_{C(\boldsymbol{\mu})} \prod_{i \in [n]} f_{IG}(\mu_i|a_i, b_i) \partial \boldsymbol{\mu}$. The weights of the gamma mixture, w_i , are assigned a Dirichlet $D(\mathbf{1}_n)$ prior, where $\mathbf{1}_n$ is a vector of ones of dimension n .

The prior of the threshold u is normal with mean chosen around a high-order sample statistic, since the GPD approximation can be expected to hold only over the tail of the data. The variance is chosen so that the bulk of the prior distribution ranges roughly over data points in the upper half. Nascimento et al. (2012) empirically showed that for large datasets the value of the variance does not affect the estimation of the threshold, whilst for smaller datasets the above restriction may need to be imposed to correctly estimate the threshold u . This is the only partially informative prior used, but the hyper-parameter choices are guided by the theoretical results of Pickands (1975).

The hyper-parameters above can be changed to effectively include expert prior information without affecting our inferential routines.

For the shape and scale of the GPD, the uninformative prior of Castellanos and Cabras (2007) is used, defined as $\pi(\xi, \sigma) = \sigma^{-1}(1 + \xi)^{-1}(1 + 2\xi)^{-1/2}$.

3 Multivariate extremes

3.1 Asymptotics and models

Modelling approaches for multivariate extremes rely on limiting results of componentwise maxima and are mainly due to de Haan and Resnick (1977). One of these limiting results is briefly discussed next (see, e.g. Beirlant et al. 2004, for a comprehensive review).

Let $X_1, \dots, X_n \in \mathbb{R}_+^d$, where $X_i = (X_{ij})_{j \in [d]}$, be independent and identically distributed random vectors with marginal unit Fréchet distributions with cdfs $\exp(-1/x)$, $x \in \mathbb{R}_+$. If $M_n = (\max_{i \in [n]} X_{ij})_{j \in [d]}$ converges in distribution as $n \rightarrow \infty$ to a non-degenerate cdf R , then $R(\mathbf{x}) = \exp(-V(\mathbf{x}))$, where $V(\mathbf{x}) = d \int_{\mathcal{S}_d} \max_{i \in [d]} \omega_i/x_i K(d\omega)$, $\omega = (\omega_i)_{i \in [d]}$, \mathcal{S}_d is the d -dimensional unit simplex, i.e. $\mathcal{S}_d = \{\omega : \omega_i \geq 0, \sum_{i \in [d]} \omega_i = 1\}$, and K is a probability measure on \mathcal{S}_d satisfying the “mean” constraint $\int_{\mathcal{S}_d} \omega_i K(d\omega) = d^{-1}$. The function V is called *exponent measure*, whilst K is the *spectral measure*. The cdf R is called *multivariate extreme-value distribution* (MEVD).

The main point here is that the limiting distribution of the componentwise maximum M_n cannot be described in a parametric closed form, but consists of a nonparametric family characterized by the spectral functions respecting the “mean” constraint. The generality of this result has led to the definition of a variety of approaches to model multivariate extreme observations. We can broadly identify three different strategies:

- define a parametric submodel for either the exponent measure (Coles and Tawn 1991, 1994) or the spectral measure (Ballani and Schlather 2011; Boldi and Davison 2007);

- model in a nonparametric fashion the class of MEVD distributions (Einmahl and Segers 2009; Guillothe et al. 2011);
- construct models based on other theoretical justifications (De Carvalho and Davison 2014; Ramos and Ledford 2009; Wadsworth et al. 2017).

In all cases, data are usually transformed via the empirical cdf into Fréchet or uniform margins and then some of the data points, those considered “extreme”, are formally retained for inference. Having already discussed the difficulty of assessing such a threshold in the univariate case, the identification of extreme data points becomes even more critical in multivariate applications since there is no unique definition of threshold.

To illustrate this, consider the different bivariate threshold choices in Fig. 2. Figure 2a, b shows that an observation is extreme if it is beyond the threshold in all or in at least one component, respectively. These thresholds are usually utilized when estimating simultaneously marginal and joint features of the data. The threshold in Fig. 2c describes as extreme an observation such that the sum of its components is larger than a specified value and is often used when only modelling dependence. The last threshold in Fig. 2d is associated with the so-called partially censored approach: an observation below a marginal threshold in any component is supposed to be censored at the threshold.

Although the theoretical limiting result of maxima can be expected to hold in the region specified by the threshold in Fig. 2a, all other thresholds are more commonly utilized to increase the sample size effectively retained for inference. Furthermore, the choice of such thresholds is often driven by the type of analysis required or computational simplifications. A flexible method that takes into account the full dataset is developed here to avoid making the arbitrary choices of thresholds location and type.

3.2 Extreme dependence

MEVDs are asymptotic distributions and the strength of dependence given by them represents a measure of the

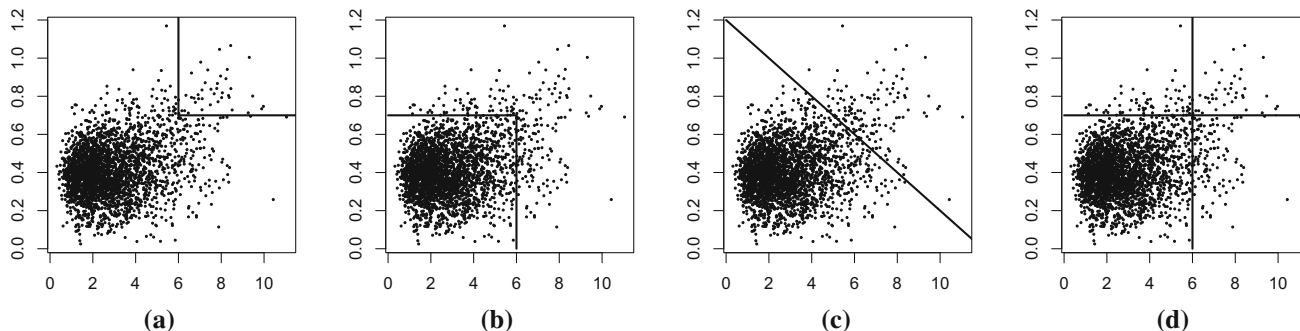


Fig. 2 Examples of bivariate threshold choices: solid lines—thresholds separating the bulk from the extreme region

asymptotic dependence in a random vector. However, in many practical applications dependent variables are observed to be asymptotically independent (Davison et al. 2013; Ledford and Tawn 1997; Ramos and Ledford 2009) and many commonly used distributions exhibit this behaviour: for example, the bivariate normal with correlation $\rho \in [-1, 1)$, $\rho \neq 0$. Due to a result of Berman (1961), multivariate extreme independence can be assessed by investigating all pairs of random variables. We thus focus on bivariate vectors. Sibuya (1960) proved that two random variables X_1 and X_2 with cdfs F_1 and F_2 are asymptotically independent iff the tail dependence χ is equal to zero, where $\chi = \lim_{u \rightarrow 1} \chi(u)$, and $\chi(u) = \mathbb{P}(F_1(X_1) > u | F_2(X_2) > u)$. For instance, for a bivariate MEVD distribution $\chi = 0$ iff X_1 and X_2 are independent, whilst $\chi = 0$ for any bivariate Gaussian with dependence $\rho \neq 1$. To address this deficiency of the MEVD distribution, novel extreme models that can take into account asymptotic dependence and independence have been proposed (e.g. Heffernan and Tawn 2004; Ramos and Ledford 2009; Wadsworth et al. 2017).

Since $\chi = 0$ for all asymptotically independent bivariate vectors, this criterion does not provide information about the subasymptotic strength of dependence for independent extremes. Coles et al. (1999) defined the subasymptotic tail dependence $\bar{\chi} = \lim_{u \rightarrow 1} \bar{\chi}(u)$, where

$$\bar{\chi}(u) = \frac{2 \log(\mathbb{P}(F_1(X_1) > u))}{\log(\mathbb{P}(F_1(X_1) > u, F_2(X_2) > u))} - 1.$$

If $\chi \in (0, 1]$ and $\bar{\chi} = 1$, then X_1 and X_2 are asymptotically dependent, whilst if $\chi = 0$ and $\bar{\chi} \in (-1, 1]$ then X_1 and X_2 are asymptotically independent. The strength of dependence increases with $\bar{\chi}$.

4 Copulae and mixtures

Given that all issues about marginals have been addressed, all is left is modelling dependence. Copulae are flexible functions to model complex relationships in a simple way. These only model the dependence structure of a random vector and allow for marginals to be defined separately (see Nelsen 2006, for a review). Their use to model multivariate extremes has been increasing over the past few years (e.g. Huser and Wadsworth 2019; Huser et al. 2017).

For a random vector $\mathbf{X} = (X_i)_{i \in [d]}$ with cdf F , whose margins have cdfs F_i , $i \in [d]$, a copula C is defined as a function $C : [0, 1]^d \rightarrow [0, 1]$ such that $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$, where $\mathbf{x} = (x_i)_{i \in [d]}$ is an instantiation of \mathbf{X} . Sklar (1959) proved that such a C linking marginal and joint distributions always exists. Notice that C is a cdf itself and, if absolutely continuous, possesses a density c called copula density and defined as $c(\mathbf{v}) =$

$\partial C(\mathbf{v})/\partial \mathbf{v}$, for $\mathbf{v} \in [0, 1]^d$. Thus the density of \mathbf{X} equals $f(\mathbf{x}) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i \in [d]} f_i(x_i)$, where f_i and f are the densities of X_i and \mathbf{X} respectively.

Copula functions and finite mixture models have recently been combined (e.g. Kim et al. 2013) to depict an even wider variety of patterns of dependence. Formally, a mixture of n copulae C_i is defined for $\sum_{i \in [n]} w_i = 1$ and $w_i \geq 0$ as $\sum_{i \in [n]} w_i C_i(F_1(x_1), \dots, F_d(x_d))$.

Our approach described in detail in Sect. 5 below is based on a finite mixture of copulae functions with marginals given by MGPLDs, thus including mixtures of gamma distributions. It therefore requires the identification of the number of components for each mixture as well as the estimation of their weights. However, this is easily handled. In practical applications, a small number of components, either copulae or gammas, is usually required as shown by Dey et al. (1995), Nascimento et al. (2012) and Rousseau and Mengersen (2011). In practice, this means that the weight of any extra component is estimated to be zero. Thus, mixture estimation can be carried out using two equivalent routines: by either separately fitting models with an increasing number of components until one is estimated to have zero weight, or by fitting one model only with a large-enough number of components so that only the required ones have nonzero weight (see Supplementary Material for an example). Alternatively, the choice of the number of components can be straightforwardly automated using a fully nonparametric approach, but this is typically not required (see, e.g. Fúquene Patiño 2015; De Waal and Van Gelder 2005).

A relevant aspect of mixture models is their inherent lack of identifiability, but this can be solved by imposing some restrictions over the parameters. Diebolt and Robert (1994) and Frühwirth-Schnatter (2001) impose order restrictions over the means in Gaussian mixtures. The same procedure is applied by Wiper et al. (2001) and Nascimento et al. (2012) over the means of a mixture of gammas, as reported in equation (2). Similarly, in our approach, we impose such an order restriction over the copulae components by ordering them according to their correlation coefficient. This guarantees that all components of our approach can be correctly identified.

5 The semiparametric approach

5.1 Likelihood

For each marginal, an MGPLD with density and cdf f_i and F_i , respectively, and parameters $\Theta_i = \{\mathbf{w}_i, \boldsymbol{\eta}_i, \boldsymbol{\mu}_i, \xi_i, \sigma_i, u_i\}$ is used, where $\mathbf{w}_i = (w_{ij})_{j \in [n_i]}$, $\boldsymbol{\eta}_i = (\eta_{ij})_{j \in [n_i]}$ and $\boldsymbol{\mu}_i = (\mu_{ij})_{j \in [n_i]}$ are the parameters of a mixture of n_i gammas as in equation (1). The dependence structure is modelled by a mixture of n copulae C_i with weights $\mathbf{w} = (w_i)_{i \in [n]}$ and

parameter set $\Theta_{D_i}, i \in [n]$. Letting $\Theta = \{\mathbf{w}, \Theta_{D_i}, \Theta_j : i \in [n], j \in [d]\}$, our cdf F is given by

$$F(\mathbf{x}|\Theta) = \sum_{i \in [n]} w_i C_i(F_1(x_1|\Theta_1), \dots, F_d(x_d|\Theta_d)|\Theta_{D_i})$$

and its density f equals

$$f(\mathbf{x}|\Theta) = \sum_{i \in [n]} w_i c_i(F_1(x_1|\Theta_1), \dots, F_d(x_d|\Theta_d)|\Theta_{D_i}) \times \prod_{j \in [d]} f_j(x_j|\Theta_j), \tag{3}$$

where c_i is the associated copula density, $i \in [n]$.

Although our approach does not require any restriction on the chosen copulae, in this work mixtures of elliptical copulae are used: more specifically, Gaussian (Song 2000), T (Demarta and McNeil 2005), skew-normal (Wu et al. 2014) and skew-T (Smith et al. 2012) copulae. Furthermore, all mixture components are assumed to belong to the same family, e.g. Gaussian. Such mixtures have the very convenient property of a known asymptotic behaviour: whilst mixtures of Gaussians and skew-normals have asymptotically independent extremes, Ts and skew-Ts exhibit extreme dependence (Kollo et al. 2017). We discuss below in our simulation study that mixtures combining different elliptical structures did not provide any additional information about the extreme structure.

Consider now bivariate vectors only. The specific form of our densities follows by substituting c_i in equation (3) with the expressions in Supplementary Material. Whilst for mixtures of Gaussian copulae, no constraints are imposed, for the other models, we impose the following constraints on the likelihood in equation (3):

- for T copulae, all components have the same number of degrees of freedom in \mathbb{R}_+ . This is to ensure that all components lead to the same asymptotic dependence structure;
- for skew-normal copulae, all components have the same skewness parameters. Our simulations showed that such parameters are particularly hard to estimate and could not be precisely estimated under more general settings;
- for skew-T copulae, one single component with integer degrees of freedom is used. This greatly speeds up computations using the formulae of Dunnett and Sobel (1954) and ensure all parameters can be correctly estimated.

As well as having closed-form expressions for marginal quantiles, bivariate quantiles can be easily deduced in our models. However, these are not uniquely defined since there are infinitely many pairs (x_1, x_2) such that $\mathbb{P}(X_1 > x_1, X_2 > x_2|\Theta)$ is equal to a specified number. Thus we look at

pairs (x_1, x_2) and compute the associated probability of joint exceedance $\mathbb{P}(X_1 > x_1, X_2 > x_2|\Theta)$. This is a function E of (x_1, x_2) and Θ defined as

$$E(x_1, x_2|\Theta) = 1 - F_1(x_1|\Theta_1) - F_2(x_2|\Theta_2) + \sum_{i \in [n]} w_i C_i(F_1(x_1|\Theta_1), F_2(x_2|\Theta_2)|\Theta_{D_i}). \tag{4}$$

Similarly, our approach leads to closed-form expressions for the probabilities $\chi(u|\Theta)$ and $\bar{\chi}(u|\Theta)$ appearing in the (subasymptotic) tail dependences. This is because, for instance, $\chi(u|\Theta) = E(F_1^{-1}(u|\Theta_1), F_2^{-1}(u|\Theta_2), \Theta) / \mathbb{P}(F_1(X_1|\Theta_1) > u)$ and these two probabilities have closed-form expressions.

5.2 Prior distribution

Our approach is completed by the introduction of a prior distribution, defined by considering separate blocks of parameters. The independent prior of each marginal parameter vector Θ_i is the one of Nascimento et al. (2012) and reported in Sect. 2.3. Our simulation study showed that to ensure the threshold location is correctly identified in all cases, 95% of the prior probability needs to roughly range between the 75th and 95th data quantile: such a prior is still in line with the theory of Pickands (1975) whilst giving enough uncertainty to allow the data to guide the choice of the threshold location.

For correlation coefficients ρ_i , a continuous uniform $\mathcal{U}[-1, 1]$ is selected. The joint $\pi(\boldsymbol{\rho})$ is defined over a restricted space as for the mean parameters of the gamma mixtures to ensure identifiability. For skew copulae, a continuous uniform $\mathcal{U}[-1 + \epsilon, 1 - \epsilon]$ is assigned to the skewness parameters δ_j , for an ϵ close to zero. The copulae mixture weights w_i are given a Dirichlet $D(\mathbf{1}_n)$. These priors are chosen to give uninformative prior beliefs.

For the degrees of freedom v of the T copula, the uninformative prior of Fonseca et al. (2008) is used, defined as

$$\pi(v) = \left(\frac{v}{v+3}\right)^{1/2} \left(\gamma\left(\frac{v}{2}\right) - \gamma\left(\frac{v+1}{2}\right) - \frac{2(v+3)}{v(v+1)^2}\right)^{1/2}, \quad v \in \mathbb{R}_+$$

where γ is the trigamma function (Abramowitz et al. 1965). For the skew-T copula with integer degrees of freedom a zero-truncated Poisson distribution with mean 25 is used. Sensitivity studies showed that this value enabled for the identification of both low and high number of degrees of freedom.

The overall *prior distribution* is then defined as

$$\pi(\Theta) = \pi(\mathbf{w})\pi(\Theta_D) \prod_{i \in [2]} \pi(\xi_i, \sigma_i)\pi(u_i)\pi(\mathbf{w}_i)\pi(\boldsymbol{\mu}_i) \times \prod_{j \in [n_i]} \pi(\eta_{ij}),$$

where $\Theta_D \subseteq \{\boldsymbol{\rho}, v, \delta_1, \delta_2\}$ and $\log(\pi(\Theta_D)) = \log(\pi(\boldsymbol{\rho})) + \mathbb{1}_{\Theta_D}(v) \log(\pi(v)) + \mathbb{1}_{\Theta_D}(\delta_1) \log(\pi(\delta_1)\pi(\delta_2))$. The set Θ_D is defined in such a way to encompass all elliptical copulae considered in this paper.

5.3 Posterior and predictive inference

For a sample $\mathbf{x} = (\mathbf{x}_i)_{i \in [m]}$, where $\mathbf{x}_i = (x_{1i}, x_{2i})$, the *posterior* log-density is then

$$\begin{aligned} \log \pi(\Theta|\mathbf{x}) &\propto \sum_{j \in [m]} \log \left(\sum_{i \in [n]} w_i c_i (F_1(x_{1j}|\Theta_1), F_2(x_{2j}|\Theta_2)|\Theta_{D_i}) \right) \\ &+ \sum_{i \in [2]} \log (f_i(x_{ij}|\Theta_i)) + \log(\pi(\Theta)). \end{aligned} \tag{5}$$

Inference cannot be performed analytically, and approximating MCMC algorithms are used. Parameters are divided into blocks and updating of the blocks follows Metropolis–Hastings steps since full conditionals have no recognizable form. Proposal variances are tuned via an adaptive algorithm as suggested in Roberts and Rosenthal (2009). Details are given in Supplementary Material. All algorithms are implemented in OX (Doornik 1996). The posterior in equation (5) is proper, and all parameters can be identified by our MCMC inferential routines.

Most quantities of interest in the analysis of extremes, e.g. $\chi(u|\Theta)$, are highly nonlinear functions of the models’ parameters. Thus, their posterior distribution cannot be derived analytically. However, the MCMC machinery enables us to derive an approximated distribution for *any* function of the models’ parameters. For instance, for I draws $\Theta^{(i)}$, $i \in [I]$, from the posterior $\pi(\Theta|\mathbf{x})$, the values $\chi(u|\Theta^{(i)})$ approximate the posterior distribution of $\chi(u|\Theta)$, given a sample \mathbf{x} . An estimate of the posterior mean is then $\frac{1}{I} \sum_{i \in [I]} \chi(u|\Theta^{(i)})$.

Estimation is an important task in extreme-value theory as much as the prediction of a new observation \mathbf{x}_{m+1} given a sample \mathbf{x} . The likelihood of a new observation can be summarized by the *predictive* distribution of joint exceedance $E(\mathbf{x}_{m+1}|\mathbf{x})$ given by

$$\begin{aligned} E(\mathbf{x}_{m+1}|\mathbf{x}) &= \int E(\mathbf{x}_{m+1}|\Theta)\pi(\Theta|\mathbf{x})d\Theta \\ &= \mathbb{E}_{\Theta|\mathbf{x}}(E(\mathbf{x}_{m+1}|\Theta)). \end{aligned}$$

This corresponds to the expectation of equation (4) with respect to the posterior $\pi(\Theta|\mathbf{x})$. This expectation cannot be computed analytically, but our Bayesian approach enables us to derive an approximated Monte Carlo estimate equal to $\frac{1}{I} \sum_{i \in [I]} E(\mathbf{x}_{m+1}|\Theta^{(i)})$.

5.4 Ascertainment of extreme independence

A critical task in the analysis of multivariate extremes is the determination of the asymptotic dependence structure. However, very few models are able to take into account both extreme dependence and independence, and consequently discriminate one from the other. Our semiparametric Bayesian approach enables us to introduce a fully probabilistic, new criterion for the ascertainment of asymptotic independence based on the posterior distribution of the number of degrees of freedom of the T copula.

Recall that T and skew-T copulae tend to Gaussian and skew-normal ones, respectively, when the number of degrees of freedom $v \rightarrow \infty$, and consequently large posterior estimates of the number of degrees of freedom may indicate asymptotically independent extremes. Thus, for a fixed $c \in \mathbb{R}_+$, we define the criterion $\phi(c) = \mathbb{P}(v \in [c, \infty)|\mathbf{x})$ which gives an uncertainty measure about the possibility that $\chi = 0$ and thus that extremes are independent. Of course, the assessment of the extreme dependence behaviour depends on the choice of c , but our simulations below give some guidance on how to choose this value and demonstrate that a fairly large interval of c values lead to similar conclusions. Values of $\phi(c)$ close to zero give an indication towards asymptotic dependence, whilst for $\phi(c)$ close to one, the evidence is towards asymptotic independence. Henceforth, we report the estimates of $\phi(c)$ only for the T mixtures since for this model the prior of the number of degrees of freedom is non-informative.

6 Simulations

A simulation study, performed to validate selection criteria for our models, is summarized next. Importantly, this exercise enables us to validate the use of the number of degrees of freedom to assess extreme dependence.

The study consists of 8 samples of size 1000 from a variety of dependence structures and marginals. Specifically, data are simulated from: a mixture of 2 Gaussian copulae with correlations (0.2, 0.9), weights (0.6, 0.4) and MGPD margins (2G); a skew-normal copula with dependence parameter 0.7, slant parameter vector $(-0.4, 0.6)$ (using the parametrization of Azzalini and Capitanio 1999) and MGPD margins (SN); a Morgenstern copula with dependence 0.75 and lognormal-GPD margins (MO); a bilogistic distribution (from Castillo et al. 2005, page 103) with depen-

dence 0.5 and lognormal margins (BL); a mixture of 2 T copulae with correlations (0.2, 0.9), weights (0.6, 0.4), 7 degrees of freedom and MGPD margins (2T); a skew-T copula with dependence parameter 0.7, slant parameter vector $(-0.4, 0.6)$, 5 degrees of freedom and MGPD margins (ST); an asymmetric logistic copula from Tawn (1988) with dependence parameter 0.5, asymmetry vector (0.3, 0.6) and lognormal-GPD margins (AL); a Cauchy copula with correlation coefficient 0.5 and lognormal margins (CA). Notice that datasets 2G, SN, MO and BL are asymptotically independent, whilst 2T, ST, AL and CA exhibit extreme dependence. For all datasets with GPD tails, the marginal thresholds are chosen at the theoretical value giving an exceedance probability of 0.1.

Priors are chosen as in Sect. 5.2. Prior means of μ_i and η_i , $i \in [2]$ are selected around the true values if available, or around reasonable values after visual investigation of the data histograms, but with large variances. The prior means of the thresholds are fixed at the 90th empirical quantile.

For all simulations, algorithms are run for 25000 iterations, with a burn-in of 5000 and thinning every 20, giving a posterior sample of 1000. Convergence is assessed by looking at trace plots of various functions of the parameters. In all cases, the number of gamma mixture components of each marginal is first chosen by fitting different MGPD models. The number of copula components is then identified by independently fitting models with an increasing number of components until the weight of one is estimated to be zero (as in Fig. 3). In all cases, no more than two components are required. Note, however, that all parameters, both those of the marginal MGPDs and those of the copula densities, are estimated jointly.

Figure 3 reports the summary $\phi(c)$ as a function of c for mixtures of T copulae estimated over the simulated datasets. With the exception of the slowest decaying solid line, which is associated with data simulated from a mixture of T with

7 degrees of freedom, $\phi(c)$ decays at a clearly different rate for asymptotically dependent and independent data. Depending on how one measures the loss of a misspecification of either behaviour, then different values of c can be chosen to devise a criterion to discriminate the extreme regimes. Here we take a neutral position and assume both regimes have the same probability ($\phi(c) = 0.5$). For this choice, values of c roughly in $[7, 15]$ discriminate the two asymptotic behaviours, where such discrimination appears to be strongest for $c = 10$. Thus, hereafter, $\phi = \phi(10)$ denotes our summary of evidence towards asymptotic independence.

This is confirmed by Table 1 which summarizes the posterior means and credibility intervals of the number of degrees of freedom: these are more concentrated around larger values in asymptotically independent datasets. Our chosen coefficient ϕ , taking notably larger values for asymptotically independent datasets in Table 1, gives a sound uncertainty statement about the possibility of asymptotic independence. The only exception is the dataset from a mixture of T copulae for which the true number of degrees of freedom is seven: thus a value for ϕ around 0.5 is not surprising.

Table 1 includes the estimates of the parameter δ of the model of Huser and Wadsworth (2019) which discriminates between asymptotic dependence and independence: for $\delta < 0.5$ extremes are independent whilst for $\delta > 0.5$ they are dependent. To fit the model, the R package SpatialADAI is used. Data are first transformed to uniform margins via the empirical cdf and thresholds at values 0.95 for each margins are used as in Huser and Wadsworth (2019). Different threshold choices did not affect the results. In most cases, the parameter estimate is close to the boundary case of 0.5 and most 95% confidence intervals include the possibility of either asymptotic behaviour, thus providing little information about the tail behaviour.

Standard model selection criteria, e.g. BIC (Schwarz 1978) and DIC (Spiegelhalter et al. 2002), reported in Table 2 and calculated using the parameters' posterior means, although giving guidance on the presence of skewness, do not provide information about extreme dependence, possibly because these are mostly influenced by the bulk of the data.

A model consisting of a mixture of Gaussian and T copulae (where all T are constrained to have the same number of degrees of freedom) is further fitted to all simulated datasets. Gaussian and T are chosen both for computational simplicity and because BIC and DIC criteria are overall lower for these models (unless data comes from skew-normal and skew-T models). These choices are natural since we validated their use in our simulation study. In all cases the weights of the Gaussian components are estimated to be zero and the estimates of the number of degrees of freedom coincide with the ones in Table 1. This exercise confirms that the main factor discriminating between asymptotic behaviours is indeed

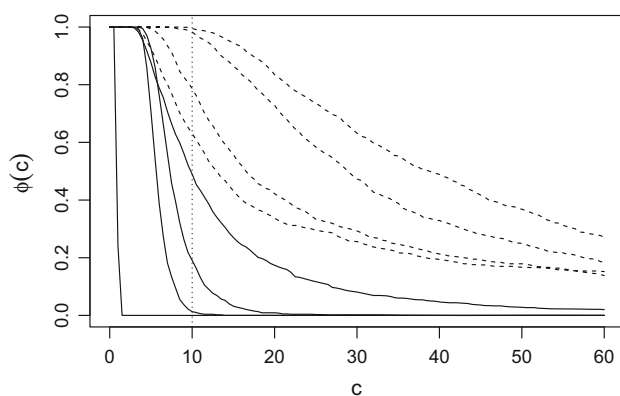


Fig. 3 Estimated $\phi(c)$ using mixtures of T copulae in our simulated dataset: solid line— asymptotically dependent data; dashed line— asymptotically independent data; dotted line— $c = 10$

Table 1 Posterior means and 95% credibility intervals of the number of degrees of freedom for the Student-T model with the largest number of nonzero weights (T), and the skew-T (ST) copula

	2G	SN	MO	BL
T	16.5 (5.8,141.5)	28.9 (10.2,135.8)	38.9 (13.0,154.3)	13.0 (4.0,157.9)
ST	4 (3,6)	19 (12,29)	20 (13,29)	23 (13,32)
ϕ	0.787	0.983	0.995	0.631
δ_{95}	0.42 (0.31,0.53)	0.38 (0.27,0.49)	0.36 (0.21,0.51)	0.18 (0,0.65)
	2T	ST	AL	CA
T	9.8 (3.6,51.9)	5.6 (3.9,9.3)	7.3 (4.4,16.0)	0.9 (0.8,1.1)
ST	3 (2,3)	6 (4,12)	8 (4,21)	1 (1,1)
ϕ	0.490	0.013	0.191	0
δ_{95}	0.48 (0.40,0.57)	0.48 (0.42,0.55)	0.13 (0,0.66)	0.60 (0.53,0.69)

The row ϕ reports our criterion for the model T and δ_{95} reports the estimate of the δ parameter of Huser and Wadsworth (2019) using a 0.95 threshold. The datasets used are: mixture of 2 Gaussian copulae (2G); skew-normal copula (SN); Morgenstern copula (MO); bilogistic copula (BL); mixture of 2 T copulae (2T); skew-T copula (ST); asymmetric logistic copula (AL); and Cauchy copula (CA)

Table 2 BIC and DIC scores of mixtures of Gaussian (G), T, skew-normal (SN) and skew-T (ST) models with the largest number of nonzero components when estimated over the simulated datasets

	2G		SN		MO		BL	
	BIC	DIC	BIC	DIC	BIC	DIC	BIC	DIC
G	9973	9604	9458	9424	9342	9255	9095	9008
T	9668	9635	9404	9489	9390	9190	9105	9004
SN	9986	9612	9609	9389	9367	9324	8988	9024
ST	9718	9632	9466	9446	9355	9260	9283	9157
	2T		ST		AL		CA	
	BIC	DIC	BIC	DIC	BIC	DIC	BIC	DIC
G	9866	9997	10832	10333	10501	10472	8972	8928
T	9882	9983	10774	10434	10492	10387	8953	9078
SN	9912	9991	10282	10010	10561	10427	8988	8932
ST	9939	10065	10278	9999	10901	10402	8940	8934

In bold are reported the lowest scores for each column. The datasets used are: mixture of 2 Gaussian copulae (2G); skew-normal copula (SN); Morgenstern copula (MO); bilogistic copula (BL); mixture of 2 T copulae (2T); skew-T copula (ST); asymmetric logistic copula (AL); and Cauchy copula (CA)

the number of degrees of freedom and consequently our ϕ criterion.

7 Applications

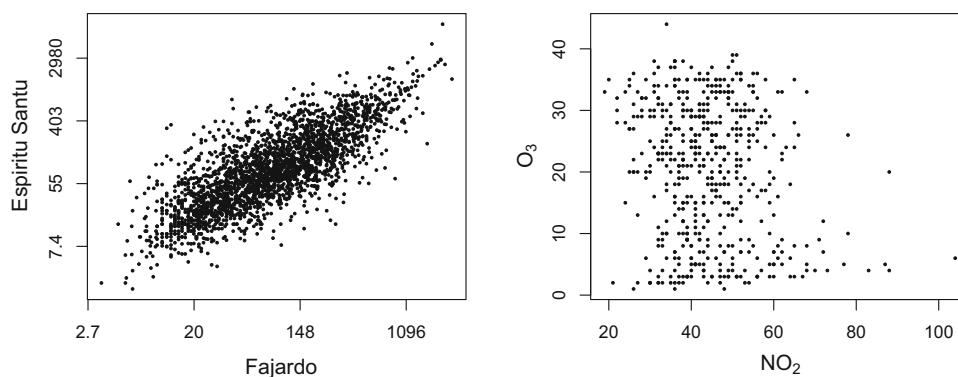
Two datasets from environmental applications are analysed next:

- weekly maxima from August 1966 to June 2016 of the flows of Fajardo and Espiritu Santu rivers in Puerto Rico, comprising 2492 observations (Nascimento et al. 2012);
- daily maxima of the hourly means during the winter months in 1994–1998 of NO₂/O₃ concentrations in Leeds, comprising 532 observations (Heffernan and Tawn 2004).

The Puerto Rico rivers dataset (Fig. 4a) is freely available at waterdata.usgs.gov, whilst the Leeds pollutants dataset (Fig. 4b) can be found in the `texmex` R package (Southworth et al. 2017). These are chosen for their apparent different asymptotic dependence: in Fig. 4 the Puerto Rico rivers seem to have strong extreme dependence, whilst the Leeds pollutants appear to have independent extremes (as noted in Heffernan and Tawn 2004). Some of the data points are not used for model fitting but to test the predictive capabilities of our and other approaches. Specifically, 1000 and 100 observations are randomly selected and discarded from the Puerto Rico rivers and Leeds contaminants datasets, respectively. These numbers are chosen to have test datasets smaller than the fitting ones and so to easily compute quantiles.

Our approach is compared against the asymptotically independent multivariate Gaussian tail model of Bortot et al. (2000), the best asymptotically dependent model in

Fig. 4 Scatterplot of the datasets from environmental applications analysed in Sect. 7



(a) Puerto Rico rivers' flows (in the log scale, with axes labels on the original scale) **(b)** Leeds pollutants concentrations in the winter months

the EVD R package (Stephenson 2002) and the models of Ramos and Ledford (2009) that can account for both dependent and independent extremes. For all these models, marginal thresholds were selected as in Ledford and Tawn (1997) at a high empirical quantile of the variable $\min(-\log(\hat{F}_1(X_1))^{-1}, -\log(\hat{F}_2(X_2))^{-1})$, where \hat{F} is the empirical cdf. In this study, different empirical quantiles of this variable are used, namely the 90, 95 and 97.5 quantiles². For each threshold and marginal, a GPD is first fitted to the exceedances using a POT approach and then the data are transformed into Fréchet margins via empirical cdf for data below the threshold and GPD cdf otherwise. Bivariate extreme models are fitted over the resulting datasets.

7.1 Model choice and measures of asymptotic dependence

The first step of the data analysis is to decide the number of mixture components for both the marginal gammas and the copulae. This is straightforwardly handled by following the procedure summarized in Sect. 4. Then the best copula mixture for each dataset is determined. Figure 5, reporting the coefficient $\phi(c)$, gives a clear indication of the asymptotic behaviour of the data: asymptotic independence for the Leeds pollutants and dependence for the Puerto Rico rivers. This is confirmed in Table 3 by the estimates of the number of degrees of freedom. The table further reports the estimates of the parameter δ of Huser and Wadsworth (2019). Whilst for the Leeds pollutants data the indication of asymptotic independence is strong, for the Puerto Rico rivers estimates are close to 0.5. Furthermore, these heavily depend on the threshold chosen: the estimate of δ indicates asymptotic dependence for a 0.9 threshold, and asymptotic independence otherwise (although in these last two cases the confidence interval includes 0.5). BIC and DIC scores are

² These values are chosen as they have been used in the literature (Ledford and Tawn 1997; Ramos and Ledford 2009)

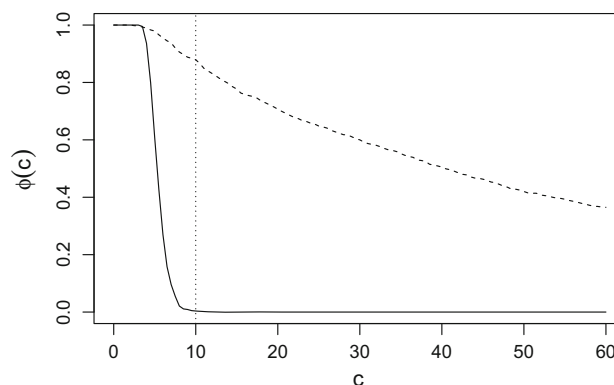


Fig. 5 Estimates of the criterion $\phi(c)$ as a function of c for applications of Sect. 7: River data (solid line); Leeds data (dashed line); $c = 10$ (dotted line)

reported in Table 4, which again strongly confirm asymptotic independence for Leeds pollutants. Since the posterior credibility intervals of the skewness parameters for all skew models include zero, we choose the Gaussian for the Leeds contaminants and the T for the Puerto Rico rivers as our favourite models.

In Fig. 6a, b, the posterior estimates of $\chi(u|\theta)$, $u \in [0.8, 1)$, for our chosen models are reported. For both applications, the posterior means give a good fit to the associated empirical estimates from the fitting and test datasets. These two diagrams give a further indication of asymptotic dependence for the Puerto Rico rivers, as $\chi(u|\theta)$ tends to a strictly positive value, and asymptotic independence for Leeds pollutants, as $\chi(u|\theta)$ goes to zero. Similar conclusions are drawn from the probabilities $\bar{\chi}(u|\theta)$ in the subsymptotic tail dependence reported in Fig. 6c, d. To the limit these confirm the asymptotic behaviour shown by $\chi(u|\theta)$, since for instance for the Leeds pollutants $\bar{\chi}(u|\theta)$ goes to -0.1 . Notice that our approach, utilizing the full dataset, can give model-based estimates of both $\chi(u|\theta)$ and $\bar{\chi}(u|\theta)$ for any $u \in (0, 1)$. These are reported in Supplementary Material.

Table 3 Measures of asymptotic dependence for our applications in Sect. 7: posterior mean and 95% credibility intervals of the number of degrees of freedom of the T copula model (column Mean); criterion ϕ ;

Puerto Rico rivers					Leeds pollutants		
Mean	ϕ	δ_{90}	δ_{95}	$\delta_{97.5}$	Mean	ϕ	δ_{80}
5.3 (3.8,7.9)	0.003	0.63 (0.59,0.67)	0.43 (0.28,0.58)	0.47 (0.36,0.58)	26.2 (7.7,133.2)	0.93	0.14 (0.02,0.26)

For the Leeds pollutants dataset, the threshold 0.8 is the largest one for which R did not return an error

Table 4 BIC and DIC scores of mixtures of Gaussian (G), T, skew-normal (SN) and skew-T (ST) models with the largest number of nonzero components when estimated over the datasets of the applications in Sect. 7

		G	T	SN	ST
River	BIC	39497	39445	39486	39518
	DIC	39618	39494	39259	39593
Leeds	BIC	7354	7359	7367	7370
	DIC	7379	7380	7382	7384

7.2 Predictions

The performance in extreme predictions of our approach is studied next. Marginally, as already noted in Nascimento et al. (2012), the MGPD can outperform the POT methodology. This is reported in Table 5 for the Puerto Rico rivers. Importantly, the table shows that joint modelling gives not only much narrower posterior credibility intervals than a simpler MGPD model, but also predicted values closer to the empirical ones. The same observation is true when comparing our approach with the results from the POT estimates and their confidence intervals.

The properties of the posterior distribution of $E(x_1, x_2|\Theta)$, the probability of joint exceedance of (x_1, x_2) , are summarized in Table 6 for a number of values of (x_1, x_2) exceeding the used thresholds, together with estimates from the other approaches considered as well as the empirical probabilities of the test data. Our approach outperforms competing ones for the Leeds pollutant dataset in all pairs, since our estimates are closer to the empirical ones than those from the other approaches. For the Puerto Rico rivers dataset, our estimates are closer to the empirical ones for all pairs but the one associated with an exceedance probability of 0.005. In all cases, the 95% posterior credibility intervals from our approach include the empirical probability. The posterior distributions of $E(x_1, x_2|\Theta)$ at a number of fixed values (x_1, x_2) are further reported in Fig. 7. Such distributions are in general not available using the approaches reviewed in Sect. 3.

Lastly, Fig. 8 reports the Monte Carlo estimates of the predictive probabilities of exceedance $E(x_{m+1}|\mathbf{x})$. Each point (x_1, x_2) of this map gives the probability of a future obser-

and estimates of the parameter δ of Huser and Wadsworth (2019) using various thresholds

vation that is larger than both x_1 and x_2 . These provide an intuitive description of the overall behaviour of the test datasets. Again, such predictive summaries are often not available for other approaches. Although the probabilities in Fig. 8b appear to be highly asymmetrical, this is mostly due to the effect of the marginals (see Supplementary Material for a map not affected by marginals). Thanks to the capability of our approach of modelling marginals and dependence simultaneously, exceedance probabilities can exhibit a variety of flexible forms.

7.3 Effect of the bulk on estimation of extreme dependence

An analysis over a subset of the full datasets, including only points considered extreme, is next carried out to ascertain whether the bulk of the data affects our tail estimation approach. The extreme points are selected as follows: first only observations that exceed the chosen thresholds in both marginals are retained (as in Fig. 2a); for the Puerto Rico rivers application, the threshold locations are chosen at the posterior means of the thresholds of the T copula model (giving 190 observations); for the Leeds pollutants the thresholds were selected to give a marginal probability of exceedance of 0.3 as in Heffernan and Tawn (2004) (giving 49 observations); lastly the margins of the resulting data points are transformed to the uniform scale via the empirical cdf.

Mixtures of T copulae are first fitted to these datasets to investigate whether the asymptotic dependence behaviour chosen by looking at the full dataset is confirmed when considering only extreme points. The results of this analysis, summarized in Table 7, confirm the asymptotic behaviours identified in Sect. 7.1, but give much larger posterior credibility intervals to the number of degrees of freedom and thus uncertainty about the true extreme regime.

Having assessed the asymptotic dependence structure over the extreme points only using our T model, the extreme-value copulae (Gudendorf and Segers 2010) associated with the T and Gaussian copulae are then fitted to the extreme datasets of the Puerto Rico rivers and Leeds pollutants applications, respectively. However, for the Puerto Rico rivers, a Gumbel copula given by

Fig. 6 Posterior estimates of $\chi(u|\Theta)$ and $\bar{\chi}(u|\Theta)$. Solid line: posterior mean—Shaded region: 95% posterior credibility interval—Dashed line: empirical estimate of fitted dataset—Dotted line: empirical estimate of test dataset. For the Leeds application, empirical estimates of $\bar{\chi}(u|\Theta)$ could be computed on a restricted interval only: (0.7, 0.89) for the fitted dataset and (0.7, 0.94) for the test dataset

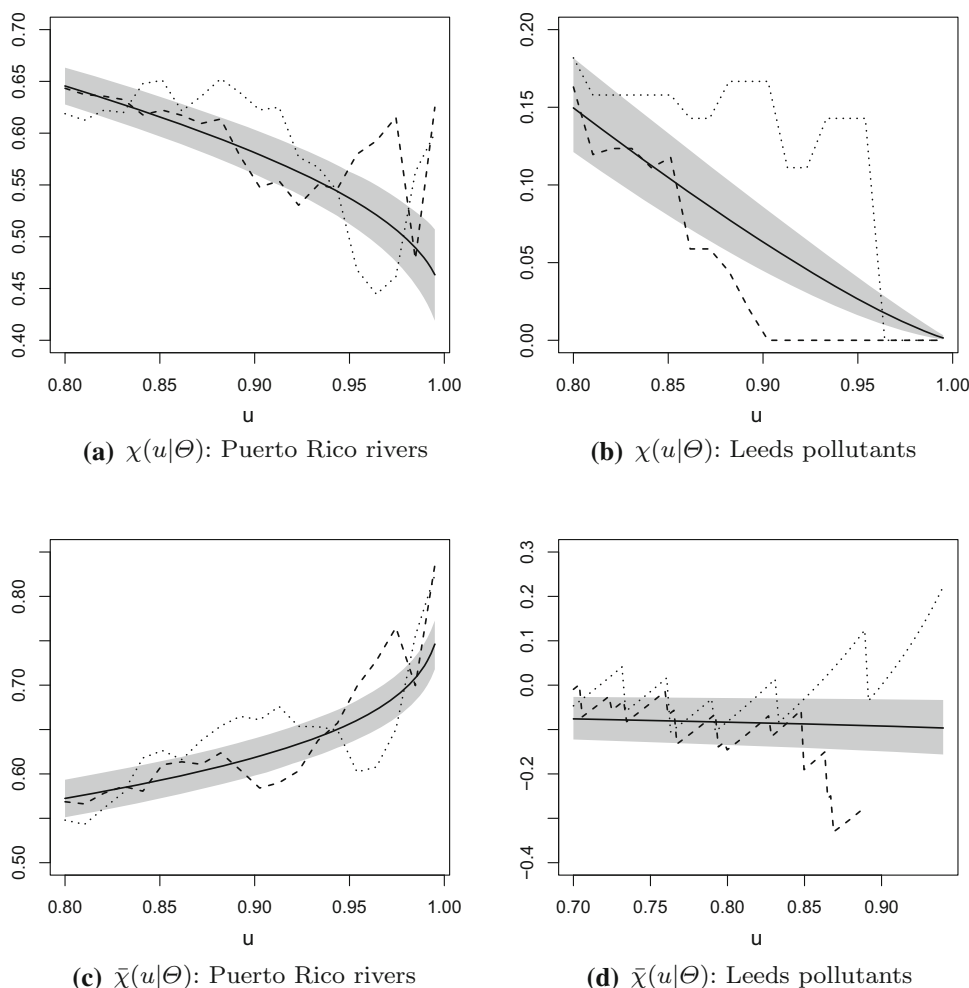


Table 5 Posterior summaries of $q(p|\Theta)$ for the Fajardo and Espiritu Santu rivers with $p = 0.005$: Empirical—empirical quantile from test dataset; Marginal—estimated quantiles using a marginal MGPD

model; Joint—estimated quantiles using our bivariate approach; POT—estimated quantiles using a POT approach at different thresholds

	Empirical	Marginal	Joint	POT 90	POT 95	POT 97.5
Fajardo	[1710,1800]	1900 (1554,2544)	1865 (1564,2289)	1881 (1583,2409)	1940 (1582,2692)	1943 (1636,2524)
Espiritu Santu	[1350,1380]	1463 (1215,1886)	1388 (1210,1663)	1465 (1237,1896)	1450 (1235,1869)	1445 (1251,1791)

$G(v_1, v_2) = \exp\left[-\left((-\log(v_1))^\theta + (-\log(v_2))^\theta\right)^{1/\theta}\right]$, for $v_1, v_2 \in [0, 1]$ and $\theta \in [1, +\infty)$ is used instead since this has an almost identical extreme dependence structure to the one of the extreme T copula, whilst being much simpler to estimate as discussed by Demarta and McNeil (2005). For the Leeds pollutants, a Gaussian copula is used since the associated extreme copula would simply be an independent one. Table 8 summarizes the posterior distributions of the relevant coefficients of tail dependence when estimated using

the full dataset or the extreme points only. In both cases, the posterior means are around the values of the empirical coefficients reported in Fig. 6, but importantly the credibility intervals are narrower for the full dataset.

8 Discussion

In this work, a new flexible approach for the estimation and prediction of extremes and joint exceedances was introduced.

The issue of model choice between model exhibiting different extreme dependence structures was investigated as well as the performance of our approach in extremes’ predictions. Furthermore, great attention was devoted to the identification of the extreme dependence behaviour by defining the new criterion ϕ which gives a probabilistic judgement on the possibility of asymptotic dependence. The results suggest that our Bayesian semiparametric approach gives a strong alternative to other bivariate approaches: the ϕ criterion robustly identifies the extreme dependence structure in our simulation study as well as in our applications.

A natural extension of the approach described here could consider two different copulae specifications in disjoint subsets of \mathbb{R}_+^2 , as for example done in Aulbach et al. (2012) and Vrac et al. (2007). Such subsets might correspond to the ones defined by the thresholds illustrated in Fig. 2. Such distinction would allow for the use of the full dataset whilst specifying a different dependence pattern for the extreme region, should one wish to do so. So, for instance, the likelihood could be defined as

$$f(x_1, x_2) = \begin{cases} c_b(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2), & \text{if } (x_1, x_2) \in B, \\ K_{c_t}(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2), & \text{otherwise,} \end{cases}$$

Table 6 Posterior summaries of $E(x_1, x_2|\Theta)$ for various (x_1, x_2) from our preferred models: T for Puerto Rico rivers and Gaussian (G) for Leeds Pollutants

	Puerto Rico rivers				Leeds pollutants	
	(x_1, x_2)				(x_1, x_2)	
	(720,730)	(900,780)	(1300,1100)		(55,32)	(58,33)
Emp. Pred.	0.015	0.010	0.005	Emp. Pred.	0.020	0.010
T	0.0175	0.0115	0.0044	G	0.0188	0.0104
95% CI	(0.0138,0.0220)	(0.0086,0.0149)	(0.0028,0.0069)	95% CI	(0.0126,0.0265)	(0.0065,0.0118)
EVD 90	0.0209	0.0141	0.0057	EVD 90	0.0549	0.0405
EVD 95	0.0214	0.0145	0.0058	EVD 95	0.0854	0.0607
EVD 97.5	0.0211	0.0154	0.0064	EVD 97.5	0.0875	0.0635
Bortot 90	0.0186	0.0122	0.0046	Bortot 90	0.0161	0.0085
Bortot 95	0.0205	0.0135	0.0050	Bortot 95	0.0133	0.0071
Bortot 97.5	0.0216	0.0153	0.0060	Bortot 97.5	0.0099	0.0050
Ramos 90	0.0203	0.0135	0.0054	Ramos 90	0.0114	0.0052
Ramos 95	0.0201	0.0136	0.0054	Ramos 95	0.0122	0.0049
Ramos 97.5	0.0207	0.0149	0.0062	Ramos 97.5	0.0093	0.0034

Emp. Pred. is the empirical estimate from the test dataset; EVD is the estimated probability from the best model in the EVD package; Bortot is the estimate from the model of Bortot et al. (2000); Ramos is the estimate from the model of Ramos and Ledford (2009). All these models are fitted using thresholds at the 90, 95 and 97.5 empirical levels

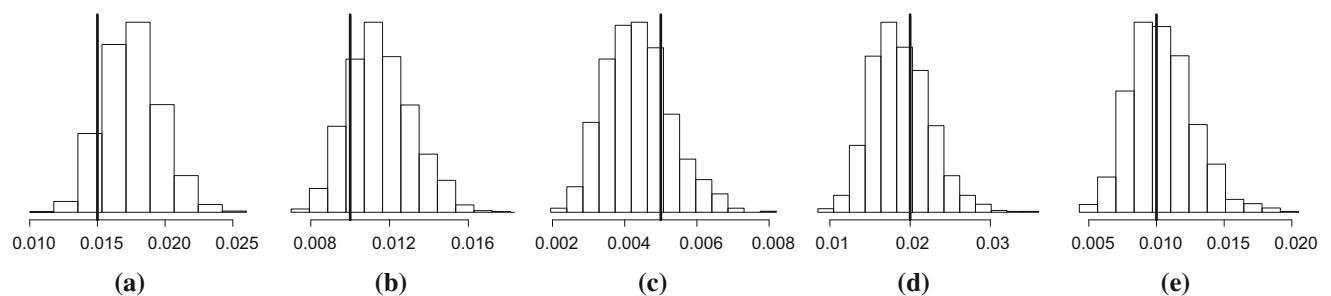


Fig. 7 Posterior distribution of the probability of joint exceedance $E(x_1, x_2|\Theta)$ at a number of values (x_1, x_2) : **a** $(x_1, x_2) = (720, 730)$; **b** $(x_1, x_2) = (900, 780)$; **c** $(x_1, x_2) = (1300, 1100)$; **d** $(x_1, x_2) = (55, 32)$; **e** $(x_1, x_2) = (58, 33)$. Figures **a–c** refer to the Puerto Rico rivers dataset, whilst figures **d–e** to the Leeds pollutants one. The vertical lines are the empirical predictive estimates

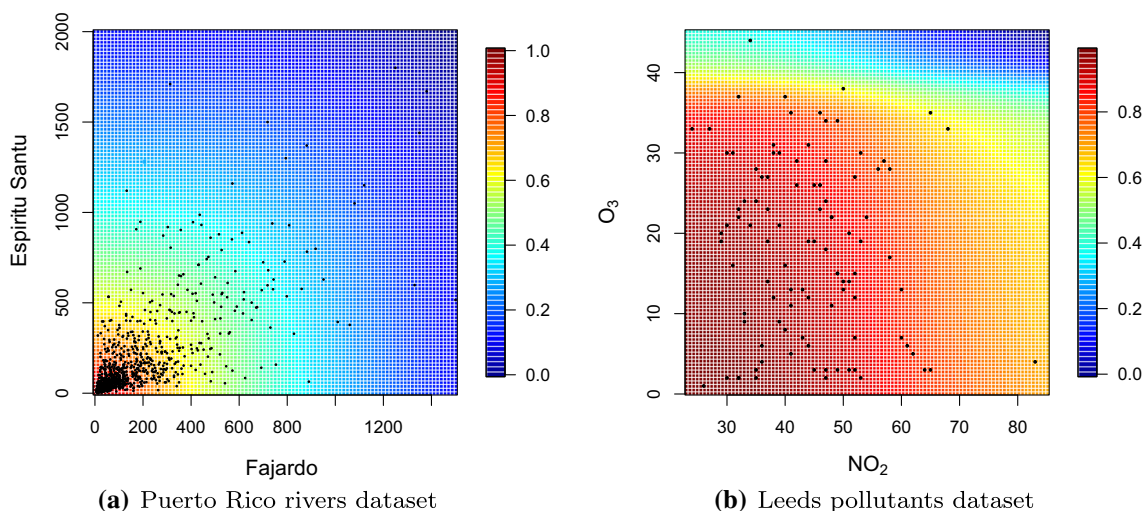


Fig. 8 Maps of the predictive probabilities of joint exceedance together with predictive datasets for the applications of Sect. 7

Table 7 Posterior mean of the number of degrees freedom of the T model, together with the 95% credibility interval, and ϕ criterion when estimated using only extreme observations in the applications of Sect. 7

	Mean	95% Int.	ϕ
Puerto Rico	9.89	(2.70,45.53)	0.25
Leeds	21.57	(2.74,107.89)	0.55

where c_b and c_t are two different copula densities, K is a normalizing constant and $B \subset \mathbb{R}^2$ is the region including non-extreme points. This more general specification brings in extra components and complications (K depends on model parameters in a non-trivial form) and handling them is not so straightforward. Solutions for these issues are the subject of ongoing research.

Although in this paper the focus was mainly on bivariate problems, multivariate extensions are readily available. For instance, mixtures of d -variate elliptical copulae could be

considered. A full definition of the approach would then be completed by an appropriate prior for the covariance matrix, for instance an inverse-Wishart, and an appropriate identification constraint for matrices, for example based on the determinant.

But more interestingly, since different pairs of variables could be defined to have a different asymptotic dependence, the overall density could be defined via vine-copulae (Bedford and Cooke 2002). For instance, in the trivariate case the overall density via a vine-copula decomposition can be written as

$$f(x_1, x_2, x_3) = c_{12}(F_1(x_1), F_2(x_2))c_{23}(F_2(x_2), F_3(x_3)) \\ c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \prod_{i \in [3]} f_i(x_i)$$

where $F_{i|j}(x_i|x_j) = \partial C_{ij}(F_i(x_i), F_j(x_j))/\partial F_j(x_j)$ and the c 's are bivariate copula densities. We are currently investigating such models.

Table 8 Posterior means and 95% credibility intervals for the tail dependences (Puerto Rico rivers) and subasymptotic dependence (Leeds pollutants)

	Puerto Rico rivers: χ		Leeds pollutants: $\bar{\chi}$	
Full dataset	0.45	(0.39,0.50)	Full dataset	-0.13 (-0.21,-0.04)
Extreme points	0.43	(0.35,0.51)	Extreme points	-0.23 (-0.48,0.08)

For Puerto Rico rivers, a T model was fitted over the full dataset and a Gumbel one over the extreme points only. For Leeds Pollutants, a Gaussian model was fitted in both cases

Acknowledgements The authors gratefully acknowledge CAPES and CNPq for financial support. Most of this work was carried out at the Instituto de Matemática at the Universidade Federal do Rio de Janeiro whilst ML held a CAPES postdoctoral fellowship. The authors gratefully thank Jonathan Tawn, Miguel de Carvalho, two anonymous referees and the Associate Editor for insightful comments on previous versions of the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables, vol. 55. Courier Corporation, Chelmsford (1965)
- Aulbach, S., Bayer, V., Falk, M.: A multivariate piecing-together approach with an application to operational loss data. *Bernoulli* **18**(2), 455–475 (2012)
- Azzalini, A., Capitanio, A.: Statistical applications of the multivariate skew normal distribution. *J. Stat. Soc. Ser. B* **61**(3), 579–602 (1999)
- Ballani, F., Schlather, M.: A construction principle for multivariate extreme value distributions. *Biometrika* **98**(3), 633–645 (2011)
- Bedford, T., Cooke, R.M.: Vines—a new graphical model for dependent random variables. *Ann. Stat.* **30**(4), 1031–1068 (2002)
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: *Statistics of Extremes: Theory and Applications*. Wiley, Chichester (2004)
- Berman, S.: Convergence to bivariate limiting extreme value distributions. *Ann. Inst. Stat. Math.* **13**(1), 217–223 (1961)
- Boldi, M.O., Davison, A.C.: A mixture model for multivariate extremes. *J. R. Stat. Soc. Ser. B* **69**(2), 217–229 (2007)
- Bortot, P., Coles, S., Tawn, J.: The multivariate Gaussian tail model: an application to oceanographic data. *J. R. Stat. Soc. Ser. C* **49**(1), 31–49 (2000)
- Camilo, D.C., de Carvalho, M.: Spectral density regression for bivariate extremes. *Stoch. Environ. Res. Risk Assess.* **31**(7), 1603–1613 (2017)
- Castellanos, M.E., Cabras, S.: A default Bayesian procedure for the generalized Pareto distribution. *J. Stat. Plan. Inference* **137**(2), 473–483 (2007)
- Castillo, E., Hadi, A.S., Balakrishnan, N., Sarabia, J.M.: *Extreme Value and Related Models with Application in Engineering and Science*. Wiley, New York (2005)
- Coles, S.G., Tawn, J.: Modelling extreme multivariate events. *J. R. Stat. Soc. Ser. B* **53**(2), 377–392 (1991)
- Coles, S.G., Tawn, J.A.: Statistical methods for multivariate extremes: an application to structural design (with discussion). *J. R. Stat. Soc. Ser. C* **43**(1), 1–48 (1994)
- Coles, S.G., Heffernan, J.E., Tawn, J.A.: Dependence measures for extreme value analyses. *Extremes* **2**(4), 339–365 (1999)
- Davison, A.C., Huser, R., Thibaud, E.: Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci.* **45**(5), 511–529 (2013)
- Davison, A.C., Smith, R.L.: Models for exceedances over high thresholds (with discussion). *J. R. Stat. Soc. Ser. B* **52**(3), 237–254 (1990)
- De Carvalho, M., Davison, A.C.: Spectral density ratio models for multivariate extremes. *J. Am. Stat. Assoc.* **109**(506), 764–776 (2014)
- De Waal, D.J., Van Gelder, P.H.A.J.M.: Modelling of extreme wave heights and periods through copulas. *Extremes* **8**(4), 345–356 (2005)
- Demarta, S., McNeil, A.J.: The t copula and related copulas. *Int. Stat. Rev.* **73**(1), 111–129 (2005)
- Dey, D., Kuo, L., Sahu, S.: A Bayesian predictive approach to determining the number of components in a mixture distribution. *Stat. Comput.* **5**(4), 297–305 (1995)
- Diebolt, J., Robert, C.: Estimation of finite mixture distributions by Bayesian sampling. *J. R. Stat. Soc. Ser. B* **56**(2), 363–375 (1994)
- Doornik, J.A.: *Ox: object oriented matrix programming, 4.1. console version*. Nuffield College, Oxford University, Oxford (1996)
- Dunnett, C.W., Sobel, M.: A bivariate generalization of Student’s t-distribution, with tables for certain special cases. *Biometrika* **41**(1/2), 153–169 (1954)
- Einmahl, J.H., Segers, J.: Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Stat.* **37**(5B), 2953–2989 (2009)
- Einmahl, J.H., Li, J., Liu, R.Y.: Thresholding events of extreme in simultaneous monitoring of multiple risks. *J. Am. Stat. Assoc.* **104**(487), 982–992 (2009)
- Engelke, S., Opitz, T., Wadsworth, J.: *Extremal dependence of random scale constructions* (2018). [arXiv:1803.04221](https://arxiv.org/abs/1803.04221)
- Fonseca, T.C., Ferreira, M.A.R., Migon, H.S.: Objective Bayesian analysis for the Student-t regression model. *Biometrika* **95**(2), 325–333 (2008)
- Frühwirth-Schnatter, S.: Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *J. Am. Stat. Assoc.* **96**(453), 194–209 (2001)
- Fúquene Patiño, J.A.: A semi-parametric Bayesian extreme value model using a Dirichlet process mixture of gamma densities. *J. Appl. Stat.* **42**(2), 267–280 (2015)
- Gamerman, D., Lopes, H.F.: *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC, Baton Rouge (2006)
- Gudendorf, G., Segers, J.: Extreme-value copulas. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) *Copula Theory and Its Applications*. Lecture Notes in Statistics, vol 198. Springer, Berlin, Heidelberg (2010)
- Guillotte, S., Perron, S., Segers, J.: Non-parametric Bayesian inference on bivariate extremes. *J. R. Stat. Soc. Ser. B* **73**(3), 377–406 (2011)
- de Haan, L., Resnick, S.I.: Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40**(4), 317–337 (1977)
- Heffernan, J.E., Tawn, J.A.: A conditional approach for multivariate extreme values (with discussion). *J. R. Stat. Soc. Ser. B* **66**(3), 497–546 (2004)
- Huser, R., Opitz, T., Thibaud, E.: Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spat. Stat.* **21**(A), 166–186 (2017)
- Huser, R., Wadsworth, J.L.: Modeling spatial processes with unknown extremal dependence class. *J. Am. Stat. Assoc.* **114**(525), 434–444 (2019)
- Kim, D., Kim, J., Liao, S., Jung, Y.: Mixture of D-vine copulas for modeling dependence. *Comput. Stat. Data Anal.* **64**, 1–19 (2013)
- Kollo, T., Gaida, P., Marju, V.: Tail dependence of skew t-copulas. *Commun. Stat.-Simul. Comput.* **46**(2), 1024–1034 (2017)
- Ledford, A.W., Tawn, J.A.: Modelling dependence within joint tail regions. *J. R. Stat. Soc. Ser. B* **59**(2), 475–499 (1997)
- Nascimento, F.F., Gamerman, D., Lopes, H.F.: Regression models for exceedance data via the full likelihood. *Environ. Ecol. Stat.* **18**(3), 495–512 (2011)
- Nascimento, F.F., Gamerman, D., Lopes, H.F.: A semiparametric Bayesian approach to extreme value estimation. *Stat. Comput.* **22**(2), 661–675 (2012)
- Nascimento, F.F., Gamerman, D., Lopes, H.F.: Time-varying extreme pattern with dynamic models. *TEST* **25**(1), 131–149 (2016)
- Nelsen, R.B.: *An Introduction to Copulas*. Springer, New York (2006)
- Pickands, J.: Statistical inference using extreme order statistics. *Ann. Stat.* **3**(1), 119–131 (1975)

- Ramos, A., Ledford, A.: A new class of models for bivariate joint tails. *J. R. Stat. Soc. Ser. B* **71**(1), 219–241 (2009)
- Roberts, G.O., Rosenthal, J.S.: Examples of adaptive MCMC. *J. Comput. Graph. Stat.* **18**(2), 349–367 (2009)
- Rousseau, J., Mengersen, K.: Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. Ser. B* **73**(5), 689–710 (2011)
- Sabourin, A., Naveau, P.: Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization. *Comput. Stat. Data Anal.* **71**, 542–567 (2014)
- Salvatori, G., de Michele, C., Kottegoda, N.T., Rosso, R.: *Extremes in Nature. An Approach Using Copulas*. Springer, Dordrecht (2007)
- Scarrott, C., MacDonald, A.: A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT* **10**(1), 33–60 (2012)
- Schwarz, G.: Estimating the dimension of a model. *Ann. Stat.* **6**(2), 461–464 (1978)
- Sibuya, M.: Bivariate extreme statistics. I. *Ann. Inst. Stat. Math.* **11**(2), 195–210 (1960)
- Sklar, M.: Fonctions de répartition à n dimension et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* **8**, 229–231 (1959)
- Smith, M.S., Gan, Q., Kohn, R.: Modelling dependence using skew T copulas: Bayesian inference and applications. *J. Appl. Econom.* **27**(3), 500–522 (2012)
- Song, P.X.K.: Multivariate dispersion models generated from Gaussian copula. *Scand. J. Stat.* **27**(2), 305–320 (2000)
- Southworth, H., Heffernan, J.E., Metcalfe, P.D.: *texmex: Statistical modelling of extreme values*. R package version 2.4 (2017)
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A.: Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* **64**(4), 583–639 (2002)
- Stephenson, A.G.: *evd: extreme value distributions*. R News **2**, (2002)
- Tawn, J.A.: Bivariate extreme value theory: models and estimation. *Biometrika* **75**(3), 397–415 (1988)
- Vrac, M., Naveau, P., Drobinski, P.: Modeling pairwise dependencies in precipitation intensities. *Nonlinear Process. Geophys.* **14**(6), 789–797 (2007)
- Wadsworth, J.L., Tawn, J.A.: Dependence modelling for spatial extremes. *Biometrika* **99**(2), 253–272 (2012)
- Wadsworth, J.L., Tawn, J.A., Davison, A.C., Elton, D.M.: Modelling across extremal dependence classes. *J. R. Stat. Soc. Ser. B* **79**(1), 149–175 (2017)
- Wan, P., Davis, R.A.: Threshold selection for multivariate heavy-tailed data. *Extremes* **22**(1), 131–166 (2019)
- Wiper, M., Rios Insua, D., Ruggeri, F.: Mixtures of gamma distributions with applications. *J. Comput. Graph. Stat.* **10**(3), 440–454 (2001)
- Wu, J., Wang, X., Walker, S.G.: Bayesian nonparametric inference for a multivariate copula function. *Methodol. Comput. Appl. Probab.* **16**(3), 747–763 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.