

A semiparametric approach for bivariate extreme exceedances

Manuele Leonelli

School of Mathematics and Statistics, University of Glasgow
and

Dani Gamerman

Instituto de Matemática, Universidade Federal do Rio de Janeiro

August 9, 2017

Abstract

Inference over tails is performed by applying only the results of extreme value theory. Whilst such theory is well defined and flexible enough in the univariate case, multivariate inferential methods often require the imposition of arbitrary constraints not fully justified by the underlying theory. In contrast, our approach uses only the constraints imposed by theory. We build on previous, theoretically justified work for marginal exceedances over a high, unknown threshold, by combining it with flexible, semiparametric copulae specifications to investigate extreme dependence. Whilst giving probabilistic judgements about the extreme regime of all marginal variables, our approach formally uses the full dataset and allows for a variety of patterns of dependence, be them extremal or not. A new probabilistic criterion quantifying the possibility that the data exhibits asymptotic independence is introduced and its robustness empirically studied. Estimation of functions of interest in extreme value analyses is performed via MCMC algorithms. Attention is also devoted to the prediction of new extreme observations. Our approach is evaluated through a series of simulations, applied to real data sets and assessed against competing approaches. Evidence demonstrates that the bulk of the data does not bias and improves the inferential process for the extremal dependence.

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 areas of applications, as for instance environmental sciences and finance amongst many. Evidence points out 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 for exposition 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 the imposition of ad-hoc assumptions about the asymptotic dependence structure: for instance by excluding the possibility of asymptotic independence. Furthermore, the application of such methods requires the arbitrary selection of datapoints considered “extreme”, usually selected as those that exceed a fixed threshold. However, this choice can greatly affect the inferential process (Scarrott and MacDonald, 2012). To overcome these difficulties a new easily interpretable, flexible approach is proposed here to investigate both marginal

and joint extreme behaviours that formally uses in a model-based fashion the full dataset. This combines some fairly new methodology for univariate extremes justified by the asymptotic theory for tails, with a flexible semiparametric dependence structure definition which does not require any assumption about the asymptotic dependence decay. Empirical evidence demonstrates in Section 4.4 below that in the applications considered the bulk of the data does not bias our inferential ascertainment of the asymptotic dependence structure.

Inference is carried out within the *Bayesian* paradigm using the *MCMC* machinery (Gamerman and Lopes, 2006), enabling us to straightforwardly deliver a wide variety of estimates and predictions of quantities of interest, e.g. high quantiles. Although our methods could be straightforwardly extended to a more general nonparametric approach, we are able to demonstrate below that our simpler and computationally less intensive methodology can capture diverse patterns of dependence, be them extremal or not.

In this work, as often in the literature, we focus on problems where extreme behaviour is of interest on the right tail only. However our approach could be easily extended to handle situations where interest is on both tails (e.g. Scarrott and MacDonald, 2012).

Before formally defining our approach, both univariate and multivariate extreme value theory and copulae functions are briefly introduced to highlight the relevance and the novelty of our methodology.

1.1 Univariate extreme value theory

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 if a random variable X with endpoint x_e is in the domain of attraction of a generalized extreme value distribution (see e.g. Beirlant et al., 2004) then $\lim_{u \rightarrow x_e} \mathbb{P}(X \leq x + u | X > u) = P(x)$, where P is the distribution function (df)

of the *generalized Pareto distribution* (GPD). The df P is defined as

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

for $u, \xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$, where the support is $x \geq u$ if $\xi \geq 0$ and $0 \leq x \leq u - \sigma/\xi$ 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 chosen threshold.

The POT approach has two serious drawbacks. First, only a small subset of the data points, those beyond the chosen 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](#)).

To overcome these deficiencies, 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 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 *MGPD* extreme value mixture model of [Nascimento et al. \(2012\)](#).

The MGPD model

The flexible MGPD model consists of a finite mixture of gamma distributions for the bulk coupled with a GPD for the tail. The parametrization of the gamma suggested

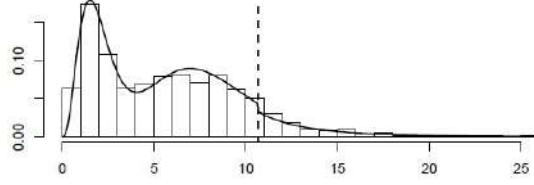


Figure 1: Example of a MGPD density fit consisting of a mixture of 2 gammas for the bulk.

in [Wiper et al. \(2001\)](#) in terms of shape, η , and mean, μ , parameters is used to avoid identifiability issues (e.g. [Richardson and Green, 1997](#)). Its density, g , is $g(x|\mu, \eta) = \Gamma(\eta)^{-1} (\eta/\mu)^\eta x^{\eta-1} \exp(-\eta x/\mu)$ and its df 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 df H of a finite mixture of n gammas are formally defined as

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

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 [Figure 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 df of a MGPD F is similarly defined in a piece-wise fashion. Whilst below the threshold u this is the df 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 $\mathbb{P}(X > q|\Theta) = 1 - p$ for p close to 1, 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) demonstrated that the MGPD can outperform standard POT models 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 method. The MGPD also provides better estimates than a standard nonparametric mixture model with an arbitrary large number of gamma components. Furthermore, a finite mixture is sufficient to model the bulk of the distribution since the weights of the required gamma components only are non-zero (Nascimento et al., 2012).

1.2 Multivariate extreme value theory

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 and refer to (see e.g. Beirlant et al., 2004, for a comprehensive review).

Let $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}_+^d$, where $\mathbf{X}_i = (X_{ij})_{j \in [d]}$, be independent and identically distributed random vectors with marginal unit Fréchet distributions with dfs $\exp(-1/x)$, $x \in \mathbb{R}_+$. If the componentwise maximum $\mathbf{M}_n = (\max_{i \in [n]} X_{ij})_{j \in [d]}$ converges in distribution as $n \rightarrow \infty$ to a non-degenerate df E , then $E(\mathbf{x}) = \exp(-V(\mathbf{x}))$, where $V(\mathbf{x}) = d \int_{\mathcal{S}_d} \max_{i \in [d]} \omega_i / x_i H(d\boldsymbol{\omega})$, $\boldsymbol{w} = (w_i)_{i \in [d]}$, \mathcal{S}_d is the d -dimensional unit simplex, i.e. $\mathcal{S}_d = \{\boldsymbol{\omega} : \omega_i \geq 0, \sum_{i \in [d]} \omega_i = 1\}$, and H is a probability measure on \mathcal{S}_d satisfying the “mean” constraint $\int_{\mathcal{S}_d} \omega_i H(d\boldsymbol{\omega}) = d^{-1}$. The function V is called *exponent measure*, whilst H is the *spectral measure*. The df E is called *multivariate extreme value distribution* (MEVD).

The main point here is that the limiting distribution of componentwise maxima 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; Jarušková, 2009) or the spectral measure (Ballani and Schlather, 2011; Boldi and Davison, 2007; Cooley et al., 2010);
- model in a nonparametric fashion the class of MEVD distributions (Einmahl and Segers, 2009; Guillotte et al., 2011);
- construct models based on alternative theoretical justifications (Bortot et al., 2000; De Carvalho and Davison, 2014; Ramos and Ledford, 2009; Wadsworth et al., 2017).

In all cases, data is usually transformed via the empirical df into Fréchet 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 Figure 2. Figures 2a and 2b state 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 contemporaneously marginal and joint features of the data. The threshold in Figure 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

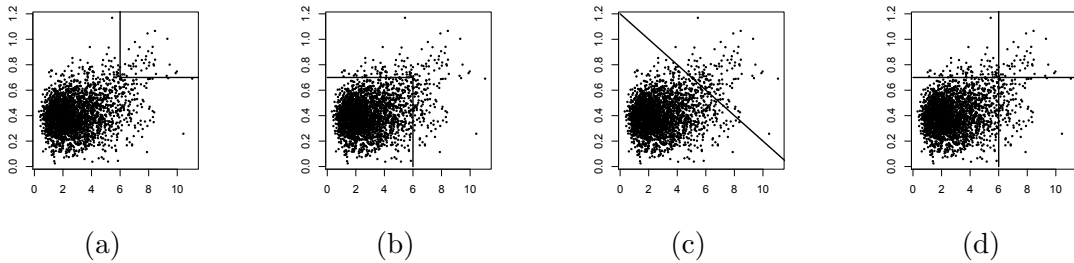


Figure 2: Examples of bivariate threshold choices.

Figure 2d is associated to the so called 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 Figure 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.

Furthermore, MGEV distributions assume a constant degree of dependence between pairs of rvs. However, in many practical applications dependent variables are observed to be asymptotically independent and many commonly used distributions exhibit this behavior: e.g. 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 dfs F_1 and F_2 are asymptotically independent iff the *coefficient of asymptotic independence* 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 MGEV 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 MGEV distribution, novel extreme models that can take into account asymptotic dependence and independence have been proposed (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 relative strength of dependence for independent extremes. Coles et al. (1999) defined the coefficient of subasymptotic 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 $\bar{\chi} = 1$ then X_1 and X_2 are asymptotically dependent, whilst if $\bar{\chi} \in (-1, 1)$ then X_1 and X_2 are asymptotically independent. The strength of dependence increases with $\bar{\chi}$.

1.3 Copulae

Having chosen to model the marginals as MGPDs, a tool to construct multivariate distributions with such given margins is needed. 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).

For a random vector $\mathbf{X} = (X_i)_{i \in [d]}$ with df F , whose margins have dfs F_i , $i \in [d]$, a *copula* C is defined as a function $C : [0, 1]^d \rightarrow [0, 1]$ such that $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$. Sklar (1959) proved that such a C linking marginal and joint distributions always exists. Notice that C is a df itself and as such 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.

Copulae 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 , $i \in [n]$, is defined as $\sum_{i \in [n]} w_i C_i(F_1(x_1), \dots, F_d(x_d))$, where $\sum_{i \in [n]} w_i = 1$ and $w_i \geq 0$.

1.4 Outline of the paper

Our approach and inferential routines are next described in Section 2. Section 3 presents a simulation study to both investigate their performance and address the issue of model choice. In Section 4 our methodology is applied to two real-world applications: river flows in Puerto Rico and NO_3/O_2 concentrations in the city of Leeds. We conclude with a discussion.

2 The semiparametric approach

2.1 Likelihood

For each marginal, an MGPD with density and df 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 Θ_{D_i} , $i \in [n]$. Letting $\Theta = \{\mathbf{w}, \Theta_{D_i}, \Theta_j : i \in [n], j \in [d]\}$, our df 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

$$f(\mathbf{x}|\Theta) = \sum_{i \in [n]} w_i c_i(F_1(x_1|\Theta_1), \dots, F_d(x_d|\Theta_d)|\Theta_{D_i}) \prod_{i \in [d]} f_i(x_i|\Theta_i), \quad (2)$$

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

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 behavior: whilst mixtures of Gaussians and skew-normals have asymptotically independent extremes, Ts and skew-Ts exhibit extreme dependence (Bortot, 2010).

Consider now bivariate vectors only. The specific form of our densities follows by substituting c_i in equation (2) with the expressions in the Supplementary Material. Simulation studies showed that, for full parameter identification, restrictions need to be imposed on the likelihood in equation (2). Whilst for mixtures of Gaussian copulae no constraints are imposed, for the other mixtures the following is assumed:

- for T-copulae all components have the same number of degrees of freedom in \mathbb{R}_+ ;
- for skew-Normal copulae all components have the same skewness parameters;
- for skew-T copulae one single component with integer degrees of freedom.¹

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

$$E(x_1, x_2 | \Theta) = 1 - F_1(x_1 | \Theta_1) - F_2(x_2 | \Theta_2) + \sum_{i \in [n]} w_i C_i(F_1(x_1 | \Theta_1), F_2(x_2 | \Theta_2) | \Theta_{D_i}). \quad (3)$$

¹This greatly speeds up computations using the formulae of Dunnett and Sobel (1954)

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

2.2 Prior distribution

Our approach is completed by the introduction of a prior distribution, defined by considering separate blocks of parameters.

For each of the marginal components Θ_i the priors specified in [Nascimento et al. \(2012\)](#) are used. Specifically, for the i -th marginal component, to each η_{ij} a gamma prior with shape c_{ij} and mean d_{ij} is assigned, where these parameters may be chosen to achieve a large prior variance. The parameter space of $\boldsymbol{\mu}_i$ is restricted to $C(\boldsymbol{\mu}_i) = \{\boldsymbol{\mu}_i : 0 < \mu_{i1} < \dots < \mu_{in_i}\}$ to address the identifiability issues of mixtures. To each μ_{ij} an inverse gamma prior with shape a_{ij} and mean b_{ij} is assigned, where again these parameters may be chosen to achieve a large variance. Therefore the prior for $\boldsymbol{\mu}_i$ is $\pi(\boldsymbol{\mu}_i) = K \prod_{j \in [n_i]} f_{IG}(\mu_{ij} | a_{ij}, b_{ij}) \mathbb{1}_{C(\boldsymbol{\mu}_i)}(\boldsymbol{\mu}_i)$, where $\mathbb{1}_A(x) = 1$ if $x \in A$ and zero otherwise, f_{IG} is the inverse gamma density and $K^{-1} = \int_{C(\boldsymbol{\mu}_i)} \prod_{j \in [n_i]} f_{IG}(\mu_{ij} | a_{ij}, b_{ij}) \partial \boldsymbol{\mu}_i$. The weights of the gamma mixture, w_{ij} , are assigned a Dirichlet $D(\mathbf{1}_{n_i})$ prior, where $\mathbf{1}_{n_i}$ is a vector of dimension n_i with ones in all entries.

The prior of the threshold u_i is normal as in [Nascimento et al. \(2012\)](#) and [Behrens et al. \(2004\)](#). Care must be exercised when specifying the hyperparameters of this distribution. The mean is chosen around a high order sample statistics. The variance is chosen so that the bulk, say 95%, of the prior distribution ranges roughly over data points larger than the median. These variances need to be slightly smaller than in the univariate MGPD model to ensure convergence.

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

For the shape and scale of the GPD distributions the uninformative prior of [Castellanos and Cabras \(2007\)](#) is used, defined as $\pi(\xi_i, \sigma_i) = \sigma_i^{-1}(1 + \xi_i)^{-1}(1 + 2\xi_i)^{-1/2}$, $i \in [d]$.

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

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

where ϕ is the trigamma function. 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 degrees of freedom.

The overall *prior distribution* is then defined as

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

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 so defined to encompass all elliptical copulae considered in this paper.

2.3 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

$$\log \pi(\Theta|\mathbf{x}) = \sum_{j \in [m]} \log \left(\sum_{i \in [n]} w_i c_i(F_1(x_{1j}|\Theta_1), F_2(x_{2j}|\Theta_2)|\Theta_{D_i}) \right) + \sum_{i \in [2]} \log(f_i(x_{ij}|\Theta_i)) + \log(\pi(\Theta)). \quad (4)$$

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 the Supplementary Material. All algorithms are implemented in OX ([Doornik, 1996](#)).

Most quantities of interest in the analysis of extremes, e.g. $\chi(u|\Theta)$, are highly non-linear 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

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

This corresponds to the expectation of equation (3) 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 $\frac{1}{I} \sum_{i \in [I]} E(\mathbf{x}_{m+1}|\Theta^{(i)})$.

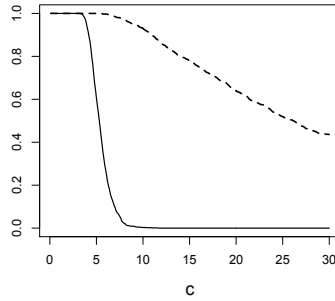


Figure 3: $\phi(c)$: River data (full line) - Leeds data (dashed line).

2.4 Ascertainment of extreme independence

A critical task in the analysis of 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. More importantly, to our knowledge none of these can deliver a probabilistic judgement about the data exhibiting either behavior. In contrast, our semiparametric Bayesian approach enables us to introduce a new probabilistic criterion for the ascertainment of asymptotic independence based on the posterior distribution of the degrees of freedom of the T copula.

Recall that for $\nu \rightarrow \infty$, T and skew-T copulae tend to Gaussian and skew-normal ones, respectively, and consequently large posterior estimates of the 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. Values of $\phi(c)$ close to zero give a strong indication of asymptotic dependence, whilst for $\phi(c)$ close to one the evidence is towards asymptotic independence. In our experience, the estimation of the number of degrees of freedom is more robust for T-copulae, possibly because not affected by prior

parameters. For these mixtures a value $c = 10$ seems to provide a sound uncertainty statement, as shown in Figure 3, where the function $\phi(c)$ from the analyses carried out in Section 4 below is reported. So for instance the solid line denotes $\phi(c)$ for a dataset that exhibits dependent extremes and $\phi(c) \approx 0$ for $c \geq 10$. Thus hereafter $\phi = \phi(10)$ denotes our summary of evidence towards asymptotic independence.

3 Simulations

A simulation study, performed to validate selection criteria for our mixtures, is summarized next. Importantly, this exercise enabled us to identify a variety of factors that together can provide a reliable toolkit to identify the strength of extreme dependence.

The study consisted of 8 samples of size 1000 from a variety of dependence structures and marginals. Specifically, data was simulated from: a mixture of 2 Gaussian copulae with MGPD margins (2G); a skew-Normal copula with MGPD margins (SN); a Morgenstern copula with lognormal-GPD margins (MO); a bilogistic copula with lognormal margins (BL); a mixture of 2 T copulae with MGPD margins (2T); a skew-T copula with MGPD margins (ST); an asymmetric logistic copula with lognormal-GPD margins (AL); a Cauchy copula with lognormal margins (CA). Notice that datasets 2G, SN, MO and BL are asymptotically independent, whilst 2T, ST, AL and CA exhibit extreme dependence.

Priors were chosen as in Section 2.2. Prior means of μ_i and η_i , $i \in [2]$, were selected around the true values if available, or around values that appeared reasonable after visual investigation of the data histograms, but with large variances. The prior means of the thresholds were fixed at the 90th empirical quantile.

For all simulations, the codes ran for 25000 iterations, with a burn-in of 5000 and thinning every 20, giving a posterior sample of 1000. Convergence was assessed by looking

Table 1: Number of non-zero copula component weights for Gaussian (G), skew-Normal (skew-N) and Student-T (T) mixtures.

	2G	SN	MO	BL	2T	ST	AL	CA
G	2	1	1	1	2	2	1	2
Skew-N	2	1	1	1	2	2	1	2
T	2	1	1	1	2	1	1	1

at trace plots of various functions of the parameters. In all cases, to reduce the number of models to be compared, the number of gamma mixture components of each marginal was first chosen by fitting different MGPD models. These numbers were then fixed when fitting various mixtures of copulae. Note however that all parameters, both those of the MGPDs and those of the copula densities, were estimated jointly.

First notice that, just as for gamma mixtures, only the required copula components have non-zero weights w_i as shown in Table 1 for the 2G and 2T datasets. Thus, more technical and computationally expensive nonparametric methods are not necessary. The number of mixture components further seems to give an indication of the data asymptotic behaviour: whilst for asymptotically independent datasets all mixtures have the same number of components (first four columns of Table 1), asymptotically independent models (Gaussian and Skew-Normal) need a larger number of components than dependent ones (T) for asymptotically dependent simulated data (last four columns of Table 1).

The posterior distributions of the degrees of freedom summarized in Table 2, being more concentrated around larger values in asymptotically independent datasets, provide a second reliable indicator of the data asymptotic behavior. This is confirmed by the coefficient ϕ which takes notably larger values for asymptotically dependent datasets (last line of Table 2). The only exception is the dataset from a mixture of T-copulae for which the true

Table 2: Posterior means and 95% credibility intervals of the degree of freedoms for the Student-T (T1), mixture of two Student-T (T2) and skew-T (ST1) copulae and ϕ criterion for the mixture of T-copulae with the most non-zero weights for each dataset.

	2G	SN	MO	BL	2T	ST	AL	CA
T1	3.2 (2.5,4.5)	28.9 (10.2,135.8)	38.9 (13.0,154.3)	13.0 (4.0,157.9)	2.4 (1.9,3.1)	5.6 (3.9,9.3)	7.3 (4.4,16.0)	0.9 (0.8,1.1)
T2	16.5 (5.8,141.5)	NA	NA	NA	9.8 (3.6,51.9)	NA	NA	NA
ST1	4 (3,6)	19 (12,29)	20 (13,29)	23 (13,32)	3 (2,3)	6 (4,12)	8 (4,21)	1 (1,1)
ϕ	0.787	0.983	0.995	0.631	0.490	0.013	0.191	0



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

Figure 4: Datasets from environmental applications.

number of degrees of freedom is seven: thus a value for ϕ around 0.5 is to be expected.

Standard model selection criteria, e.g. BIC (Schwarz, 1978) and DIC (Spiegelhalter et al., 2002), although giving guidance on the number of mixture components and on the presence of skewness, do not provide information about extreme dependence, possibly because these are mostly influenced by the bulk of the data (see the Supplementary Material).

4 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 (Figure 4a) is freely available at waterdata.usgs.gov, whilst the Leeds pollutants dataset (Figure 4b) can be found in R packages. These were chosen for their apparent different asymptotic dependence: in Figure 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](#)). In both cases some of the data points were not used for model fitting but to test predictive capabilities of both our and other approaches. Specifically, 1000 and 100 observations were selected at random and discarded from the Puerto Rico rivers and Leeds contaminants datasets, respectively.

Our approach is compared against the asymptotically independent multivariate Gaussian tail model of [Bortot et al. \(2000\)](#), the best asymptotically dependent model in the EVD R package ([Stephenson, 2002](#)) and the model 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 df. In this study different empirical quantiles of this variable were used, namely the 90, 95 and 97.5 quantiles². For each threshold and marginal, a GPD was first fitted to the exceedances using a POT approach and then the data was transformed into Frechét margins via empirical df for data below the threshold and GPD df otherwise. Bivariate extreme models were lastly fitted over the resulting datasets.

²These values were chosen as they have been used in the literature ([Ledford and Tawn, 1997](#); [Ramos and Ledford, 2009](#)).

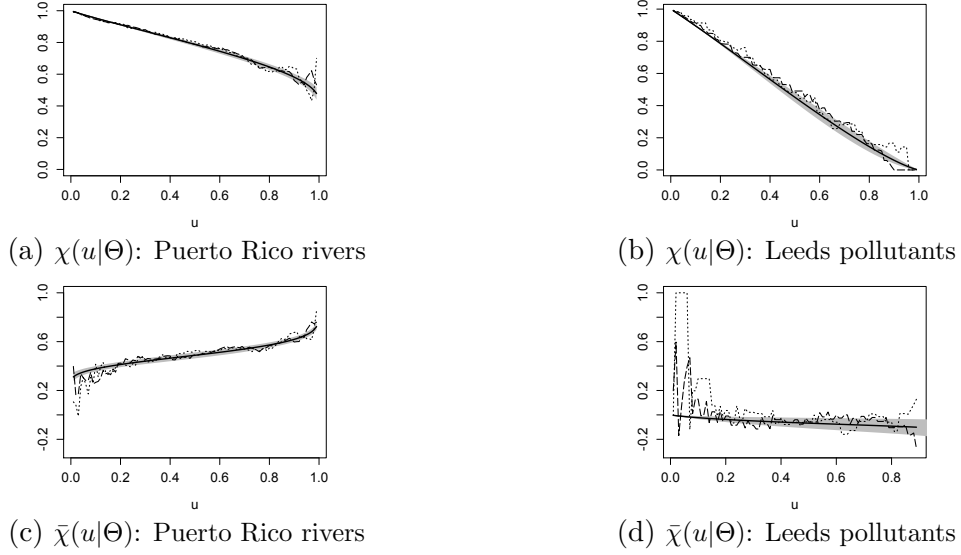


Figure 5: Posterior estimates of $\chi(u|\Theta)$ and $\bar{\chi}(u|\Theta)$. Full line: posterior mean - Shaded region: 95% posterior credibility interval - Dashed line: empirical estimate of fitted dataset - Dotted line: empirical estimate of test dataset.

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 coefficient of subasymptotic dependence reported in Figures 5c and 5d. To the limit these confirm the asymptotic behaviour shown by $\chi(u|\Theta)$, since for instance for the Puerto Rico rivers $\bar{\chi}(u|\Theta)$ goes to one.

4.3 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 4 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 properties of the posterior distributions of $E(x_1, x_2|\Theta)$ for various pairs (x_1, x_2)

Table 4: 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 the 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)	1875	1975	2031
Espiritu Santu	[1350,1380]	1463 (1215,1886)	1388 (1210,1663)	1464	1459	1477

Table 5: Posterior summaries of $E(x_1, x_2|\Theta)$ for various (x_1, x_2) and estimates from competing models.

Models	Puerto Rico rivers			Leeds pollutants		
	(x_1, x_2) (720,730)	(x_1, x_2) (900,780)	(x_1, x_2) (1300,1100)	(x_1, x_2) (55,32)	(x_1, x_2) (58,33)	
Emp. Pred.	0.015	0.010	0.005	Emp. Pred.	0.020	0.010
T1	0.0175	0.0115	0.0044	G1	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	0.071
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

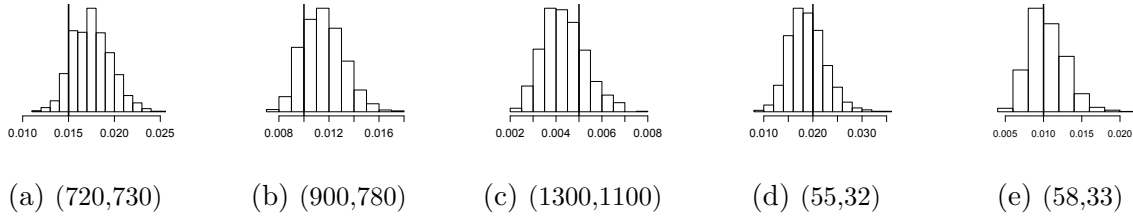


Figure 6: Posterior distribution of $E(x_1, x_2|\Theta)$ for various (x_1, x_2) - Figures (a)-(c): Puerto Rico rivers - Figures (d)-(e): Leeds pollutants - Vertical lines: empirical predictive estimates.

whose elements exceed the used thresholds are summarized in Table 5 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. For the Puerto Rico rivers dataset, our estimates are more accurate for all pairs but the one associated to an exceedance probability of 0.005. In all cases, the 95% posterior credibility intervals from our mixtures include the empirical probability. In Figure 6 is further reported the posterior distributions of $E(x_1, x_2|\Theta)$ for the pairs considered in Table 5: these are in general not available using the approaches reviewed in Section 1.2.

Lastly, Figure 7 reports the Monte Carlo estimates of the predictive probabilities of exceedance $E(\mathbf{x}_{m+1}|\mathbf{x})$. Each point (x_1, x_2) of this map gives the probability of a future observation that is larger than both x_1 and x_2 . These provide an intuitive description of the overall behavior of the test datasets. Again, such predictive summaries are often not available for other approaches.

4.4 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

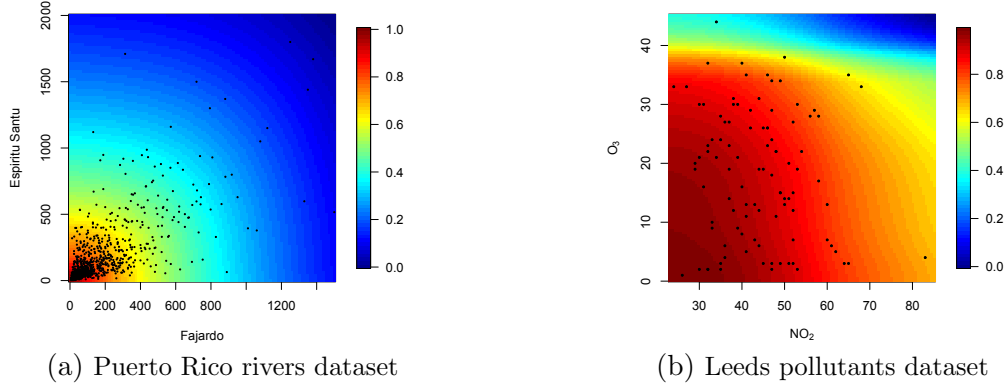


Figure 7: Predictive probabilities of joint exceedance together with predictive datasets.

approach. The extreme points are selected as follows: first only observations that exceed the chosen thresholds in both marginals are retained (as in Figure 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 marginals of the resulting data points are transformed to the uniform scale via the empirical df.

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 3b confirm the asymptotic behaviors identified in Section 4.2, but give much larger posterior credibility intervals to the degrees of freedom and thus uncertainty about the true extreme regime.

Having assessed the asymptotic dependence structure over the extreme points only, the extreme-value copulae (Gudendorf and Segers, 2010) associated to the T and Gaussian copulae are fitted to the extreme datasets of the Puerto Rico rivers and Leeds pollutants

Table 6: Posterior means and 95% credibility intervals for the coefficients of asymptotic dependence (Puerto Rico rivers) and sub-asymptotic 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)

applications, respectively. However, for the Puerto Rico rivers a Gumbel copula given by

$$G(v_1, v_2) = \exp\left[-\left(\left(-\log(v_1)\right)^\theta + \left(-\log(v_2)\right)^\theta\right)^{1/\theta}\right], \quad v_1, v_2 \in [0, 1], \theta \in [1, +\infty)$$

is used instead since this has an almost identical Pickands dependence function to the one of the extreme T-copula (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 6 summarizes the posterior distributions of the relevant coefficients of 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 Figure 5, but importantly the credibility intervals are narrower for the full dataset.

5 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 the various mixtures was investigated as well as the performance of our approach in extremes' predictions. The results suggest that our Bayesian semiparametric approach outperformed other bivariate approaches in predicting new extreme observations for the applications considered, whilst also allowing for the study of not only extreme but also overall dependence structures.

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.

A natural extension of the approach described here could consider two different copulae specification in disjoint subsets of \mathbb{R}_+^2 . Such subsets might correspond to the ones defined by the thresholds illustrated in Figure 2. Such distinction would allow for the use of the full dataset whilst specifying a different dependence pattern for the extreme region, should one wishes to do so. So for instance the likelihood could be defined as

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

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. The investigation of such models will be the focus of future research.

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 and Miguel de Carvalho for insightful comments on previous versions of the manuscript.

References

- Ballani, F. and M. Schlather (2011). A construction principle for multivariate extreme value distributions. *Biometrika* *98*, 633–645.
- Bedford, T. and R. M. Cooke (2002). Vines - a new graphical model for dependent random variables. *Ann. Statist.* *30*, 1031–1068.
- Behrens, C. N., H. F. Lopes, and D. Gamerman (2004). Bayesian analysis of extreme events with threshold estimation. *Stat. Model.* *4*, 227–244.
- Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of extremes: theory and applications*. Wiley, Chichester.

- Berman, S. (1961). Convergence to bivariate limiting extreme value distributions. *Ann. Inst. Statist. Math.* 13, 217–223.
- Boldi, M. O. and A. C. Davison (2007). A mixture model for multivariate extremes. *J. R. Statist. Soc. Ser. B* 69, 217–229.
- Bortot, P. (2010). Tail dependence in bivariate skew-Normal and skew-t distributions. Available online: www2.stat.unibo.it/bortot/ricerca/paper-sn-2.pdf.
- Bortot, P., S. Coles, and J. Tawn (2000). The multivariate Gaussian tail model: an application to oceanographic data. *J. R. Statist. Soc. Ser. C* 49, 31–49.
- Castellanos, M. E. and S. Cabras (2007). A default Bayesian procedure for the generalized Pareto distribution. *J. Statist. Plann. Inference* 137, 473–483.
- Coles, S. G., J. E. Heffernan, and J. A. Tawn (1999). Dependence measures for extreme value analyses. *Extremes* 2, 339–365.
- Coles, S. G. and J. Tawn (1991). Modelling extreme multivariate events. *J. R. Statist. Soc. Ser. B* 53, 377–392.
- Coles, S. G. and J. A. Tawn (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion). *J. R. Statist. Soc. Ser. C* 43, 1–48.
- Cooley, D., R. A. Davis, and P. Naveau (2010). The pairwise beta distribution: a flexible parametric multivariate model for extremes. *J. Multivariate Anal.* 101, 2103–2117.
- Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds (with discussion). *J. R. Statist. Soc. B* 52, 237–254.

- De Carvalho, M. and A. C. Davison (2014). Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association* 109(506), 764–776.
- de Haan, L. and S. I. Resnick (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 40, 317–337.
- Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *Int. Stat. Rev.* 73, 111–129.
- Doornik, J. A. (1996). *Ox: object oriented matrix programming, 4.1. console version*. Oxford: Nuffield College, Oxford University.
- Dunnett, C. W. and M. Sobel (1954). A bivariate generalization of Student’s t-distribution, with tables for certain special cases. *Biometrika* 41, 153–169.
- Einmahl, J. H. and J. Segers (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.* 37, 2953–2989.
- Einmahl, J. H. J., J. Li, and R. Y. Liu (2009). Thresholding events of extreme in simultaneous monitoring of multiple risks. *Journal of the American Statistical Association* 104(487), 982–992.
- Fonseca, T. C., M. A. R. Ferreira, and H. S. Migon (2008). Objective Bayesian analysis for the Student-*t* regression model. *Biometrika* 95, 325–333.
- Gamerman, D. and H. F. Lopes (2006). *Markov chain monte carlo: stochastic simulation for Bayesian inference*. CRC, Baton Rouge.
- Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In *Copula theory and its applications*, pp. 127–145. Springer.

- Guillotte, S., S. Perron, and J. Segers (2011). Non-parametric Bayesian inference on bivariate extremes. *J. R. Stat. Soc. Ser. B* 73, 377–406.
- Heffernan, J. E. and J. A. Tawn (2004). A conditional approach for multivariate extreme values (with discussion). *J. R. Statist. Soc. B* 66, 497–546.
- Jarušková, D. (2009). Modeling multivariate extremes of precipitation series in northern Moravia. *Environmetrics* 20, 751–775.
- Kim, D., J. Kim, S. Liao, and Y. Jung (2013). Mixture of D-vine copulas for modeling dependence. *Comput. Statist. Data Anal.* 64, 1–19.
- Ledford, A. W. and J. A. Tawn (1997). Modelling dependence within joint tail regions. *J. R. Statist. Soc. Ser. B* 59, 475–499.
- Nascimento, F. F., D. Gamerman, and H. F. Lopes (2012). A semiparametric Bayesian approach to extreme value estimation. *Stat. Comput.* 22, 661–675.
- Nelsen, R. B. (2006). *An introduction to copulas*. Springer, New York.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* 3, 119–131.
- Ramos, A. and A. Ledford (2009). A new class of models for bivariate joint tails. *J. R. Stat. Soc. Ser. B* 71, 219–241.
- Richardson, S. and P. J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. R. Statist. Soc. B* 59, 731–792.
- Roberts, G. O. and J. S. Rosenthal (2009). Examples of adaptive mcmc. *J. Comput. Graph. Statist.* 18, 349–367.

- Salvatori, G., C. de Michele, N. T. Kottegoda, and R. Rosso (2007). *Extremes in nature. An approach using copulas*. Springer, Dordrecht.
- Scarrott, C. and A. MacDonald (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT 10*, 33–60.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist. 6*, 461–464.
- Sibuya, M. (1960). Bivariate extreme statistics, I. *Ann. Inst. Statist. Math. Tokyo 11*, 195–210.
- Sklar, M. (1959). Fonctions de répartition à n dimension et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.
- Smith, M. S., Q. Gan, and R. Kohn (2012). Modelling dependence using skew T copulas: Bayesian inference and applications. *J. Appl. Econometrics 27*, 500–522.
- Song, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scand. J. Statist. 27*, 305–320.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B 64*, 583–639.
- Stephenson, A. G. (2002). evd: extreme value distributions. *R News 2*.
- Wadsworth, J. L., J. A. Tawn, A. C. Davison, and D. M. Elton (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B 79*(1), 149–175.
- Wiper, M., D. Rios Insua, and F. Ruggeri (2001). Mixtures of gamma distributions with applications. *J. Comput. Graph. Statist. 10*, 440–454.

Wu, J., X. Wang, and S. G. Walker (2014). Bayesian nonparametric inference for a multivariate copula function. *Methodol. Comput. Appl. Probab.* 16, 747–763.

A Copula densities

For all the copulae below we let $F_i(x_i|\Theta_i)$, $i \in [2]$, be the df of an MGPD.

Gaussian copula

In the bivariate case the Gaussian copula density depends on a correlation parameter $\rho \in [-1, 1]$ and can be written as

$$c(F_1(x_1|\Theta_1), F_2(x_2|\Theta_2)|\rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{2\rho z_1 z_2 - \rho^2(z_1^2 + z_2^2)}{2(1-\rho^2)}\right),$$

where $z_i = \Phi^{-1}(F_i(x_i|\Theta_i))$ and Φ is the standard univariate normal df.

T-copula

In the bivariate case the T-copula density depends on a correlation parameter $\rho \in [-1, 1]$ and degrees of freedom $v \in \mathbb{R}_+$ and can be written as

$$c(x_1, x_2|\rho, v) = \frac{\Gamma\left(\frac{v}{2}\right) \Gamma\left(\frac{v+2}{2}\right) \left(1 + \frac{z_1^2}{v} + \frac{z_2^2}{v} + \frac{z_1^2 z_2^2}{v^2}\right)^{(v+1)/2}}{\sqrt{1-\rho^2} \Gamma\left(\frac{v+1}{2}\right)^2 \left(1 + \frac{z_1^2 + z_2^2 - 2\rho z_1 z_2}{v(1-\rho^2)}\right)^{(v+2)/2}},$$

where $z_i = T_v^{-1}(F_i(x_i|\Theta_i))$ and T_v is the standard univariate T df with v degrees of freedom.

Skew-Normal copula

For $\rho \in [-1, 1]$ and $\delta_i \in (-1, 1)$, $i \in [2]$, define

$$\begin{aligned} \lambda_i &= \delta_i / \sqrt{1 - \delta_i^2}, & \psi &= \rho \sqrt{1 - \delta_1^2} \sqrt{1 - \delta_2^2} + \delta_1 \delta_2, \\ \alpha_1 &= \frac{\delta_1 - \delta_2 \psi}{((1-\psi^2)(1-\psi^2 - \delta_1^2 - \delta_2^2 + 2\psi\delta_1\delta_2))^{1/2}}, & \alpha_2 &= \frac{\delta_2 - \delta_1 \psi}{((1-\psi^2)(1-\psi^2 - \delta_1^2 - \delta_2^2 + 2\psi\delta_1\delta_2))^{1/2}}. \end{aligned} \tag{5}$$

The density of a bivariate skew-normal copula then depends on the parameters defined in equation (5) and can be written as

$$c(x_1, x_2 | \psi, \alpha_1, \alpha_2, \lambda_1, \lambda_2) = \frac{sn(z_1, z_2 | \psi, \alpha_1, \alpha_2)}{sn(z_1 | \lambda_1) sn(z_2 | \lambda_2)},$$

where $sn(z | \lambda) = 2\phi(z)\Phi(\lambda z)$ - with ϕ the density of a standard normal distribution - $z_i = SN^{-1}(F_i(x_i | \Theta_i) | \lambda_i)$ - with SN the df associated to the density sn - and $sn(z_1, z_2 | \psi, \alpha_1, \alpha_2) = 2\phi_\psi(x_1, x_2 | \psi)\Phi(\alpha_1 x_1 + \alpha_2 x_2)$ - with ϕ_ψ the density of a bivariate standard normal distribution with correlation ψ .

Skew-T copula

For $\rho \in [-1, 1]$, $v \in \mathbb{R}$ and $\delta_i \in (-1, 1)$, $i \in [2]$, define λ_i , ψ and α_i as in equation (5). The density of a bivariate skew-T copula can be written as

$$c(x_1, x_2 | \psi, v, \alpha_1, \alpha_2, \lambda_1, \lambda_2) = \frac{st(z_1, z_2 | \psi, v, \alpha_1, \alpha_2)}{st(z_1 | \lambda_1, v) st(z_2 | \lambda_2, v)},$$

where $st(z | \lambda, v) = 2t_v(z | v)T_v(\lambda z \sqrt{(v+1)/(z^2+v)})$ - with t_v the density of a standard univariate T with v degrees of freedom - $z_i = ST^{-1}(F_i(x_i | \Theta_i) | \lambda_i, v)$ - with ST the df associated to the density st - and

$$st(x_1, x_2 | \cdot) = 2t_{\psi, v}(x_1, x_2 | \psi, v)T_{v+2} \left(\frac{\alpha_1 x_1 + \alpha_2 x_2}{\sqrt{(x_1^2 + x_2^2 - 2\psi x_1 x_2 + v(1 - \psi^2)) / ((v+2)(1 - \psi^2))}} \right)$$

- with $t_{\psi, v}$ the density of a bivariate standard T distribution with v degrees of freedom and correlation ψ .

B MCMC algorithm

Sampling is carried out in blocks with Metropolis-Hastings proposals. At each iteration we first sample the copula parameters and copula mixture weights, and then the marginal

parameters for each of the marginals. For the marginals we use the steps outlined in [Nascimento et al. \(2012\)](#) and therefore we do not report them here. However, in our case the acceptance probabilities are computed with respect to the posterior in equation (4).

At iteration s parameters are updated as follows.

- Sampling ρ_i , for $i \in [k]$.

Since the correlation $\rho_i \in [-1, 1]$, the proposal kernel is taken as the truncated Normal distribution $\rho_i^* | \rho_i^{(s)} \sim \mathcal{N}(\rho_i(s), V_{\rho_i}) \mathbb{1}_{-1 \leq \rho_1^{(s+1)} < \dots < \rho_{i-1}^{(s+1)} < \rho_i^{(s)} < \dots < \rho_n^{(s)} \leq 1}$, where $\rho_i^{(s)}$ is the value of ρ_i at iteration s and V_{ρ_i} is the variance chosen to ensure appropriate chain mixing. The value $\rho_i^{(s+1)} = \rho_i^*$ is accepted with probability α_{ρ_i} , where

$$\alpha_{\rho_i} = \min \left\{ 1, \frac{\pi(\Theta^* | \mathbf{x}) f_N(\rho_i^{(s)} | \rho_i^*, V_{\rho_i}) \mathbb{1}_{\rho_1^{(s+1)} < \dots < \rho_i^* < \dots < \rho_k^{(s)}}}{\pi(\tilde{\Theta} | \mathbf{x}) f_N(\rho_i^* | \rho_i^{(s)}, V_{\rho_i}) \mathbb{1}_{\rho_1^{(s+1)} < \dots < \rho_i^{(s)} < \dots < \rho_k^{(s)}}} \right\},$$

where $\Theta^* = \{\boldsymbol{\rho}_{<i}^{(s+1)}, \rho_i^*, \boldsymbol{\rho}_{>i}^{(s)}, \Theta_C^{(s)}, \Theta_M^{(s)}\}$, $\tilde{\Theta} = \{\boldsymbol{\rho}_{<i}^{(s+1)}, \rho_i^{(s)}, \boldsymbol{\rho}_{>i}^{(s)}, \Theta_C^{(s)}, \Theta_M^{(s)}\}$, $\boldsymbol{\rho}_{<i}^{(s+1)} = (\rho_j^{(s+1)})_{j < i}$, $\boldsymbol{\rho}_{>i}^{(s)} = (\rho_j^{(s)})_{j > i}$, $\Theta_C^{(s)} \subseteq \{\mathbf{w}^{(s)}, \delta_1^{(s)}, \delta_2^{(s)}, v^{(s)}\}$ denotes the remaining copula parameters, which depend on the considered copula, at iteration s and $\Theta_M^{(s)}$ denotes all the marginal parameters at iteration s .

- Sampling \mathbf{w} .

The vector of copula weights is proposed from a Dirichlet distribution $\mathbf{w}^* \sim D(V_w \mathbf{w}^{(s)})$, where V_w is chosen to be equal to 50. So $\mathbf{w}^{(s+1)} = \mathbf{w}^*$ with probability α_w equal to

$$\alpha_w = \min \left\{ 1, \frac{\pi(\Theta^* | \mathbf{x}) f_D(\mathbf{w}^{(s)} | \mathbf{w}^*)}{\pi(\tilde{\Theta} | \mathbf{x}) f_D(\mathbf{w}^* | \mathbf{w}^{(s)})} \right\},$$

where $\Theta^* = \{\boldsymbol{\rho}^{(s+1)}, \mathbf{w}^*, \Theta_C^{(s)}, \Theta_M^{(s)}\}$, $\tilde{\Theta} = \{\boldsymbol{\rho}^{(s+1)}, \mathbf{w}^{(s)}, \Theta_C^{(s)}, \Theta_M^{(s)}\}$, and $\Theta_C^{(s)} \subseteq \{\delta_1^{(s)}, \delta_2^{(s)}, v^{(s)}\}$.

- Sampling δ_1 (skew-Normal and skew-T).

Since $\delta_1 \in (-1, 1)$, the proposal kernel is taken as the truncated Normal

$$\delta_1^* | \delta_1^{(s)} \sim \mathcal{N}(\delta_1^{(s)}, V_{\delta_1}) \mathbb{1}_{-1+\epsilon, 1-\epsilon}, \quad (6)$$

where V_{δ_1} is the variance of the proposal distribution chosen to ensure chain mixing.

So $\delta_1^{(s+1)} = \delta_1^*$ with probability α_{δ_1} equal to

$$\alpha_{\delta_1} = \min \left\{ 1, \frac{\pi(\Theta^* | \mathbf{x}) f_{N(-1+\epsilon, 1-\epsilon)}(\delta_1^{(s)} | \delta_1^*, V_{\delta_1})}{\pi(\tilde{\Theta} | \mathbf{x}) f_{N(-1+\epsilon, 1-\epsilon)}(\delta_1^* | \delta_1^{(s)}, V_{\delta_1})} \right\}, \quad (7)$$

where $\Theta^* = \{\boldsymbol{\rho}^{(s+1)}, \mathbf{w}^{(s+1)} \delta_1^*, \Theta_C^{(s)}, \Theta_M^{(s)}\}$, $\tilde{\Theta} = \{\boldsymbol{\rho}^{(s+1)}, \mathbf{w}^{(s+1)} \delta_1^{(s)}, \Theta_C^{(s)}, \Theta_M^{(s)}\}$, $\Theta_C^{(s)} = \{\delta_2^s, v^{(s)}\}$ and $f_{N(-1, -1)}$ denotes the density of a Normal truncated in $(-1, 1)$.

- Sampling δ_2 (skew-Normal and skew-T).

The proposal and acceptance of δ_2 follows the same steps as in equations (6)-(7) by substituting δ_1 with δ_2 and defining $\Theta^* = \{\boldsymbol{\rho}^{(s+1)}, \mathbf{w}^{(s+1)} \delta_1^{(s+1)}, \delta_2^*, \Theta_C^{(s)}, \Theta_M^{(s)}\}$, $\tilde{\Theta} = \{\boldsymbol{\rho}^{(s+1)}, \mathbf{w}^{(s+1)} \delta_1^{(s+1)}, \delta_2^{(s)}, \Theta_C^{(s)}, \Theta_M^{(s)}\}$ and $\Theta_C^{(s)} \subseteq \{v^{(s)}\}$.

- Sampling v (T and skew-T).

For $v \in \mathbb{R}_+$, v^* is proposed from a gamma $G(v^{(s)}, (v^{(s)})^2/V_v)$, where V_v is the variance of the proposal distribution chosen to ensure chain mixing. So $v^{(s+1)} = v^*$ with probability α_v equal to

$$\alpha_v = \min \left\{ 1, \frac{\pi(\Theta^* | \mathbf{x}) f_G(v^{(s)} | v^*, (v^*)^2/V_v)}{\pi(\tilde{\Theta} | \mathbf{x}) f_G(v^* | v^{(s)}, (v^{(s)})^2/V_v)} \right\},$$

where $\Theta^* = \{\Theta_C^{(s+1)}, v^*, \Theta_M^{(s)}\}$, $\tilde{\Theta} = \{\Theta_C^{(s+1)}, v^{(s)}, \Theta_M^{(s)}\}$ and $\Theta_C^{(s+1)} \subseteq \{\boldsymbol{\rho}^{(s+1)}, \mathbf{w}^{(s+1)}, \delta_1^{(s+1)}, \delta_2^{(s+1)}\}$.

For $v \in \mathbb{N}$, v^* is proposed from a discrete uniform distribution in $\{v^{(s)} - 2, v^{(s)} - 1, v^{(s)}, v^{(s)} + 1, v^{(s)} + 2\}$. So $v^{(s+1)} = v^*$ with probability α_v equal to

$$\alpha_v = \min \left\{ 1, \frac{\pi(\Theta^* | \mathbf{x})}{\pi(\tilde{\Theta} | \mathbf{x})} \right\}.$$

C BIC and DIC scores

Table 7: BIC and DIC scores of mixtures whose components have non-zero weights in our simulation study.

	2G		SN		MO		BL		2T		ST		AL		CA	
	BIC	DIC	BIC	DIC	BIC	DIC	BIC	DIC	BIC	DIC	BIC	DIC	BIC	DIC	BIC	DIC
G1	9998	9860	9458	9424	9342	9255	9095	9008	10012	10175	10846	10705	10501	10472	8923	9072
G2	9973	9604	NA	NA	NA	NA	NA	NA	9866	9997	10832	10333	NA	NA	8972	8928
T1	9884	9657	9404	9489	9390	9190	9105	9004	9900	10007	10774	10434	10492	10387	8953	9078
T2	9668	9635	NA	NA	NA	NA	NA	NA	9882	9983	NA	NA	NA	NA	NA	NA
SN1	10050	9693	9609	9389	9367	9324	8988	9024	10064	10226	10279	9865	10561	10427	8938	9091
SN2	9986	9612	NA	NA	NA	NA	NA	NA	9912	9991	10282	10010	NA	NA	8988	8932
ST1	9718	9632	9466	9446	9355	9260	9283	9157	9939	10065	10278	9999	10901	10402	8940	8934

Table 8: BIC and DIC scores of mixtures whose components have non-zero weights in our applications.

		G1	G2	T1	SN1	SN2	ST1
River	BIC	39518	39497	39445	39538	39486	39518
	DIC	39747	39618	39494	39896	39259	39593
Leeds	BIC	7354	NA	7359	7367	NA	7370
	DIC	7379	NA	7380	7382	NA	7384