



HAL
open science

Estimation of extreme quantiles conditioning on multivariate critical layers

Elena Di Bernardino, Fátima Palacios-Rodríguez

► **To cite this version:**

Elena Di Bernardino, Fátima Palacios-Rodríguez. Estimation of extreme quantiles conditioning on multivariate critical layers . 2016. hal-01201727v2

HAL Id: hal-01201727

<https://hal.science/hal-01201727v2>

Preprint submitted on 18 Feb 2016

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

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

Estimation of extreme quantiles conditioning on multivariate critical layers

E. Di Bernardino* and F. Palacios-Rodríguez^{†‡}

Thursday 11th February, 2016

Abstract

Let $T_i := [X_i | \mathbf{X} \in \partial L(\alpha)]$, for $i = 1, \dots, d$, where $\mathbf{X} = (X_1, \dots, X_d)$ is a risk vector and $\partial L(\alpha)$ is the associated multivariate critical layer at level $\alpha \in (0, 1)$. The aim of this work is to propose a non-parametric extreme estimation procedure for the $(1 - p_n)$ -quantile of T_i for a fixed α and when $p_n \rightarrow 0$, as the sample size $n \rightarrow +\infty$. An extrapolation method is developed under the Archimedean copula assumption for the dependence structure of \mathbf{X} and the von Mises condition for marginal X_i . The main result is the Central Limit Theorem for our estimator for $p = p_n \rightarrow 0$, when n tends towards infinity. A set of simulations illustrates the finite-sample performance of the proposed estimator. We finally illustrate

*Département IMATH, Laboratoire Cédric EA4629, CNAM, Paris, France.

[†]Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Sevilla, Sevilla, España.

[‡]*Correspondence to: Ms. F. Palacios-Rodríguez, Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Sevilla, Calle Tarfia sin número, Sevilla, 41012, España. E-mail: fpalacios2@us.es.*

how the proposed estimation procedure can help in the evaluation of extreme multivariate hydrological risks.

Keywords: Multivariate risk measures, return levels, critical layers, extreme quantile.

1 Introduction

Multivariate extreme events and multivariate critical layers In hydrology, phenomena are usually characterized by extreme events, and quantification of the risk of the occurrence of a specific extreme event is gaining attention in environmental sciences (e.g., see Hochrainer-Stigler and Pflug (2012)). The classic univariate measure in environmental sciences is that of the *return level*. This quantity represents the magnitude of an event that occurs at a given time and at a given site. More precisely, the return level is the quantile x_p which expresses the magnitude of the event that is exceeded with a probability equal to p , with $p = 1/T$ (T is called the *return period*). Environmental risks frequently involve several variables that are often correlated. For instance, a flood can be described by the volume, the peak and the duration (see Chebana and Ouarda (2011a,b)). It is therefore crucial to identify extreme risks in a multivariate setting.

The notion of multivariate return level is not univalent (Vandenberghe et al. (2012)) and several definitions can be found in the recent literature. Since different combinations of probabilities may produce the same return period, a multivariate return level is inherently ambiguous. Events that have equal probability of exceedance define iso-hyper-surfaces, otherwise known as *critical layers*. Salvadori et al. (2011) provide the following definition.

Definition 1.1. Let $\mathbf{X} = (X_1, \dots, X_d)$ be a random risk vector with joint distribution function F . For $\alpha \in (0, 1)$ and $d \geq 2$, the critical layer $\partial L(\alpha)$ at level α is defined as $\partial L(\alpha) = \{\mathbf{x} \in \mathbb{R}^d : F(\mathbf{x}) = \alpha\}$.

Definition 1.1 provides a partition composed of three probability regions: $\partial L^<(\alpha) = \{\mathbf{x} \in \mathbb{R}^d : F(\mathbf{x}) < \alpha\}$ (the sub-critical region); $\partial L(\alpha)$ (the critical region where all the events have a constant F); and $\partial L^>(\alpha) = \{\mathbf{x} \in \mathbb{R}^d : F(\mathbf{x}) > \alpha\}$ (the super-critical region).

In practice, at any occurrence of the phenomenon, only these three mutually exclusive events may occur (see Belzunce et al. (2007)). The multivariate return level can be defined with respect to one of the above three areas. For instance, in hydrology, the sub-critical region may be of interest if droughts are to be investigated, while the study of floods may require the use of super-critical regions (Salvadori et al. (2012)). Chebana and Ouarda (2011b) propose a parametric estimator for the critical layer and apply it to a bivariate flood real data-set. An estimation for the bivariate critical layers $\partial L(\alpha_n)$ by assuming $\alpha_n \rightarrow 1$, as $n \rightarrow \infty$, is presented in de Haan and Huang (1995). In the last decade, Embrechts and Puccetti (2006), Nappo and Spizzichino (2009), Prékopa (2012), and Cousin and Di Bernardino (2013) have used the notion of critical layers (i.e., multivariate quantile curves) to generalize the notion of univariate return level in the multivariate setting.

Conditional distribution on critical layers and considered multivariate extreme return levels We now introduce the conditional random variable X_i on the critical layer $\partial L(\alpha)$

for $i = 1, \dots, d$, that is,

$$T_i := [X_i | \mathbf{X} \in \partial L(\alpha)], \quad \text{for } \alpha \in (0, 1), \quad (1)$$

where $\partial L(\alpha)$ is as in Definition 1.1, and the associated conditional distribution function is given by $F_i(x | \alpha) = \mathbb{P}[X_i \leq x | \mathbf{X} \in \partial L(\alpha)]$. One can interpret the random variable T_i in Equation (1) as the contribution (or the responsibility) of the marginal risk X_i in the case where the whole risk vector \mathbf{X} belongs to the multivariate stress scenario represented by the critical layer $\partial L(\alpha)$, for some suitable level $\alpha \in (0, 1)$. It should be borne in mind that Equation (1) can also be written by using the Multivariate Probability Integral Transformation (MPIT) $Z := F(X_1, \dots, X_d)$. Indeed, under regularity conditions on F , one can write $T_i := [X_i | Z = \alpha]$, for $\alpha \in (0, 1)$ (see, e.g., Cousin and Di Bernardino (2013), Nelsen (2006), Genest and Rivest (2001)).

Using the above notation, in this paper we consider the multivariate return level, based on the critical layers, that was recently proposed by Di Bernardino et al. (2015). This risk measure is defined by the $(1 - p)$ -quantile of the random variable T_i in (1), that is,

$$x_p^i := U_{T_i} \left(\frac{1}{p} \right), \quad \text{for } p \in (0, 1), \quad (2)$$

where $U_{T_i}(t) := F_i^{\leftarrow}(1 - \frac{1}{t} | \alpha)$ for $t > 1$, and $F_i^{\leftarrow}(\cdot | \alpha)$ denotes the left-continuous inverse of $F_i(\cdot | \alpha)$.

Proposal of the present work The goal of this paper is to estimate the multivariate extreme return levels $x_{p_n}^i$ defined by Equation (2). To this end, two problematic points can be identified:

- i)* Di Bernardino et al. (2015) analyse this measure and introduce a semi-parametric estimation procedure. However, the aforementioned semi-parametric estimation and empirical quantile estimators perform well only if the threshold is not too high. These methods cannot handle extreme events, that is $p_n \ll 1/n$, which are specifically required for hydrological and environmental risk measures.
- ii)* As pointed out before, the considered conditional random variable T_i relies on the latent MPIT Z , which is not observed. Therefore, in order to apply a quantile estimation procedure, Z has to be previously estimated. This type of plug-in procedure increases the variance of the final estimation and introduces statistical difficulties (see, e.g., Di Bernardino et al. (2013a)).

In order to overcome the drawback outlined in item *i)*, in the present work we provide an estimator of $x_{p_n}^i$ in Equation (2), for a fixed α and when $p_n \rightarrow 0$, as $n \rightarrow +\infty$, by using Extreme Value Theory (EVT). In order to model the dependence structure of the multivariate risk vector \mathbf{X} , we consider Archimedean copulas. Recall that Archimedean copulas can be written as $C_\phi(u_1, \dots, u_d) = \phi^{-1}(\phi(u_1) + \dots + \phi(u_d))$, for all $(u_1, \dots, u_d) \in [0, 1]^d$, where the function ϕ is called the generator of the copula, and ϕ is a continuous, strictly decreasing function from $[0, 1]$ to $[0, \infty]$ such that $\phi(1) = 0$ and $\phi(0) = \infty$ (see, e.g., Definition 2 in McNeil and Nešlehová (2009)). Furthermore, the generator of an Archimedean copula also satisfies several additional d -monotony properties (for further details, see Theorem 2.2 in McNeil and Nešlehová (2009)). The rationale for employing Archimedean copulas is motivated by the fact that under this

assumption the distribution of T_i and its tail index can be easily obtained (see Proposition 2.1). Furthermore, in this framework, one can avoid having to previously estimate the latent variable Z (see item *ii*). Indeed, the proposed estimator procedure is only based on quantities that can be directly estimated by using the observed d -dimensional independent and identically distributed (*i.i.d.*) sample of (\mathbf{X}_j) , for $j = 1, \dots, n$ (see Equation (8)).

We remark that Archimedean copulas play a central role in the understanding of dependencies of multivariate random vectors (see Nelsen (2006), McNeil and Nešlehová (2009), Durante and Salvadori (2010)).

Frequently, hydrological phenomena are characterized by upper tail dependence described by Gumbel-Logistic models (e.g., see Fawcett and Walshaw (2012), Chebana and Ouarda (2011a), de Waal et al. (2007)).

Following these considerations, under a regular variation condition for the generator ϕ and the von Mises condition for the marginal X_i , we develop an extreme extrapolation technique in order to estimate $x_{p_n}^i$ (see, e.g., Cai et al. (2015)).

Organization of the paper In Section 2.1, we derive the tail index of the distribution of T_i . Under suitable assumptions, a non-parametric estimation procedure for $x_{p_n}^i$ is obtained, for a fixed level α and when $p_n \rightarrow 0$, as $n \rightarrow +\infty$ (Section 2.2). The main result is the asymptotic convergence of our estimator with $p = p_n \rightarrow 0$, as $n \rightarrow \infty$ (Section 3). In Section 4, the performance of the estimator $\hat{x}_{p_n}^i$ is illustrated on simulated data. Finally, Section 5 concludes with an application to a 3-dimensional rainfall data-set in order to illustrate how the proposed

estimation procedure can help in the evaluation of multivariate extreme return levels.

2 Proposed extreme estimator for the multivariate return level x_p^i

2.1 Tail index for T_i

In this section, we aim to study the tail behaviour of T_i , for $i = 1, \dots, d$. We assume the existence of the limit in $[1, \infty]$ of

$$\rho = - \lim_{s \uparrow 1} \frac{(1-s)\phi'(s)}{\phi(s)}. \quad (3)$$

Equation (3) is equivalent to regular variation of ϕ at 1 with index ρ , that is, $\phi \in RV_\rho(1)$ (see Charpentier and Segers (2009) for details). Furthermore, $\rho \geq 1$ due to the convexity of ϕ . When $\rho > 1$, the upper tail of the copula exhibits asymptotic dependence, while if $\rho = 1$, then the upper tail exhibits asymptotic independence. Under condition (3) for the generator ϕ , we now study the maximum domain of attraction (MDA) of T_i , for $i = 1, \dots, d$.

Proposition 2.1 (The von Mises condition for T_i). *Let (X_1, \dots, X_d) be a random vector with Archimedean copula with twice differentiable generator ϕ . Assume that $\phi \in RV_\rho(1)$, with $\rho \in [1, +\infty]$. Let $i \in \{1, \dots, d\}$ and F_i be the twice differentiable distribution function of X_i . Assume that F_i verifies the von Mises condition with index $\gamma_i \in \mathbb{R}$. Let T_i be as in (1) with distribution function $F_i(\cdot|\alpha)$.*

i) If $\rho \in [1, +\infty)$, then $F_i(\cdot|\alpha)$ verifies the von Mises condition with tail index $\gamma^{T_i} = \frac{\gamma_i}{\rho}$.

Specifically, $T_i \in MDA(\gamma^{T_i})$.

ii) If $\rho = +\infty$, then $F_i(\cdot|\alpha)$ verifies the von Mises condition with tail index $\gamma^{T_i} = 0$. In particular, $T_i \in MDA(0)$.

The proof and illustrations of Proposition 2.1 are given in the Supporting Information. Furthermore, the von Mises condition and the explicit form of the distribution $F_i(\cdot|\alpha)$ can also be found in the Supporting Information.

Remark 2.1. *Note that γ^{T_i} depends on neither the risk level α nor on the dimension d . However, γ^{T_i} depends on the domain of attraction of the respective margin X_i and on the regularly varying index ρ of the generator of the Archimedean copula considered. It should be borne in mind that assumptions of Proposition 2.1 can be easily satisfied. Indeed, in Table 1 in Charpentier and Segers (2009), various copula models with associated ρ index can be found and the von Mises condition is verified for a large class of marginal distributions F_i (see illustrations in the Supporting Information).*

The relationship between the quantile functions U_{T_i} and U_{X_i} is established in the following result. The proof of Proposition 2.2 is given in the Supporting Information.

Proposition 2.2 (Relation between U_{T_i} and U_{X_i}). *Let (X_1, \dots, X_d) be a random vector with Archimedean copula with generator ϕ . Assume that $\phi \in RV_\rho(1)$, with $\rho \in [1, +\infty]$. Let T_i be as in (1) with distribution function $F_i(\cdot|\alpha)$. Let $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$, as $n \rightarrow \infty$, and*

$$k_U(n) := n \left\{ 1 - \phi^{-1} \left[\left(1 - \left(1 - \frac{k(n)}{n} \right)^{1/(d-1)} \right) \phi(\alpha) \right] \right\}. \quad (4)$$

Therefore,

i) $k_U(n)$ is an intermediate sequence, that is, $k_U(n) \rightarrow \infty$, $k_U/n \rightarrow 0$ as $n \rightarrow \infty$.

ii) $U_{T_i} \left(\frac{n}{k} \right) = U_{X_i} \left(\frac{n}{k_U} \right)$, where U_{X_i} is the marginal quantile function, that is, $U_{X_i}(t) := F_i^{*-}(1 - 1/t)$.

2.2 Proposed estimator using an extrapolation method

Henceforth, we will focus on the case: $\gamma_i > 0$ and $\rho \in [1, +\infty)$ (in particular, this implies $\gamma^{T_i} > 0$). This choice is motivated by our applications in hydrology and in particular in real rainfall data-sets. Indeed, in these real-life applications, we can easily observe heavy tailed distributions (see, for instance, Pavlopoulos et al. (2008) and Papalexiou et al. (2013)). For the marginal distribution X_i , we therefore assume that there exists $\gamma_i > 0$ such that for all $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{U_{X_i}(tx)}{U_{X_i}(t)} = x^{\gamma_i}. \quad (5)$$

In this case, Propositions 2.1 and 2.2 yield, as $n \rightarrow \infty$,

$$x_{p_n}^i = U_{T_i} \left(\frac{1}{p_n} \right) \sim U_{X_i} \left(\frac{n}{k_U} \right) \left(\frac{k}{n p_n} \right)^{\gamma_i/\rho} = U_{X_i} \left(\frac{n}{k_U} \right) \left(\frac{k}{n p_n} \right)^{\gamma^{T_i}}, \quad (6)$$

where k_U is as in Equation (4) and $k = k(n) \rightarrow \infty$, $k(n)/n \rightarrow 0$, as $n \rightarrow \infty$.

Let (X_1, \dots, X_d) be a d -dimensional random vector with continuous distribution function F and Archimedean copula with generator ϕ . The goal is to estimate $x_{p_n}^i$ in (6) based on d -dimensional *i.i.d.* observations, (\mathbf{X}_j) , for $j = 1, \dots, n$, from F , where $p_n \rightarrow 0$, as $n \rightarrow +\infty$. Let $X_{n-[k_U],n}^i$ be the $(n - [k_U])$ -th order statistic of (X_1^i, \dots, X_n^i) . Therefore, the natural estimator of $U_{X_i} \left(\frac{n}{k_U} \right)$ is its empirical counterpart, that is, $X_{n-[k_U],n}^i$ (e.g., see de Haan and Ferreira (2006)).

From Equation (6), in order to define the estimator of $x_{p_n}^i$, it thus remains to estimate γ_i and ρ . We estimate γ_i with the Hill estimator (see Hill (1975)):

$$\hat{\gamma}_i = \frac{1}{k_1} \sum_{j=0}^{k_1-1} \log X_{n-j,n}^i - \log X_{n-k_1,n}^i, \quad (7)$$

where k_1 is an integer sequence such that $k_1(n) \rightarrow \infty$, $k_1/n \rightarrow 0$, $n \rightarrow \infty$, and that $X_{n-k_1, n}^i$ is the intermediate order statistic at level $n - k_1$. In addition, the regularly varying index ρ is estimated by taking into account the estimator of the upper tail dependence coefficient proposed by Schmidt and Stadtmüller (2006) (for details, see the Supporting Information).

Let $\hat{\gamma}^{T_i} := \frac{\hat{\gamma}_i}{\hat{\rho}}$. We can therefore estimate $x_{p_n}^i$ in (6) by

$$\hat{x}_{p_n}^i = X_{n-\lfloor k_U \rfloor, n}^i \left(\frac{k}{n p_n} \right)^{\hat{\gamma}^{T_i}}. \quad (8)$$

Remark 2.2. Notice that the proposed estimator in Equation (8) does not rely on the latent MPIT $Z := F(X_1, \dots, X_d)$, which is not directly observed. Under assumptions of Proposition 2.1, the application of the proposed extrapolation technique precludes the necessity to previously estimate Z . Indeed, the estimator $\hat{x}_{p_n}^i$ in Equation (8) is only based on quantities that can be directly estimated by using the observed d -dimensional i.i.d. sample (\mathbf{X}_j) , for $j = 1, \dots, n$. In Section 4, we provide a comparison with an empirical quantile estimation of T_i constructed by using the empirical multivariate distribution function $F_n(\mathbf{X}_j)$ (see Equation (10)).

3 Asymptotic normality

In order to prove the asymptotic normality of $\hat{x}_{p_n}^i$, we need to quantify the rate of convergence in Equation (5). We therefore assume the following second-order regularity condition.

Assumption 3.1 (2RV condition on U_{X_i}). *There exist $\tau_i < 0$ and an eventually positive or negative function A_i such that, as $t \rightarrow \infty$, $A_i(tx)/A_i(t) \rightarrow x^{\tau_i}$ for all $x > 0$ and*

$$\sup_{x>1} \left| x^{-\gamma_i} \frac{U_{X_i}(tx)}{U_{X_i}(t)} - 1 \right| = \mathcal{O}(A_i(t)), \text{ (see Condition (3.2.4) in de Haan and Ferreira (2006)).}$$

For the sake of brevity, the auxiliary results necessary to obtain Theorem 3.1 are presented in the Supporting Information. In the following, our main result is presented: the asymptotic normality for the estimator $\widehat{x}_{p_n}^i$ in Equation (8).

Theorem 3.1 (Asymptotic normality of $\widehat{x}_{p_n}^i$ in the upper tail dependence case, $\rho > 1$). *Let (X_1, \dots, X_d) be a random vector with Archimedean copula with twice differentiable generator ϕ . Assume that $\phi \in RV_\rho(1)$, with $\rho \in (1, +\infty)$. Let $i \in \{1, \dots, d\}$ and F_i be the twice differentiable distribution function of X_i . Assume that F_i verifies the von Mises condition with index $\gamma_i > 0$. Let T_i be as in (1) with distribution function $F_i(\cdot|\alpha)$. Assume:*

1. *For (X_i, X_j) , with $i \neq j$, the upper tail copula Λ_U exists, has continuous partial derivatives, and satisfies the second-order condition in Equation (10) in the Supporting Information with auxiliary function $A_\rho(\cdot)$.*
2. *U_{X_i} satisfies Assumption 3.1 with auxiliary function $A_i(\cdot)$, $\gamma_i > 0$ and $\tau_i < 0$.*
3. *$k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$, $n \rightarrow \infty$ such that Theorem 2.1 in the Supporting Information is satisfied.*
4. *$k_1 = k_1(n) \rightarrow \infty$, $k_1/n \rightarrow 0$, and $\sqrt{k_1}A_i(n/k_1) \rightarrow \lambda$, $n \rightarrow \infty$.*
5. *$k_2 = k_2(n) \rightarrow \infty$, $k_2/n \rightarrow 0$, and $\sqrt{k_2}A_\rho(n/k_2) \rightarrow 0$, $n \rightarrow \infty$.*

Let $r = \lim_{n \rightarrow +\infty} \frac{\sqrt{k_1(n)}}{\sqrt{k_2(n)}}$, $r' = \lim_{n \rightarrow +\infty} \frac{\sqrt{k_U(n)} \log(d_n)}{\sqrt{k_1(n)}}$ and $r'' = \lim_{n \rightarrow +\infty} \frac{\sqrt{k_U(n)} \log(d_n)}{\sqrt{k_2(n)}}$ with r, r' and $r'' \in [0, \infty]$.

Hence, as $n \rightarrow \infty$, if $r \leq 1$ and $\lim_{n \rightarrow +\infty} \frac{\log(d_n)}{\sqrt{k_1(n)}} = 0$, then

$$\min \left(\sqrt{k_U}, \frac{\sqrt{k_1}}{\log(d_n)} \right) \left(\frac{\widehat{x}_{p_n}^i}{x_{p_n}^i} - 1 \right) \xrightarrow{d} \begin{cases} B + r'(\Theta_1 + r\Theta_2), & r' \leq 1; \\ \frac{1}{r'}B + \Theta_1 + r\Theta_2, & r' > 1, \end{cases}$$

and, if $r > 1$ and $\lim_{n \rightarrow +\infty} \frac{\log(d_n)}{\sqrt{k_2(n)}} = 0$, then

$$\min \left(\sqrt{k_U}, \frac{\sqrt{k_2}}{\log(d_n)} \right) \left(\frac{\widehat{x}_{p_n}^i}{x_{p_n}^i} - 1 \right) \xrightarrow{d} \begin{cases} B + r''(\frac{1}{r}\Theta_1 + \Theta_2), & r'' \leq 1; \\ \frac{1}{r''}B + \frac{1}{r}\Theta_1 + \Theta_2, & r'' > 1, \end{cases}$$

where $d_n := k/(np_n)$, $B \sim N(0, \gamma_i^2)$, $\Theta_1 \sim N(\mu/\gamma_i, 1)$ with $\mu = \lambda/(1 - \tau_i)$ and $\Theta_2 \sim N(0, \sigma^2/\rho^2)$, with $\sigma^2 = \sigma_U^2 \left(\frac{\log(2)}{(2-\lambda_U)\log^2(2-\lambda_U)} \right)^2$, $\lambda_U := \Lambda(1, 1)$ the upper tail dependence coefficient and $\sigma_U^2 = \lambda_U + \left(\frac{\partial}{\partial x} \Lambda_U(1, 1) \right)^2 + \left(\frac{\partial}{\partial y} \Lambda_U(1, 1) \right)^2 + 2\lambda_U \left(\left(\frac{\partial}{\partial x} \Lambda_U(1, 1) - 1 \right) \left(\frac{\partial}{\partial y} \Lambda_U(1, 1) - 1 \right) - 1 \right)$.

The proof of Theorem 3.1 is presented in the Supporting Information.

Remark 3.1 (Asymptotic consistency of $\widehat{x}_{p_n}^i$ in the upper tail independence case, $\rho = 1$).

Notice that, if $\rho = 1$ (i.e. tail copula $\Lambda_U \equiv 0$), then the asymptotic variance σ_U^2 in Theorem 3.1 vanishes (see the Supporting Information). However, in the upper tail independence case, the consistency of the proposed estimator $\widehat{x}_{p_n}^i$ can be obtained. To be precise, if $\phi \in RV_1(1)$ and the second, third, and fourth conditions of Theorem 3.1 hold, then $\frac{\widehat{x}_{p_n}^i}{x_{p_n}^i} \xrightarrow{\mathbb{P}} 1$, for $n \rightarrow \infty$.

4 Simulation study

The aim of this section is to evaluate the performance of $\widehat{x}_{p_n}^i$ in finite-size samples. Although we restrict ourselves to a 3-dimensional case in this study, these illustrations could be adapt-

able in any dimension d . The performance of our extreme estimator $\widehat{x}_{p_n}^i$ is also compared with a pseudo-empirical estimator (denoted $\widehat{x}_{p_n}^{pseudo}$), an empirical estimator ($\widehat{x}_{p_n}^{emp}$) and a semi-parametric empirical estimator ($\widehat{x}_{p_n}^*$). The construction of these three competitor estimators is now described. In order to attain $\widehat{x}_{p_n}^{pseudo}$, it is assumed that the distribution function of T_i is known (see Lemma 2.1 in the Supporting Information). We can then sample from the random variable T_i by using the fact that $T_i \stackrel{d}{=} F_i^{-1} \{ \phi^{-1} [(1 - U^{1/(d-1)}) \phi(\alpha)] \}$, where U is a uniform random variable. Finally, the pseudo-empirical estimator $\widehat{x}_{p_n}^{pseudo}$ can be defined as the $(n - \lfloor np_n \rfloor)$ -th order statistic of the sample obtained from T_i ,

$$\widehat{x}_{p_n}^{pseudo} = T_{n - \lfloor np_n \rfloor, n}^i. \quad (9)$$

On the other hand, an empirical estimator ($\widehat{x}_{p_n}^{emp}$) can be proposed without the need for any information about T_i . To this end, we sample from the latent random variable $T_i = [X_i | F(\mathbf{X}) = \alpha]$ by using the empirical multivariate distribution function. Let (\mathbf{X}_j) , $j = 1, \dots, n$, be a d -dimensional *i.i.d.* sample of \mathbf{X} . For all $t \in \mathbb{R}^d$, the d -dimensional empirical distribution function of \mathbf{X} is defined as $F_n(t) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\mathbf{X}_j \leq t\}}$. \widetilde{T}_i is then obtained by collecting the points (\mathbf{X}_j^i) , for $j = 1, \dots, n$, such that $F_n(\mathbf{X}_j) \in [\alpha - h, \alpha + h]$ for a positive sufficiently small value h . The quantity h is adjusted to each considered model and each sample size. The competitor estimator $\widehat{x}_{p_n}^{emp}$ is given by

$$\widehat{x}_{p_n}^{emp} = \widetilde{T}_{n - \lfloor np_n \rfloor, n}^i. \quad (10)$$

Finally, from Definition 5.1. in Di Bernardino et al. (2015), a semi-parametric empirical competitor estimator is presented, denoted as $\widehat{x}_{p_n}^*$. Let $\widetilde{B}^i = X_{\lfloor s \rfloor, n}^i$ be the s -th order statistic of X_i

with $s = n \phi_{\hat{\theta}_n}^{-1}(S_i \phi_{\hat{\theta}_n}(\alpha))$ where S_i is a random variable with $Beta(1, d-1)$ distribution. Bear in mind that $\phi_{\hat{\theta}_n}$ is the semi-parametric estimator of the generator of the copula ϕ_θ obtained by considering the maximum pseudo-likelihood estimator of the parameter θ associated to ϕ_θ (e.g., see Genest et al. (1995)). The competitor semi-parametric empirical estimator is given by

$$\hat{x}_{p_n}^* = \tilde{B}_{n - \lfloor np_n \rfloor, n}^i \quad (11)$$

We now consider the following 3-dimensional distributional models:

1. Joe copula and Fréchet margins: $F_i(t) = \exp\{-t^{-\beta}\}$, $i = 1, 2, 3$, and $C_\theta(u_1, u_2, u_3) = 1 - [1 - \exp\{\log(1 - (1 - u_1)^\theta) + \log(1 - (1 - u_2)^\theta) + \log(1 - (1 - u_3)^\theta)\}]^{1/\theta}$. In this section, we take the dependence copula parameter $\theta = 3$ and the marginal parameter $\beta = 3$. Bear in mind that the assumptions of Theorem 3.1 are satisfied. Indeed, $\phi \in RV_3(1)$ and U_{X_i} satisfies Assumption 3.1 with $\gamma_i = 1/3$ and $\tau_i = -1$, for $i = 1, 2, 3$. In addition, the associated tail-logistic model is given by $\Lambda_U(x, y) = x + y - (x^3 + y^3)^{1/3}$, which satisfies the second-order condition in Equation (10) in the Supporting Information, and $\lambda_U = 0.74$.
2. Independence copula with Fréchet margins: $F_i(t) = \exp\{-t^{-\beta}\}$, $i = 1, 2, 3$, and $C_\theta(u_1, u_2, u_3) = u_1 u_2 u_3$. In this case, $\phi \in RV_1(1)$ and the upper tail copula is given by $\Lambda_U = \lambda_U = 0$. However, using Remark 3.1, the consistency of $\hat{\gamma}^{T_i}$ and $\hat{x}_{p_n}^i$ is illustrated in this section.
3. Gumbel copula with Pareto margins: $F_i(t) = 1 - (\delta_1/(t + \delta_1))^{\delta_2}$, $i = 1, 2, 3$, and $C_\theta(u_1, u_2, u_3) = \exp\left\{-\left((-\log(u_1))^\theta + (-\log(u_2))^\theta + (-\log(u_3))^\theta\right)^{1/\theta}\right\}$. In the simulation study we take the dependence copula parameter $\theta = 2$ and the marginal parameters $\delta_1 = 1$ and $\delta_2 = 2$. Bear in mind that the assumptions of Theorem 3.1 are satisfied. Indeed, $\phi \in RV_2(1)$ and

U_{X_i} verifies the 2RV condition with $\gamma_i = 1/2$ and $\tau = -1/2$, for $i = 1, 2, 3$. Furthermore, $\Lambda_U(x, y) = x + y - (x^2 + y^2)^{1/2}$ verifies the second-order condition in Equation (10) in the Supporting Information, and $\lambda_U = 0.59$.

Various sample sizes are taken and we consider $p_n = 1/n$ and $p_n = 1/2n$, the critical layer level $\alpha = 0.9$, and 500 Monte Carlo simulations. Note that specific values for auxiliary sequences of our procedure (k_1 , k_2 , and k) are chosen for each sample size as indicated in the figures. In order to choose k_1 , the estimator of γ_i is plotted against various values of k_1 . By balancing the potential estimation bias and variance, a common practice is to choose k_1 from the first stable region of the plots (see e.g., Cai et al. (2015)). Finally, in order to gain stability in the estimates, the obtained values are averaged in this region. A similar procedure is developed for the auxiliary sequence k_2 (for the estimation of ρ). The sequence k is selected by observing the stability of the final ratio $\hat{x}_{p_n}^i/x_{p_n}^i$.

Boxplots of ratio $\hat{\gamma}^{T_i}/\gamma^{T_i}$ In Figure 1, we present the boxplots of ratio $\hat{\gamma}^{T_i}/\gamma^{T_i}$ for the three distributional models considered, and for different sample sizes.

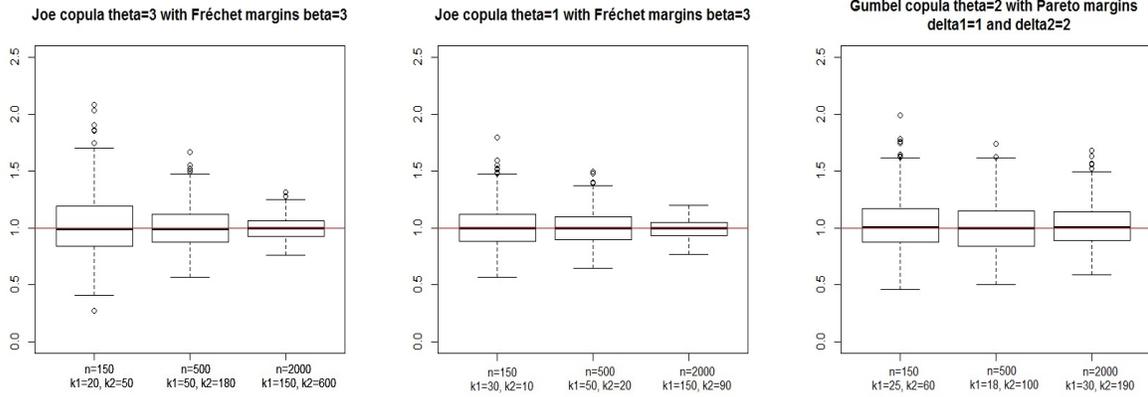


Figure 1: Boxplots of the ratio $\hat{\gamma}^{T_i} / \gamma^{T_i}$: for Joe copula with $\theta = 3$ and Fréchet margins with $\beta = 3$ (left panel); for Independence copula and Fréchet margins with $\beta = 3$ (centre panel); and for Gumbel copula with $\theta = 2$ and Pareto margins with $\delta_1 = 1$ and $\delta_2 = 2$. We consider $n = 150$, $n = 500$, and $n = 2000$, and 500 Monte Carlo simulations.

Boxplots of ratio $\hat{x}_{p_n}^i / x_{p_n}^i$ Using Remark 3.1, an illustration of the consistency of the proposed estimator is provided in the independent copula case. In Figure 2, the obtained boxplots for the ratio $\hat{x}_{p_n}^i / x_{p_n}^i$ are presented for $p_n = 1/n$ and $p_n = 1/2n$ for $n = 150$ (left panel) and $n = 1000$ (right panel). For $p_n = 1/n$, we also provide the boxplots of the ratios $\hat{x}_{p_n}^{pseudo} / x_{p_n}^i$, $\hat{x}_{p_n}^{emp} / x_{p_n}^i$, and $\hat{x}_{p_n}^* / x_{p_n}^i$, with empirical competitor estimators previously defined in Equations (9), (10) and (11), respectively.

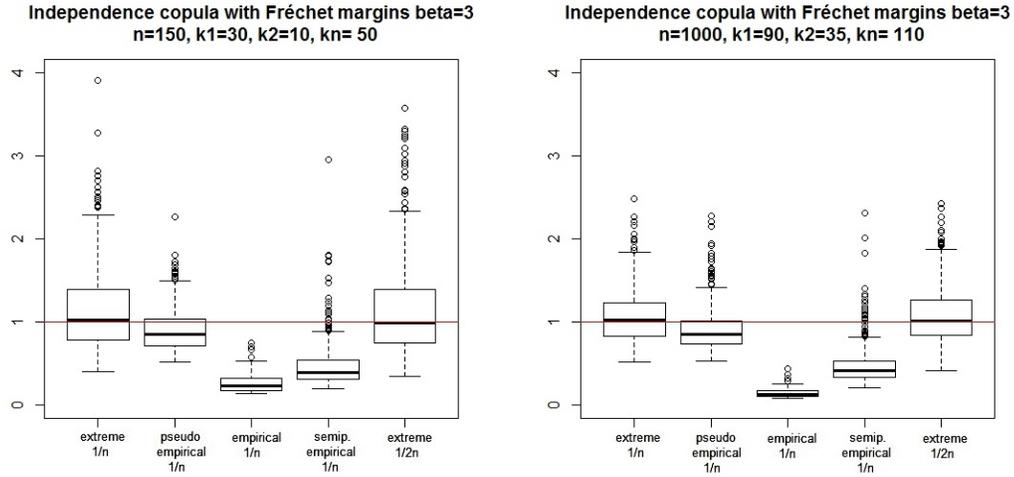


Figure 2: Independence copula and Fréchet margins with $\beta = 3$. Boxplots for the ratio $\hat{x}_{p_n}^i / x_{p_n}^i$ with $p_n = 1/n$, $p_n = 1/2n$, and $n = 150$ (left panel), $n = 1000$ (right panel). Boxplots for the competitor empirical estimators with $p_n = 1/n$ are also displayed. We consider $\alpha = 0.9$ and 500 Monte Carlo simulations.

In Figure 3, the obtained boxplots for the ratio $\hat{x}_{p_n}^i / x_{p_n}^i$ are presented for $p_n = 1/n$ and $p_n = 1/2n$, in the Joe (first row) and Gumbel copula model (second row). For $p_n = 1/n$, the comparison with the empirical competitor estimators is also provided.

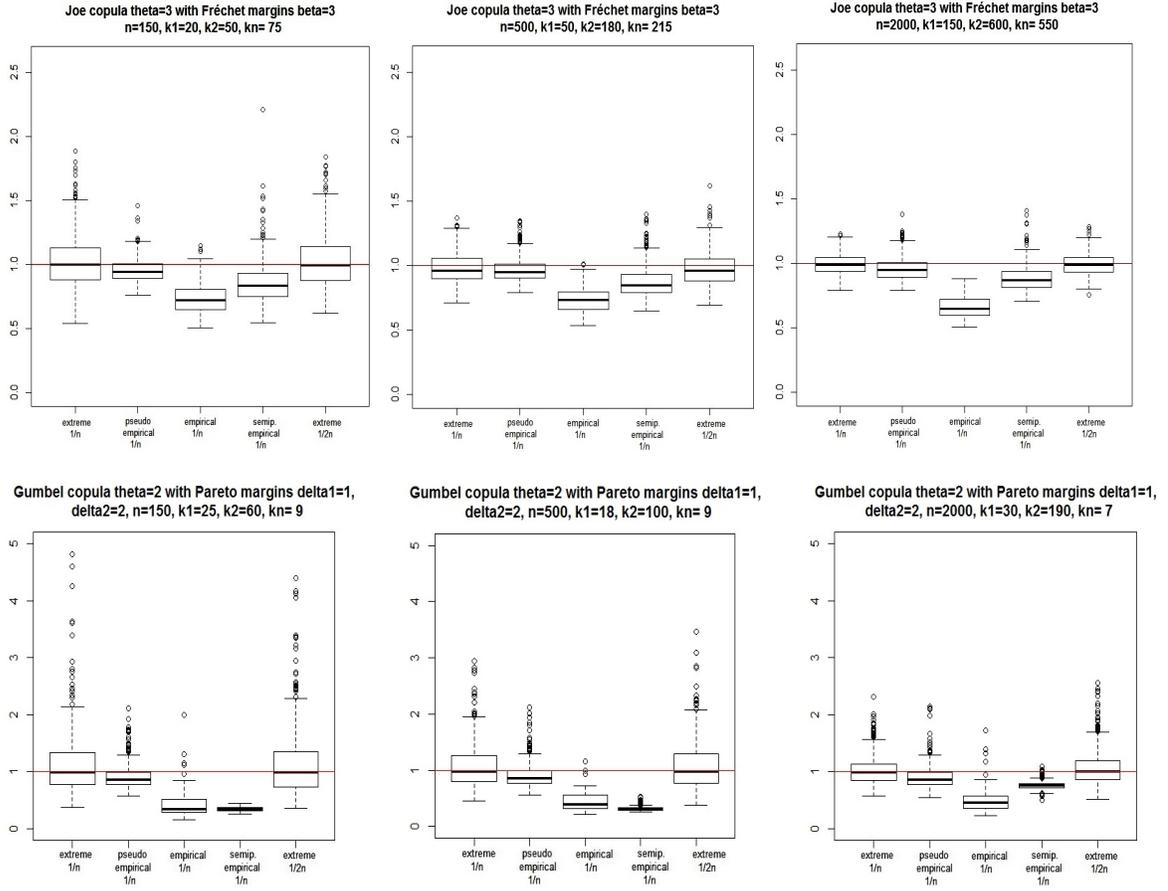


Figure 3: First row: Trivariate Joe copula with $\theta = 3$ and Fréchet margins with $\beta = 3$. Second row: Trivariate Gumbel copula with $\theta = 2$ and Pareto margins with $\delta_1 = 1$ and $\delta_2 = 2$. Boxplots for the ratio $\hat{x}_{p_n}^i/x_{p_n}^i$ with $p_n = 1/n, 1/2n$ for $n = 150, n = 500$ and $n = 2000$. Boxplots for the competitor empirical estimators with $p_n = 1/n$ are also displayed. We consider $\alpha = 0.9$ and 500 Monte Carlo simulations.

In Figures 2 and 3, it can be noticed that the empirical ($\hat{x}_{p_n}^{emp}$), pseudo-empirical ($\hat{x}_{p_n}^{pseudo}$) and semi-parametric empirical ($\hat{x}_{p_n}^*$) competitor estimators underestimate the conditional quantile

$x_{p_n}^i$ and are consistently outperformed by the proposed EVT estimator $\widehat{x}_{p_n}^i$. In addition, the empirical estimators are not applicable for $p < 1/n$. Furthermore, the behaviour of the EVT estimator $\widehat{x}_{p_n}^i$ remains stable when p_n changes from $1/n$ to $1/2n$.

Asymptotic normality Finally, the asymptotic normality in Theorem 3.1 is illustrated in Figure 4 for the Joe copula model. The Q-Q plots in Figure 4 gather the sample quantiles of $\min\left(\sqrt{k_U}, \frac{\sqrt{k_1}}{\log(d_n)}, \frac{\sqrt{k_2}}{\log(d_n)}\right)\left(\frac{\widehat{x}_{p_n}^i}{x_{p_n}^i} - 1\right)$ versus the theoretical normal quantiles for various sample sizes with $p_n = 1/n$. Since the scatterplots line up on the line in Figure 4, this indicates that the sample quantiles coincide largely with the theoretical quantiles from the asymptotic distribution. Hence, Theorem 3.1 provides an adequate approximation for finite sample sizes.

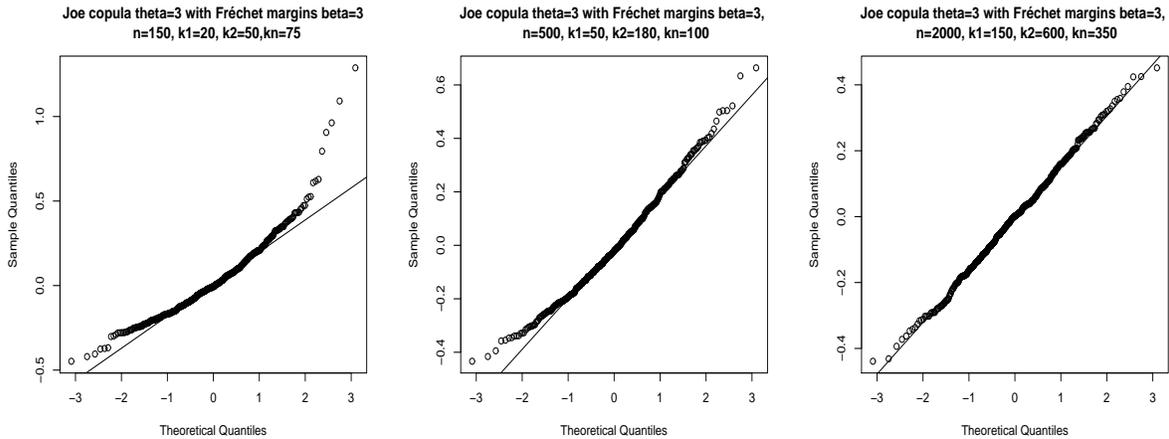


Figure 4: Joe copula with parameter $\theta = 3$ and Fréchet margins with $\beta = 3$. Q-Q plots for $\min\left(\sqrt{k_U}, \frac{\sqrt{k_1}}{\log(d_n)}, \frac{\sqrt{k_2}}{\log(d_n)}\right)\left(\frac{\widehat{x}_{p_n}^i}{x_{p_n}^i} - 1\right)$ for $p_n = 1/n$. We take $n = 150$, $n = 500$ and $n = 2000$. We consider $\alpha = 0.9$ and 500 Monte Carlo simulations.

Behaviour of ratio $\widehat{x}_{p_n}^i$ in terms of α In Figure 5, the boxplots of ratio $\widehat{x}_{p_n}^i/x_{p_n}^i$ are presented for a Joe copula with $\theta = 4$ and Fréchet margins with $\beta = 4$ by considering different values of the critical layer level α . Note that the convergence rate k_U in Proposition 2.2 (see Equation (4)) satisfies $\frac{\partial k_U(\alpha)}{\partial \alpha} \leq 0$, for fixed values of sample size n and dimension d . Therefore, as can be observed in Figure 5, the performance of the proposed estimators decreases when α increases.

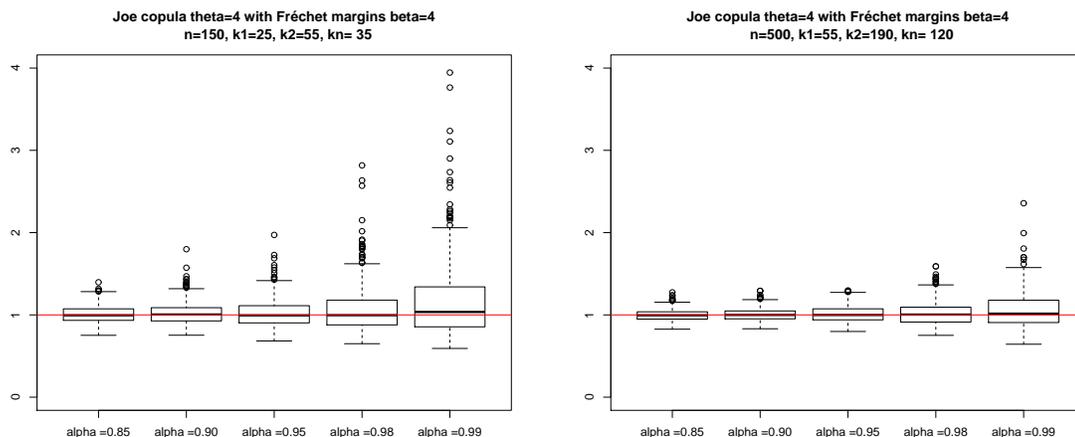


Figure 5: Trivariate Joe copula with $\theta = 4$ and Fréchet margins with $\beta = 4$. Boxplots for the ratio $\widehat{x}_{p_n}^i/x_{p_n}^i$ with $p_n = 1/n$, $n = 150$ (left panel) and $n = 500$ (right panel). Different values of α are taken, and 500 Monte Carlo simulations.

5 Multivariate and univariate extreme return levels: an illustration for hydrological data-set

In this section, we focus on estimating the risk of a flood in the *Bièvre* region in the south of Paris (France) by using both the proposed multivariate extreme return level (see Equation (2)) and the classic univariate return level (see the Introduction section).

Presentation of the hydrological data-set The data-set contains the monthly mean of the rainfall measurements recorded in 3 different stations of the *Bièvre* region, from 2003 to 2013. The unit of measurement is *mm*. The size of the data-set is $n = 125$. The localization of the 3 stations is presented in Figure 6 and the data-set is represented in Figure 7 (left panel). Let X_i denote the temporal series of the monthly mean of the rainfall measurements for station i , for $i = 1, 2, 3$. Station 1 is called *Geneste* (denoted X_1), station 2 *Loup Pendu* (X_2), and station 3 *Trou salé* (X_3). The data-set considered was provided by the *Syndicat Intercommunal pour l'Assainissement de la Valle de la Bièvre* (SIAVB, see <http://www.siavb.fr/>). Di Bernardino and Prieur (2014) discussed the plausibility of the temporal independence assumption for these 3-dimensional monthly rainfall measurements.



Figure 6: Localization of the three stations in the *Bièvre* region (in the south of Paris, France).

For the sake of completeness, a test of exchangeability is developed (e.g., see Genest et al. (2012)) for the three pairs (X_1, X_2) , (X_1, X_3) , and (X_2, X_3) : we obtain p -values of 0.511, 0.206 and 0.181, respectively. The test is performed with the function `exchTest` of the R package `copula` and suggests exchangeability for all pairs (see Figure 7, left panel). Furthermore, by using a goodness-of-fit test for various parametrical multivariate distributions, Di Bernardino and Prieur (2014) proposed for this data-set a 3-dimensional Gumbel copula with dependence parameter $\theta = 3.93$. The critical layers $\partial L(\alpha)$ of this data-set for different values of α are displayed in Figure 7 (right panel).

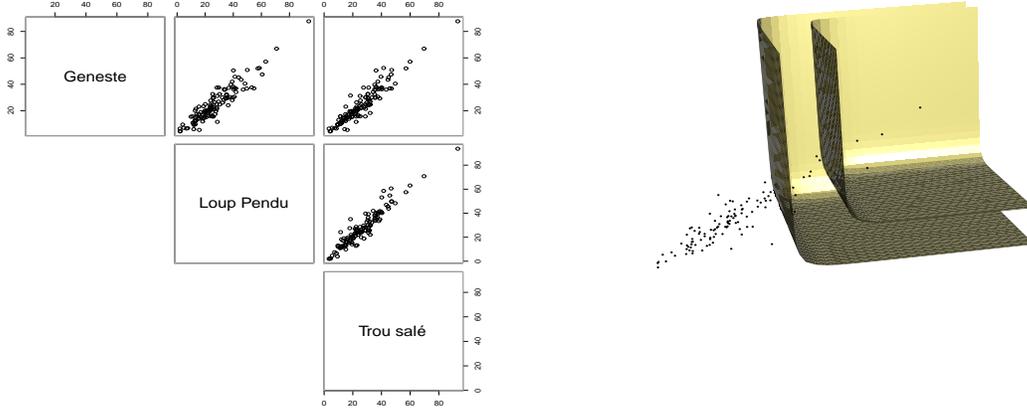


Figure 7: Left panel: Scatterplot of the 3-dimensional hydrological data-set considered. Right Panel: Associated critical layers $\partial L(\alpha)$, for $\alpha = 0.75$ and 0.9 .

Univariate versus multivariate return levels Given the temporal series X_i of the monthly mean of the rainfall measurements for station i , one can define the classic univariate return level with associated probability p as the quantity:

$$x_p^{i,univ} = U_{X_i} \left(\frac{1}{p} \right), \quad \text{for } i = 1, 2, 3,$$

where $p = 1/T$ and T is called the *return period*. As proposed by Salvadori et al. (2011), the return level associated to the three stations can be obtained by considering the vector $\vec{x}_p^{univ} := (x_p^{1,univ}, x_p^{2,univ}, x_p^{3,univ})$, that is, the aggregation of univariate quantiles.

However, \vec{x}_p^{univ} does not take into account the dependence structure between the three temporal series. As discussed in the introduction section, while the return level in the univariate setting is

usually identified without ambiguity (see, for instance, Corbella and Stretch (2012) and Salvadori et al. (2011)), in the multivariate setting, it is a troublesome task (Vandenberghe et al. (2012), Gräler et al. (2013)).

In this present paper, a possible procedure is proposed for the identification of the contribution of the margins to the global (regional) multivariate risk. As mentioned in the introduction section, the information concerning the dependence structure of the three considered rainfall measurements is integrated in order to calculate the associated multivariate return levels. We consider $x_p^i := U_{T_i} \left(\frac{1}{p} \right)$, for $i = 1, 2, 3$, where $T_i := [X_i | (X_1, X_2, X_3) \in \partial L(\alpha)]$, with $\alpha \in (0, 1)$ (see Equation (2)), and we define our multivariate return level as $\vec{x}_p := (x_p^1, x_p^2, x_p^3)$. In this case, x_p^i represents the return level associated to the i -th station conditioned to the fact that the 3-dimensional rainfall data-set belongs to the iso-surface $\partial L(\alpha)$.

Estimation procedure and obtained results In the following, the return levels $x_p^{i,univ}$ and x_p^i on the considered rainfall data are estimated, for $i = 1, 2, 3$. We consider here $\alpha = 0.9$. To estimate x_p^i , firstly we deal with the estimation of $\hat{\gamma}^{T_i}$ for each station. In Figure 8 (left panel), the Hill estimator $\hat{\gamma}_i$ is presented versus k_1 for each station; $k_1 \in [7, 27]$ is chosen since this window is the first stable region of this plot. Similarly, the stable region chosen for the considered $\hat{\rho}$ corresponds to $k_2 \in [25, 50]$ (see Figure 8, centre panel). Furthermore, the adaptive sequence $\hat{k}_U(n)$ is estimated as described in the Supporting Information. In addition, the stable region chosen for $\hat{x}_{p_n}^i$ is $k \in [38, 80]$ (see Figure 8, right panel). Finally, to gain stability, the estimations $\hat{\gamma}_i$, $\hat{\rho}$ and $\hat{x}_{p_n}^i$ are averaged in the chosen stable regions (see also Cai et al. (2015)).

The obtained extreme estimation for \vec{x}_p is presented in Table 1 for different probability levels p . In Table 1, the empirical estimator $\hat{x}_{1/n}^{i,emp}$ given in Equation (10) is also included. Unsurprisingly, $\hat{x}_{1/n}^{i,emp}$ underestimates the risk (see also the simulation study in Section 4). Furthermore, using the extreme quantile estimator proposed in Theorem 4.3.8 in de Haan and Ferreira (2006), the univariate return level $x_p^{i,univ}$ is estimated for different probability levels p (see Table 2).

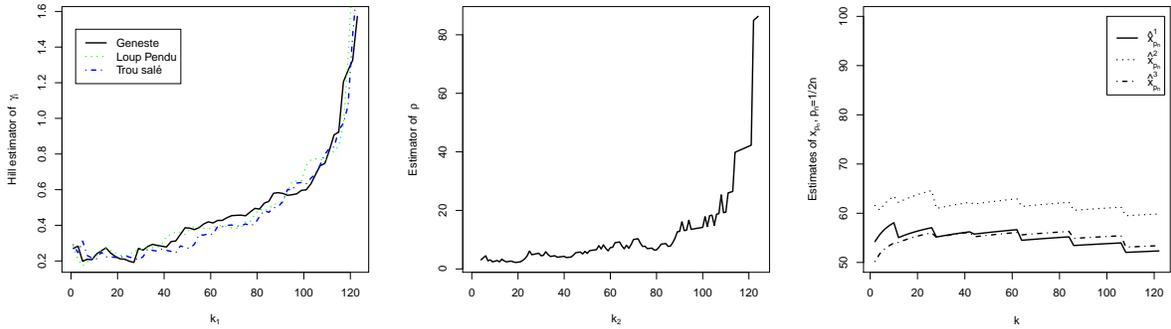


Figure 8: Hill estimators $\hat{\gamma}_i$ versus k_1 , for $i = 1, 2, 3$ (left panel). Estimator $\hat{\rho}$ based on the subvector (X_1, X_3) versus k_2 (centre panel). Estimates $\hat{x}_{p_n}^i$ against various values of the intermediate sequence k , for $i = 1, 2, 3$, and $p_n = 1/2n$ (right panel).

Note that in Table 1, $\hat{\gamma}_i > 0$ (i.e., the monthly mean of the rainfall measurements for each station i belongs to the Fréchet MDA) and $\hat{\rho} > 1$ (i.e., upper tail dependency). Values gathered in Table 1 (*resp.* in Table 2) represent the estimated multivariate return levels (*resp.* univariate return levels) in *mm* with an associated return period of around 10 years (for $p = 1/n$), 20 years (for $p = 1/2n$), 52 years (for $p = 1/5n$) and 105 years (for $p = 1/10n$).

i	Station	$\hat{\gamma}_i$	$\hat{\rho}$	$\hat{x}_{1/n}^{i,emp}$	$\hat{x}_{1/n}^i$	$\hat{x}_{1/2n}^i$	$\hat{x}_{1/5n}^i$	$\hat{x}_{1/10n}^i$
1	Geneste	0.227	4.852	52.33912	53.82129	55.59621	58.03267	59.94646
2	Loup Pendu	0.239	4.852	58.55755	59.98556	62.07613	64.95194	67.21560
3	Trou salé	0.222	4.852	41.49933	54.04171	55.78875	58.18519	60.06618

Table 1: Estimated extreme multivariate return level $\hat{x}_{p_n}^i$, for different values of p_n . Hill estimator $\hat{\gamma}_i$ is calculated by taking the average for $k_1 \in [7, 27]$; $\hat{\rho}$ is obtained by taking the average for $k_2 \in [25, 50]$; $\hat{x}_{p_n}^i$ are calculated by taking the average for $k \in [38, 80]$. The empirical estimator $\hat{x}_{1/n}^{i,emp}$ in Equation (10) is also displayed.

i	Station	$\hat{x}_{1/n}^{i,univ}$	$\hat{x}_{1/2n}^{i,univ}$	$\hat{x}_{1/5n}^{i,univ}$	$\hat{x}_{1/10n}^{i,univ}$
1	Geneste	71.63235	83.84691	103.24705	120.85247
2	Loup Pendu	81.12761	95.79985	119.34466	140.92860
3	Trou Salé	72.41996	84.51017	103.64423	120.94719

Table 2: Estimated extreme univariate return level $\hat{x}_p^{i,univ}$, for different values of p_n .

In Tables 1 and 2, a major contribution of the second station (i.e., X_2) can be observed. One can interpret that the manager of the *Bièvre* region needs to pay more attention to this station since it contributes towards producing a flood in the region to a greater degree, both in the univariate and multivariate return level cases.

Conclusions The proposed multivariate return level approach in the present paper has the advantage of using a mathematically consistent way of defining the multivariate probability of dangerous events by relying on the iso-curves $\partial L(\alpha)$. However, there is no universal choice of

an appropriate approach to all real-world problems (see also the introduction section). It is necessary to address to the problem from a probabilistic point of view and to be aware of the practical implications of the approach chosen.

It is also evident in our hydrological study, but not necessarily the case, that the more variables/information included, the smaller the design quantiles become (see multivariate and univariate return levels in Tables 1 and 2). Indeed, marginal components of the multivariate levels (i.e., \widehat{x}_p^i) are considerably lower than the corresponding univariate return levels (i.e., $\widehat{x}_p^{i,univ}$) (see Tables 1 and 2). This fact can be intuitively interpreted: the probability of an extreme event which simultaneously exceeds a return level in all margins is liable to be much lower than the probability of an event which exceeds the same level in any one of the margins considered alone. Therefore, the univariate levels $x_p^{i,univ}$ should be much higher to obtain the same small exceedance probability p . Salvadori and De Michele (2013) discuss this *dimensionality paradox* and provide a theoretical explanation. The interested reader is also referred to Salvadori et al. (2011) and Gräler et al. (2013) for analogous considerations.

In particular, in our study, the large discrepancy between the estimated \widehat{x}_p^i and $\widehat{x}_p^{i,univ}$ depends on the considered parameter setting ($\alpha = 0.9$, $p_n = \frac{1}{n}, \frac{1}{2n}, \frac{1}{5n}, \frac{1}{10n}$, with $n = 125$) and on the theoretical properties of the considered multivariate return level x_p^i . For further details about the properties of this risk measure, the interested reader is referred to Propositions 2.3-2.5 and Corollary 4.4 in Di Bernardino et al. (2015).

Furthermore, one should also be aware of the fact that our T_i -quantile approach (see Equation

(2)) is only applied to variables that are positively associated and with a focus on extremes in terms of large values. In all other cases, adaptations should be made in order to operate in the correct area of the copula (such as the *directional* multivariate return levels proposed by Torres et al. (2015)).

From a practical perspective, it is impossible to provide a general suggestion for an appropriate approach to estimate multivariate design events applicable to a vast set of design exercises. Hitherto, many applications have been based on the concept of univariate return level, since the concept of multivariate return level has a different meaning and is potentially less conservative (as can be observed by comparing Tables 1 and 2).

On the other hand, when the analyst estimates the extension of flood inundation, a joint return period approach could prove appealing. Indeed, an ensemble of equally rare scenarios (i.e. those with the same return probability p) could be used to assess the variability of the flood maps obtained due to the selection of a single design event.

Supporting Information

1 Proofs

Proof of Proposition 2.1

We first prove item *i*). Let $F_i(\cdot|\alpha)$ as in Lemma 2.1 below. Since, by assumptions, $\phi \in RV_\rho(1)$, $\phi' \in RV_{\rho-1}(1)$ and F_i verifies the von Mises condition with index γ_i (see Definition 2.1 below), then,

$$\begin{aligned} \lim_{x \uparrow x_{F_i}(\alpha)} \frac{(1 - F_i(x|\alpha))F_i''(x|\alpha)}{(F_i')^2(x|\alpha)} &= \lim_{z \uparrow 1} \frac{d-2}{d-1} \left[\left(1 - \frac{\phi(z)}{\phi(\alpha)}\right)^{-(d-1)} - 1 \right] \\ &+ \frac{\phi(\alpha)[(-\rho+1) - (\gamma_i+1)]}{(d-1)\phi'(z)(1-z)} \left[- \left(1 - \frac{\phi(z)}{\phi(\alpha)}\right)^{2-d} + 1 \right] \\ &- \left(-\frac{1}{\rho}\right) \frac{1}{d-1} [(-\rho+1) - (\gamma_i+1)]. \end{aligned}$$

Since $\phi(1) = 0$, the first summand approaches 0 when z approaches 1. We denote $C = \left(-\frac{1}{\rho}\right) \frac{[(-\rho+1) - (\gamma_i+1)]}{d-1}$. For the second summand it is verified that:

$$\lim_{z \uparrow 1} \frac{\phi(\alpha)C}{\phi(z)} \left[- \left(1 - \frac{\phi(z)}{\phi(\alpha)}\right)^{2-d} + 1 \right] = \frac{2\rho + 2\gamma_i - d\rho - d\gamma_i}{\rho(d-1)}.$$

Hence,

$$\lim_{x \uparrow x_{F_i}(\alpha)} \frac{(1 - F_i(x|\alpha))F_i''(x|\alpha)}{(F_i')^2(x|\alpha)} = - \left(\frac{\gamma_i}{\rho} + 1\right).$$

The random variable T_i therefore verifies the von Mises condition with $\gamma^{T_i} = \frac{\gamma_i}{\rho}$. Similar to the proof of item *i*), the von Mises condition for T_i when $\rho = +\infty$ is satisfied with $\gamma^{T_i} = 0$.

Therefore item *ii*) is also proved. From Theorem 1.1.8 in de Haan and Ferreira (2006), others assertions of Proposition 2.1 are shown directly. \square

Proof of Proposition 2.2

For item *i*), since $k(n)/n \rightarrow 0$, as $n \rightarrow \infty$, and $\phi^{-1}(0) = 1$, $k_U/n \rightarrow 0$ holds, as $n \rightarrow \infty$.

Furthermore, we have the following asymptotic approximation

$$k_U(n) \sim n(\phi(\alpha)(d-1))^{1/\rho} \left(\frac{k(n)}{n}\right)^{1/\rho}, \quad (12)$$

as $n \rightarrow +\infty$. From Equation (12), it holds that $k(n)/k_U(n) \rightarrow 0$, as $n \rightarrow \infty$, for $\rho \in (1, +\infty]$.

Then, $k_U(n) \rightarrow +\infty$ as $n \rightarrow +\infty$.

Since $U_{X_i}(t) = F_i^{t-}(1-1/t)$ and using Lemma 2.1 below,

$$U_{T_i} \left(\frac{n}{k(n)} \right) = U_{X_i} \left(\frac{1}{1 - \phi^{-1} \left[\left(1 - (1 - k(n)/n)^{1/(d-1)}\right) \phi(\alpha) \right]} \right).$$

Therefore,

$$U_{T_i} \left(\frac{n}{k(n)} \right) = U_{X_i} \left(\frac{n}{k_U(n)} \right), \text{ where } k_U(n) = n \left\{ 1 - \phi^{-1} \left[\left(1 - \left(1 - \frac{k(n)}{n}\right)^{1/(d-1)}\right) \phi(\alpha) \right] \right\}.$$

Therefore item *ii*) of Proposition 2.2 is also proved. \square

Proof of Theorem 3.1

Firstly, we provide a normality result for the ratio $\frac{\widehat{\gamma}_i^{T_i}}{\gamma_i^{T_i}}$. Since U_{X_i} satisfies Assumption 3.1 with

$\gamma_i > 0$ and $\tau_i < 0$, from Theorem 3.2.5 in de Haan and Ferreira (2006) and Slutsky's Theorem

(e.g., see Serfling (1980)), it is verified that

$$\sqrt{k_1} \left(\frac{\widehat{\gamma}_i}{\gamma_i} - 1 \right) \xrightarrow{d} N(\mu/\gamma_i, 1), \quad (13)$$

with $\mu = \lambda/(1-\tau_i)$ and $\lim_{n \rightarrow \infty} \sqrt{k_1(n)} A_i(n/k_1(n)) = \lambda < +\infty$. Since the distribution function of the random vector (X_1, \dots, X_d) is given by a d -dimensional Archimedean copula C_ϕ with generator ϕ , then the distribution function of every bivariate subvector (X_i, X_j) , $i \neq j$, is given by the same bivariate Archimedean copula. In addition, since under the assumptions of Theorem 3.1, conditions of Corollary 2.1 below are satisfied, by using the Delta Method, it is verified that

$$\sqrt{k_2} \left(\frac{\rho}{\hat{\rho}} - 1 \right) \xrightarrow{d} N(0, \sigma^2/\rho^2), \quad (14)$$

with σ^2 provided by

$$\sigma^2 := \sigma_U^2 \left(\frac{\log(2)}{(2 - \lambda_U) \log^2(2 - \lambda_U)} \right)^2, \quad (15)$$

with $\lambda_U := \Lambda(1, 1)$ the upper tail dependence coefficient and

$$\sigma_U^2 := \lambda_U + \left(\frac{\partial}{\partial x} \Lambda_U(1, 1) \right)^2 + \left(\frac{\partial}{\partial y} \Lambda_U(1, 1) \right)^2 + 2\lambda_U \left(\left(\frac{\partial}{\partial x} \Lambda_U(1, 1) - 1 \right) \left(\frac{\partial}{\partial y} \Lambda_U(1, 1) - 1 \right) - 1 \right).$$

We now write

$$\frac{\hat{\gamma}_i^T}{\gamma_i^T} = \frac{\hat{\gamma}_i}{\gamma_i} \times \frac{\rho}{\hat{\rho}} =: M_1 \times M_2$$

and we deal with the two factors separately.

- From (13), $M_1 = \frac{\Theta_1}{\sqrt{k_1}} + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k_1}} \right) + 1$, with $\Theta_1 \sim N(\mu/\gamma_i, 1)$.
- From (14), $M_2 = \frac{\Theta_2}{\sqrt{k_2}} + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k_2}} \right) + 1$, with $\Theta_2 \sim N(0, \sigma^2/\rho^2)$.

Hence,

$$\left(\frac{\hat{\gamma}_i^T}{\gamma_i^T} - 1 \right) = M_1 \times M_2 - 1 = \frac{\Theta_1}{\sqrt{k_1}} + \frac{\Theta_2}{\sqrt{k_2}} + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k_1}} \right) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k_2}} \right).$$

Let $r = \lim_{t \rightarrow +\infty} \frac{\sqrt{k_1(n)}}{\sqrt{k_2(n)}} \in [0, \infty]$ and $\gamma^{T_i} := \frac{\gamma_i}{\rho}$. Then, as $n \rightarrow \infty$, we get

$$\min(\sqrt{k_1}, \sqrt{k_2}) \left(\frac{\widehat{\gamma}^{T_i}}{\gamma^{T_i}} - 1 \right) \xrightarrow{d} \begin{cases} \Theta_1 + r\Theta_2, & \text{if } r \leq 1, \\ \frac{1}{r}\Theta_1 + \Theta_2, & \text{if } r > 1. \end{cases} \quad (16)$$

We now write

$$\frac{\widehat{x}_{p_n}^i}{x_{p_n}^i} = \frac{X_{n - \lfloor k_U \rfloor, n}^i}{U_{X_i} \left(\frac{n}{k_U} \right)} \times \left(\frac{k}{n p_n} \right)^{\widehat{\gamma}^{T_i} - \gamma^{T_i}} =: N_1 \times N_2.$$

From Theorem 2.1 below,

$$N_1 \xrightarrow{d} \frac{B}{\sqrt{k_U}} + 1 + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k_U}} \right), \quad \text{where } B \sim N(0, \gamma_i^2). \quad (17)$$

By using the normality for the ratio $\frac{\widehat{\gamma}^{T_i}}{\gamma^{T_i}}$ in Equation (16), we can get

$$\frac{\min(\sqrt{k_1}, \sqrt{k_2})}{\log(d_n)} \left(d_n^{\widehat{\gamma}^{T_i} - \gamma^{T_i}} - 1 \right) \xrightarrow{d} \begin{cases} \Theta_1 + r\Theta_2, & \text{if } r \leq 1, \\ \frac{1}{r}\Theta_1 + \Theta_2, & \text{if } r > 1, \end{cases}$$

where $d_n = \frac{k}{n p_n}$. The interested reader is also referred to the proof of Theorem 4.3.8 in de Haan and Ferreira (2006). Consequently,

$$N_2 \xrightarrow{d} \begin{cases} \frac{\log(d_n)}{\sqrt{k_1}} (\Theta_1 + r\Theta_2) + 1 + o_{\mathbb{P}} \left(\frac{\log(d_n)}{\sqrt{k_1}} \right), & \text{if } r \leq 1, \\ \frac{\log(d_n)}{\sqrt{k_2}} \left(\frac{1}{r}\Theta_1 + \Theta_2 \right) + 1 + o_{\mathbb{P}} \left(\frac{\log(d_n)}{\sqrt{k_2}} \right), & \text{if } r > 1. \end{cases} \quad (18)$$

Combining the asymptotic relations in (17) and (18), if $r \leq 1$, we have

$$\frac{\widehat{x}_{p_n}^i}{x_{p_n}^i} - 1 = \frac{B}{\sqrt{k_U}} + \frac{\log(d_n)}{\sqrt{k_1}} (\Theta_1 + r\Theta_2) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k_U}} \right) + o_{\mathbb{P}} \left(\frac{\log(d_n)}{\sqrt{k_1}} \right). \quad (19)$$

Similarly, if $r > 1$, then

$$\frac{\widehat{x}_{p_n}^i}{x_{p_n}^i} - 1 = \frac{B}{\sqrt{k_U}} + \frac{\log(d_n)}{\sqrt{k_2}} \left(\frac{1}{r}\Theta_1 + \Theta_2 \right) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k_U}} \right) + o_{\mathbb{P}} \left(\frac{\log(d_n)}{\sqrt{k_2}} \right). \quad (20)$$

Hence, Theorem 3.1 comes down from Equations (19) and (20). \square

2 Auxiliary results

In this section, some brief reminders and auxiliary results are described. These are only intended to outline some notation and references, and to help in the proofs developed in Section 1.

2.1 The von Mises condition and the distribution of T_i

Definition 2.1 and Lemma 2.1 introduced below are crucial in the proof of Propositions 2.1 and 2.2. In particular, Lemma 2.1 can be obtained by adapting Lemma 3.4 in Brechmann (2014) in the case of $j = 1$.

Definition 2.1 (*the von Mises condition*). *Let F be a distribution function and x^* its right endpoint. Let F' and F'' be the first and the second derivatives of F , respectively. Suppose $F''(x)$ exists and $F'(x)$ is positive for all x in some left neighborhood of x^* . The von Mises condition for F holds if*

$$\lim_{t \uparrow x^*} \frac{(1 - F(t))F''(t)}{(F'(t))^2} = -\gamma - 1.$$

Under the *von Mises condition* in Definition 2.1, the maximum domain of attraction (MDA) of the distribution function F can be determined by using the tail parameter γ (e.g., see Theorem 1.1.8 in de Haan and Ferreira (2006)).

Lemma 2.1. *Let (X_1, \dots, X_d) be a random vector which follows an Archimedean copula C_ϕ with generator ϕ . Let $F_i(x|\alpha) = \mathbb{P}[X_i \leq x | \mathbf{X} \in \partial L(\alpha)]$. Therefore, for $i = 1, \dots, d$,*

$$F_i(x|\alpha) = \begin{cases} \left(1 - \frac{\phi(F_i(x))}{\phi(\alpha)}\right)^{d-1}, & \text{if } x > Q_i(\alpha); \\ 0, & \text{if } x \leq Q_i(\alpha), \end{cases}$$

where F_i is the marginal distribution of X_i and $Q_i(\alpha)$ is the associated quantile function at level $\alpha \in (0, 1)$.

Tail index and the von Mises condition for T_i In the following, certain tail indexes for T_i are derived. The ρ indexes for the classic bivariate Archimedean copulas are collected in Table 1 in Charpentier and Segers (2009). From this table and from Proposition 2.1, Table 3 below is constructed. Table 3 (left panel) contains the tail index γ^{T_i} of T_i when X_i is in the Weibull domain (i.e., $\gamma_i < 0$), Gumbel domain (i.e., $\gamma_i = 0$) and Fréchet domain (i.e., $\gamma_i > 0$), for different values of ρ . In Table 3 (right panel), some specific models are considered.

ρ	$\gamma_i < 0$	$\gamma_i = 0$	$\gamma_i > 0$
$(1, +\infty)$	γ_i/ρ	0	γ_i/ρ
1	γ_i	0	γ_i
$+\infty$	0	0	0

Copula	$U(0, 1)$	$Exp(\lambda)$	$Par(\delta, 1)$
Gumbel	$-1/\theta$	0	$1/\delta\theta$
Ali-Mikhail-Haq	-1	0	$1/\delta$
18	0	0	0

Table 3: Left panel: The tail index γ^{T_i} when (X_1, \dots, X_d) follows an Archimedean copula with $\phi \in RV_\rho(1)$ and F_i verifies the von Mises condition with index γ_i . Right panel: The tail index γ^{T_i} for some specific models.

2.2 Estimation of the regularly varying index ρ

We now deal with the estimation of the regularly varying index ρ of the Archimedean generator ϕ . The corollary obtained below constitutes a major auxiliary result in the proof of Theorem 3.1. In this paper, we use an estimator of ρ derived by the estimator of the *upper tail dependence coefficient* proposed by Schmidt and Stadtmüller (2006). This procedure is recalled in this section.

Let G be a d -dimensional distribution function with margin distributions G_i , $i = 1, \dots, d$. If, for the subsets $I, J \in \{1, \dots, d\}$, $I \cap J = \emptyset$, the following limit exists everywhere on $\bar{\mathbb{R}}_+^d = [0, \infty]^d \setminus (\infty, \dots, \infty)$

$$\Lambda_U^{I,J}(\mathbf{x}) := \lim_{t \rightarrow \infty} P \left[X_i > G_i^{-1}(1 - x_i/t), \forall i \in I \mid X_j > G_j^{-1}(1 - x_j/t), \forall j \in J \right],$$

then the function $\Lambda_U^{I,J} : \bar{\mathbb{R}}_+^d \rightarrow \mathbb{R}$ is called an *upper tail copula* associated with G w.r. to I, J .

Let (X_i, X_j) , $i \neq j$, be a bivariate random vector with distribution functions G_i and G_j . It is said to be *upper tail dependent* if $\Lambda_U(1, 1)$ exists and

$$\lambda_U := \Lambda_U(1, 1) = \lim_{t \rightarrow 1^-} P[X_i > G_i^{-1}(t) \mid X_j > G_j^{-1}(t)] > 0.$$

Conversely, if $\lambda_U = 0$, (X_i, X_j) is called *upper tail independent*. Further, λ_U is referred to as the *upper tail dependence coefficient*. If the distribution function of a random vector (X_1, \dots, X_d) is given by a d -dimensional Archimedean copula C_ϕ with generator ϕ , then the distribution function of every bivariate subvector (X_i, X_j) is given by the bivariate Archimedean copula with the same generator. As a consequence, to estimate ρ , we focus on the bivariate subvector

(X_i, X_j) . Furthermore, under our assumption, one can prove that $\lambda_U = 2 - 2^{1/\rho}$ (e.g., see Corollary 2.1. in Di Bernardino and Rullière (2014)). Therefore, we can consider the estimator $\hat{\rho} := \frac{\log(2)}{\log(2-\hat{\lambda}_U)}$. In this paper, we use the nonparametric rank-based estimator of λ_U proposed by Schmidt and Stadtmüller (2006). Assume that $(X_i, X_j), (X_i^{(1)}, X_j^{(1)}), \dots, (X_i^{(n)}, X_j^{(n)})$, $i \neq j$, are *i.i.d.* bivariate random vectors with distribution function G having marginal distribution functions G_i and G_j . The estimator of λ_U in Schmidt and Stadtmüller (2006) is given by $\hat{\lambda}_U = \hat{\Lambda}_{U,n}(1, 1)$, where

$$\hat{\Lambda}_{U,n}(x, y) := \frac{1}{k_2} \sum_{w=1}^n \mathbf{1}_{\{R_i^{(w)} > n - k_2 x \text{ and } R_j^{(w)} > n - k_2 y\}},$$

with $k_2 = k_2(n) \rightarrow \infty$, $k_2/n \rightarrow 0$, as $n \rightarrow \infty$ and $R_i^{(w)} = \sum_{h=1}^n \mathbf{1}_{\{X_i^{(h)} \leq X_i^{(w)}\}}$ (*resp.* $R_j^{(w)} = \sum_{h=1}^n \mathbf{1}_{\{X_j^{(h)} \leq X_j^{(w)}\}}$) is the rank of $X_i^{(w)}$ in $X_i^{(1)}, \dots, X_i^{(n)}$ (*resp.* is the rank of $X_j^{(w)}$ in $X_j^{(1)}, \dots, X_j^{(n)}$), for $w = 1, \dots, n$.

Under a second-order condition for the bivariate *upper tail copula* $\Lambda_U(x, y)$ (see condition in Equation (21)), we can obtain an asymptotic normality result for the estimator $\hat{\rho}$ (see Corollary 2.1 below). The proof of Corollary 2.1 follows from Corollary 2 in Schmidt and Stadtmüller (2006) and the Delta Method technique.

Corollary 2.1 (Asymptotic normality of $\hat{\rho}$). *Let G be a bivariate distribution function of (X_i, X_j) with continuous marginal distribution functions G_i and G_j . Let C_ϕ be the copula of (X_i, X_j) with generator $\phi \in RV_\rho(1)$, with $\rho \in (1, +\infty)$. Let $k_2 = k_2(n) \rightarrow \infty$ and $k_2/n \rightarrow 0$ as $n \rightarrow \infty$. Assume that the bivariate upper tail copula $\Lambda_U(x, y)$ exists and has continuous partial derivatives. Furthermore, let $A_\rho : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be an auxiliary function such that $A_\rho(t) \rightarrow 0$ as*

$t \rightarrow \infty$ and

$$\lim_{t \rightarrow \infty} \frac{\Lambda_U(x, y) - t \bar{C}(x/t, y/t)}{A_\rho(t)} = g(x, y) < \infty, \quad (21)$$

locally uniformly for $(x, y)^2 \in \bar{\mathbb{R}}_+^2$ for some nonconstant function g , where \bar{C} represents the survival copula. Therefore, if $\sqrt{k_2} A_\rho(n/k_2) \rightarrow 0$ as $n \rightarrow \infty$, then

$$\sqrt{k_2}(\hat{\rho} - \rho) \xrightarrow{d} N(0, \sigma^2),$$

where $N(0, \sigma^2)$ is a centred normal-distributed random variable with $\sigma^2 = \sigma_U^2 \left(\frac{\log(2)}{(2-\lambda_U) \log^2(2-\lambda_U)} \right)^2$ and $\sigma_U^2 = \lambda_U + \left(\frac{\partial}{\partial x} \Lambda_U(1, 1) \right)^2 + \left(\frac{\partial}{\partial y} \Lambda_U(1, 1) \right)^2 + 2\lambda_U \left(\left(\frac{\partial}{\partial x} \Lambda_U(1, 1) - 1 \right) \left(\frac{\partial}{\partial y} \Lambda_U(1, 1) - 1 \right) - 1 \right)$.

Note that the asymptotic variance in Corollary 2.1, vanishes in the asymptotically independent case. Therefore, in the case $\Lambda_U = 0$, it is verified that $\hat{\lambda}_U \xrightarrow{\mathbb{P}} 0$ (for more details see Theorem A.1. and Corollary A.1. in Di Bernardino et al. (2013b)). Consequently, $\frac{\hat{\rho}}{\rho} \xrightarrow{\mathbb{P}} 1$.

Second-order condition for the bivariate upper tail copula $\Lambda_U(x, y)$ In Table 4, the second-order condition for the bivariate upper tail copula $\Lambda_U(x, y)$ in Equation (21) is illustrated for some classic Archimedean copula models with $\Lambda_U(x, y) = x + y - (x^\theta + y^\theta)^{1/\theta}$. We consider the Gumbel copula, Joe copula, and Copulas (12), (14), (15) and (21) in Table 4.1 in Nelsen (2006). Observe that the property in Equation (21) is not verified for Copula (2) in Table 4.1 in Nelsen (2006).

Copula	$\phi(t)$	$A_\rho(t)$
Gumbel	$(-\log(t))^\theta$	t^{-1}
Joe	$-\log(1 - (1 - t)^\theta)$	$t^{-\theta}$
(12)	$(1/t - 1)^\theta$	t^{-1}
(14)	$(t^{-1/\theta} - 1)^\theta$	t^{-1}
(15)	$(1 - t^{1/\theta})^\theta$	t^{-1}
(21)	$1 - (1 - (1 - t)^\theta)^{1/\theta}$	$t^{-\theta}$

Table 4: Bivariate Archimedean copula models with $\rho = \theta$ and $\lambda_U = 2 - 2^{1/\theta}$.

2.3 Intermediate Order Statistics

In the following, we adapt in our setting the well-known Central Limit Theorem for the intermediate order statistics. This result follows easily from Theorems 2.4.1 and 2.4.2 in de Haan and Ferreira (2006). Further details are given in Theorem 2.1 in Drees (1998). Theorem 2.1 below is crucial in the proof of our main result (see Theorem 3.1).

Theorem 2.1 (Theorem 2.1 in Drees (1998)). *Let (X_1, \dots, X_d) be a random vector with Archimedean copula C_ϕ with twice differentiable generator ϕ . Assume that $\phi \in RV_\rho(1)$, with $\rho \in [1, +\infty]$. Let $i \in \{1, \dots, d\}$. Assume that U_{X_i} satisfies Assumption 3.1 with auxiliary function $A_i(\cdot)$, $\gamma_i > 0$ and $\tau_i < 0$. Let $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$, $n \rightarrow \infty$ such that $\lim_{n \rightarrow \infty} \sqrt{k_U} A_i(n/k_U)$ exists and is finite with the sequence k_U defined by $k_U(n) := n \left\{ 1 - \phi^{-1} \left[\left(1 - \left(1 - \frac{k(n)}{n} \right)^{1/(d-1)} \right) \phi(\alpha) \right] \right\}$.*

Then, it holds that, as $n \rightarrow \infty$,

$$\sqrt{k_U(n)} \left(\frac{X_{n-\lfloor k_U \rfloor, n}^i}{U_{X_i} \left(\frac{n}{k_U(n)} \right)} - 1 \right) \xrightarrow{d} \gamma_i N(0, 1).$$

Proof of Theorem 2.1

From Proposition 2.2, it is verified that $U_{T_i} \left(\frac{n}{k} \right) = U_{X_i} \left(\frac{n}{k_U} \right)$, and $k_U(n) \rightarrow \infty$, $k_U/n \rightarrow 0$ as $n \rightarrow \infty$. Since, by assumptions, U_{X_i} satisfies Assumption 3.1 with auxiliary function $A_i(\cdot)$, $\gamma_i > 0$, $\tau_i < 0$ and $\sqrt{k_U} A_i(n/k_U) \rightarrow \lambda' < \infty$, as $n \rightarrow \infty$, then from Theorem 2.4.2 in de Haan and Ferreira (2006), the result is attained. \square

2.4 Adaptive version of the estimator $\widehat{x}_{p_n}^i$

The intermediate sequence $k_U(n)$ in Proposition 2.2 and Theorem 3.1 is an unknown sequence which depends on the generator of the considered Archimedean copula. In this section, a plug-in procedure based on the estimation of k_U is presented. This can be seen as an adaptive version of the results of Section 3. For this purpose, the notion of self-nested diagonal of a copula and the associated nonparametric estimator proposed by Di Bernardino and Rullière (2013) are recalled in the following.

Recall that the diagonal section of a d -dimensional copula C is given by $\delta_1(u) = C(u, \dots, u)$, $u \in [0, 1]$, and δ_{-1} is the inverse function of δ_1 , such that $\delta_1 \circ \delta_{-1}$ is the identity function. From Lemma 3.4 in Di Bernardino and Rullière (2013), one can write the family of self-nested diagonals of an Archimedean copula C_ϕ at each order $r \in \mathbb{R}$ as: $\delta_r(u) = \phi^{-1}(d^r \phi(u))$, for $u \in (0, 1)$, $r \in \mathbb{R}$.

Di Bernardino and Rullière (2013) introduce the following estimation of a self-nested diagonal

δ_r , by using an interpolation procedure (see also Lemma 3.6 in the aforementioned paper).

Definition 2.2 (Definition 4.2 in Di Bernardino and Rullière (2013)). *Let $\widehat{\delta}_1$ be an estimator of δ_1 , and $\widehat{\delta}_{-1}$ be an estimator of the inverse function δ_{-1} . Estimators of δ_h and δ_{-h} can be obtained for any $h \in \mathbb{N} \setminus \{0\}$ by setting*

$$\begin{cases} \widehat{\delta}_h(u) &= \widehat{\delta}_1 \circ \dots \circ \widehat{\delta}_1(u), & (h \text{ times}) \\ \widehat{\delta}_{-h}(u) &= \widehat{\delta}_{-1} \circ \dots \circ \widehat{\delta}_{-1}(u), & (h \text{ times}) \\ \widehat{\delta}_0(u) &= u. \end{cases}$$

At any order $r \in \mathbb{R}$, an estimator of $\widehat{\delta}_r$ of δ_r is

$$\widehat{\delta}_r(u) = z \left(\left(z^{-1} \circ \widehat{\delta}_h(u) \right)^{1-\eta} \left(z^{-1} \circ \widehat{\delta}_{h+1}(u) \right)^\eta \right), \text{ for } u \in [0, 1], \quad (22)$$

with $\eta = r - [r]$ and $h = [r]$ where $[r]$ denotes the integer part of r , and where z is a strictly monotone function driving the interpolation, ideally the inverse of the generator of the copula.

Several different estimators for δ_1 can be found in the literature. In particular, one can propose $\widehat{\delta}_1(u) = F_{Y,n}(u)$, where $F_{Y,n}(u)$ is the empirical distribution function of $Y := \max(U_1, U_2, \dots, U_d)$. Similarly, we consider $\widehat{\delta}_{-1}(u) = F_{Y,n}^{-1}(u)$, with $F_{Y,n}^{-1}(u)$ the empirical quantile function of Y .

Using the self-nested diagonal family δ_r , we write the sequence $k_U(n)$ as: $k_U(n) = n(1 - \delta_{r(n)}(\alpha))$, where $r(n) := \log \left(1 - \left(1 - \frac{k(n)}{n} \right)^{1/(d-1)} \right) / \log(d)$ is a negative real sequence. Therefore, using the nonparametric estimator $\widehat{\delta}_{r(n)}$ in Definition 2.2, we introduce the estimator

$$\widehat{k}_U(n) = n(1 - \widehat{\delta}_{r(n)}(\alpha)), \text{ for } \alpha \in (0, 1). \quad (23)$$

The following consistency result for $\widehat{k}_U(n)$ can now be proved.

Lemma 2.2. *Let $k_U(n)$ be the intermediate sequence defined as in Theorem 2.1. Let $\widehat{\delta}_1(u) = F_{Y,n}(u)$, with $F_{Y,n}(u)$ the empirical distribution function of $Y := \max(U_1, U_2, \dots, U_d)$ and $\widehat{\delta}_{-1}(u) = F_{Y,n}^{-1}(u)$, with $F_{Y,n}^{-1}(u)$ the empirical quantile function of Y . Let $\widehat{k}_U(n)$ be the associated estimator proposed in Equation (23) for a fixed $\alpha \in (0, 1)$ and where z is a strictly monotone function driving the interpolation. Then,*

$$\frac{\widehat{k}_U(n)}{k_U(n)} \xrightarrow{\mathbb{P}} 1, \quad \text{as } n \rightarrow \infty.$$

Proof of Lemma 2.2

Firstly, we prove that $\frac{\widehat{\delta}_h(u)}{\delta_h(u)} \xrightarrow{\mathbb{P}} 1$, for $u \in (0, 1)$ and for fixed $h \in \mathbb{Z}$, where δ_h is introduced at the beginning of this section and $\widehat{\delta}_h(u)$ is defined in Definition 2.2. Consider that $h \in \mathbb{Z}^+$. Since $\widehat{\delta}_1(u) := F_{Y,n}(u)$, where $F_{Y,n}(u)$ is the empirical distribution function of $Y := \max(U_1, U_2, \dots, U_d)$, then from Glivenko Cantelli's Theorem, it is verified that

$$\sup_{u \in [0,1]} |\widehat{\delta}_1(u) - \delta_1(u)| = \sup_{u \in [0,1]} |F_{Y,n}(u) - F_Y(u)| \xrightarrow{\mathbb{P}} 0, \quad \text{as } n \rightarrow \infty.$$

By induction, we assume that $\sup_{u \in [0,1]} |\widehat{\delta}_{m-1}(u) - \delta_{m-1}(u)| \xrightarrow{\mathbb{P}} 0$. Since C is a Lipschitz function (see Definition 6.2.6 in Nelsen (2006)), from Theorem 1 in Kasy (2015) and from the uniform convergence of $\widehat{\delta}_1(u)$, then $\sup_{u \in [0,1]} |\widehat{\delta}_m(u) - \delta_m(u)| \xrightarrow{\mathbb{P}} 0$, as $n \rightarrow \infty$. Let $h \in \mathbb{Z}^-$. We have $\widehat{\delta}_{-1}(u) := F_{Y,n}^{-1}(u)$, where $F_{Y,n}^{-1}(u)$ is the empirical quantile function of Y . From Theorem 3 in Mason (1982),

$$\sup_{u \in (0,1)} |\widehat{\delta}_{-1}(u) - \delta_{-1}(u)| = \sup_{u \in (0,1)} |F_{Y,n}^{-1}(u) - F_Y^{-1}(u)| \xrightarrow{\mathbb{P}} 0, \quad \text{as } n \rightarrow \infty.$$

By induction, we suppose that $\sup_{u \in (0,1)} |\widehat{\delta}_m(u) - \delta_m(u)| \xrightarrow{\mathbb{P}} 0$. Since C^{-1} is a uniformly conti-

nuous function in $[0, 1]$, then from Theorem 1 in Kasy (2015) and from the uniform convergence of $\widehat{\delta}_{-1}(u)$, we obtain $\sup_{u \in (0,1)} |\widehat{\delta}_{m-1}(u) - \delta_{m-1}(u)| \xrightarrow{\mathbb{P}} 0$, as $n \rightarrow \infty$. Therefore, $\frac{\widehat{\delta}_h(u)}{\delta_h(u)} \xrightarrow{\mathbb{P}} 1$, for $u \in (0, 1)$ and for fixed $h \in \mathbb{Z}$. Furthermore, by using the Slutsky's Theorem (e.g., see Serfling (1980)), one can prove that $\frac{\widehat{\delta}_r(u)}{\delta_r(u)} \xrightarrow{\mathbb{P}} 1$, $\forall u \in (0, 1)$ and $\forall r \in \mathbb{R}$ fixed. Therefore, since δ_r is also a continuous and bounded function in r , from Polya's Theorem (e.g., see Section A.1.1 in Embrechts et al. (1997)), then, for $u \in (0, 1)$, $\sup_{r \in \mathbb{R}} |\widehat{\delta}_r(u) - \delta_r(u)| \xrightarrow{\mathbb{P}} 0$, as $n \rightarrow \infty$. By using this uniform consistency we obtain the assertion of Lemma 2.2. \square

Using Lemma 2.2, it can be proved that $X_{n-\lfloor \widehat{k}_U \rfloor, n}^i$ is asymptotically as efficient as $X_{n-\lfloor k_U \rfloor, n}^i$. To be more precise, an adaptive plug-in version of Theorem 2.1 can be obtained, i.e.,

$$\sqrt{\widehat{k}_U(n)} \left(\frac{X_{n-\lfloor \widehat{k}_U \rfloor, n}^i}{U_{X_i} \left(\frac{n}{\widehat{k}_U(n)} \right)} - 1 \right) \xrightarrow{d} \gamma_i N(0, 1), \quad \text{as } n \rightarrow \infty. \quad (24)$$

Further details are given in Hall and Welsh (1985), Drees and Kaufmann (1998), and Danielsson et al. (2001). Then, an adaptive version of Theorem 3.1 for $\widehat{x}_{p_n}^i$ can also be provided. The proof is a slightly modified version of the proof of Theorem 3.1, by using the result in Equation (24) instead of Theorem 2.1. Illustrations of this plug-in estimation of $\widehat{x}_{p_n}^i$, by using \widehat{k}_U instead of k_U , can be found in Section 5.

In particular, to estimate the adaptive sequence $\widehat{k}_U(n)$ in Section 5, we consider $z(x) = \exp(-x)$. This choice is recommended in Di Bernardino and Rullière (2013) when there is positive dependence, since it is the best choice for any Gumbel copula, whatever the parameter of the copula (see Corollary 3.7 in Di Bernardino and Rullière (2013)). Another natural choice could be any

estimator of the inverse of the generator of the copula. Finally, it should be borne in mind that this function z does not change values of any δ_k , for $k \in \mathbb{Z}$. Therefore, the global shape of δ_r , as a function of $r \in \mathbb{R}$, is not heavily impacted by the choice of z . For an in-depth analysis of the weak impact of the interpolation function z in the evaluation of δ_r , the reader is referred to Section 4.3.1 in Di Bernardino and Rullière (2013).

Illustrations of estimators $\widehat{\delta}_r$ and \widehat{k}_U In Figure 9, illustrations of $\widehat{\delta}_r$ with $r \in \mathbb{R}$ are provided for two different interpolation functions. As in Di Bernardino and Rullière (2013), a 2-dimensional Gumbel copula is generated with $\theta = 3$ and sample size $n = 2000$ and $n = 7000$. We consider $z(x) = \exp(-x)$ (first row of Figure 9) and $z(x) = \exp(-x^{1/\theta})$, i.e., the inverse of the Gumbel generator copula (see second row of Figure 9) with $r = -3.5, -2.4, -1.2, 0.6, 1.2, 2.4, 3.5$. As pointed out before, it can be observed that the modification of the interpolation function z does not produce significant differences in the estimation of δ_r .

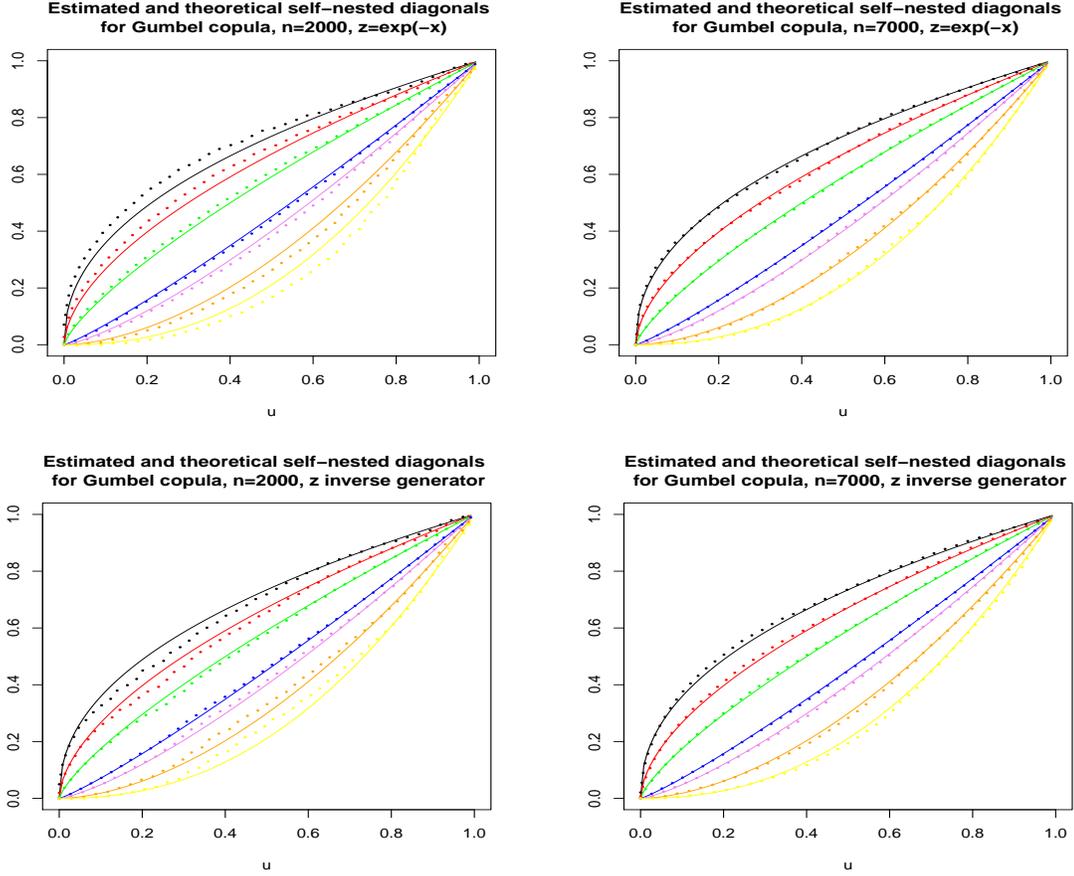


Figure 9: Gumbel copula with dependence parameter $\theta = 3$. Estimation of $\delta_r(x)$ by considering $z(x) = \exp(-x)$ (first row) and $z(x) = \exp(-x^{1/\theta})$ (second row) with $r = -3.5, -2.4, -1.2, 0.6, 1.2, 2.4, 3.5$, for $n = 2000$ (left panels) and $n = 7000$ (right panels).

Finally, an illustration of Lemma 2.2 is provided in Figure 10 where the boxplots of the ratio $\widehat{k}_U(n)/k_U(n)$ are gathered for a Joe copula $\theta = 3$ with Fréchet margins $\beta = 3$ by considering different sample sizes, with $k(n) = \sqrt{n}$ (left panel) and $k(n) = n^{0.9}$ (right panel). In this case, we choose z the inverse of the generator of the considered Joe copula.

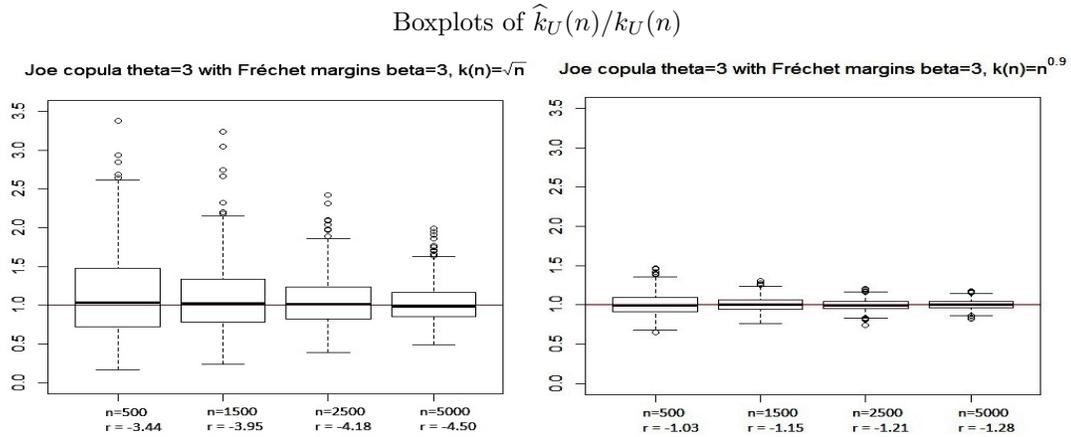


Figure 10: Joe copula with dependence parameter $\theta = 3$ and Fréchet margins with $\beta = 3$. Boxplots for the ratio $\hat{k}_U(n)/k_U(n)$ for various values of n , $\alpha = 0.9$, $k(n) = \sqrt{n}$ (left panel) and $k(n) = n^{0.9}$ (right panel). 500 Monte Carlo simulations are taken.

Acknowledgements The authors wish to thank the reviewers for their numerous and very useful comments and suggestions. The authors would like to thank Latouche A. for fruitful discussions. This work was partly supported by a grant from Junta de Andalucía (Spain) for research group (FQM- 328) and by a pre-doctoral contract (Palacios Rodríguez, F.) from the “V Plan Propio de Investigación” of the University of Seville.

References

Belzunce, F., Castaño, A., Olvera-Cervantes, A., and Suárez-Llorens, A. (2007). Quantile curves and dependence structure for bivariate distributions. *Computational Statistics & Data Analysis*, 51:5112–5129.

- Brechmann, E. C. (2014). Hierarchical kendall copulas: properties and inference. *The Canadian Journal of Statistics*, 42(1):78–108.
- Cai, J. J., Einmahl, J. H. J., de Haan, L., and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journals of the Royal Statistical Society. Series B. Statistical Methodology*, 77(2):417–442.
- Charpentier, A. and Segers, J. (2009). Tails of multivariate Archimedean copulas. *Journal of Multivariate Analysis*, 100(7):1521–1537.
- Chebana, F. and Ouarda, T. B. M. J. (2011a). Multivariate extreme value identification using depth functions. *Environmetrics*, 22:441–455.
- Chebana, F. and Ouarda, T. B. M. J. (2011b). Multivariate quantiles in hydrological frequency analysis. *Environmetrics*, 22:63–78.
- Corbella, S. and Stretch, D. D. (2012). Multivariate return periods of sea storms for coastal erosion risk assessment. *Natural Hazards and Earth System Sciences*, 12(8):2699–2708.
- Cousin, A. and Di Bernardino, E. (2013). On multivariate extensions of Value-at-Risk. *Journal of Multivariate Analysis*, 119:32 – 46.
- Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76:226–248.

- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory. An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer: New York.
- de Haan, L. and Huang, X. (1995). Large quantile estimation in a multivariate setting. *Journal of Multivariate Analysis*, 53(2):247–263.
- de Waal, D. J., van Gelder, P. H. A. J. M., and Nel, A. (2007). Estimating joint tail probabilities of river discharges through the logistic copula. *Environmetrics*, 18:621–631.
- Di Bernardino, E., Fernández-Ponce, J., Palacios-Rodríguez, F., and Rodríguez-Griñolo, M. (2015). On Multivariate Extensions of the Conditional Value-at-Risk measure. *Insurance: Mathematics and Economics*, 61:1–16.
- Di Bernardino, E., Laloë, T., Maume-Deschamps, V., and Prieur, C. (2013a). Plug-in estimation of level sets in a non-compact setting with applications in multivariable risk theory. *ESAIM: Probability and Statistics*, 17:236–256.
- Di Bernardino, E., Maume-Deschamps, V., and Prieur, C. (2013b). Estimating a bivariate tail: a copula based approach. *Journal of Multivariate Analysis*, 119:81–100.
- Di Bernardino, E. and Prieur, C. (2014). Estimation of multivariate Conditional-Tail-Expectation using Kendall’s process. *Journal of Nonparametric Statistics*, 26(2):241–267.
- Di Bernardino, E. and Rullière, D. (2013). On certain transformations of Archimedean copulas: Application to the non-parametric estimation of their generators. *Dependence Modelling*, 1:1–36.

- Di Bernardino, E. and Rullière, D. (2014). On tail dependence coefficients of transformed multivariate Archimedean copulas. *Working paper*, <https://hal.archives-ouvertes.fr/hal-00992707v1/document>.
- Drees, H. (1998). On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, 25:187–210.
- Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75:149–172.
- Durante, F. and Salvadori, G. (2010). On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21:143–161.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer. Berlin.
- Embrechts, P. and Puccetti, G. (2006). Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, 97(2):526–547.
- Fawcett, L. and Walshaw, D. (2012). Estimating return levels from serially dependent extremes. *Environmetrics*, 23:272–283.
- Genest, C., Ghoudi, K., and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.
- Genest, C., Nešlehová, J., and Quessy, J. F. (2012). Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, 64:811–834.

- Genest, C. and Rivest, L. P. (2001). On the multivariate probability integral transformation. *Statistics & Probability Letters*, 53:391–399.
- Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. E. C. (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4):1281–1296.
- Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1):331–341.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174.
- Hochrainer-Stigler, S. and Pflug, G. (2012). Risk management against extremes in a changing environment: a risk-layer approach using copulas. *Environmetrics*, 23:663–672.
- Kasy, M. (2015). Uniformity and the delta method. *Harvard University Working Paper*.
<http://scholar.harvard.edu/kasy/publications/uniformity-and-delta-method>.
- Mason, D. M. (1982). Some Characterizations of Almost Sure Bounds for Weighted Multidimensional Empirical Distributions and a Glivenko-Cantelli Theorem for Sample Quantiles. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 59:505–513.
- McNeil, A. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and l_1 -norm symmetric distributions. *The Annals of Statistics*, 37(5B):3059–3097.

- Nappo, G. and Spizzichino, F. (2009). Kendall distributions and level sets in bivariate exchangeable survival models. *Information Sciences*, 179:2878–2890.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Series in Statistics. Springer: New York.
- Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C. (2013). How extreme is extreme? an assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, 17:851–862.
- Pavlopoulos, H., Picek, J., and Jurečková, J. (2008). Heavy tailed durations of regional rainfall. *Applications of Mathematics*, 53(3):249–265.
- Prékopa, A. (2012). Multivariate Value at Risk and related topics. *Annals of Operation Research*, 193:49–69.
- Salvadori, G. and De Michele, C. (2013). Multivariate extreme value methods. In AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., and Sorooshian, S., editors, *Extremes in a Changing Climate*. Springer: Dordrecht.
- Salvadori, G., De Michele, C., and Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15:3293–3305.
- Salvadori, G., Durante, F., and Perrone, E. (2012). Semi-parametric approximation of Kendall’s distribution function and multivariate return periods. *Journal de la Société Française de Statistique*, 153(3):151–173.

- Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics. Theory and Applications*, 33(2):307–335.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- Torres, R., Lillo, R. E., and Laniado, H. (2015). A directional multivariate value at risk. *Insurance: Mathematics and Economics*, 65:111–123.
- Vandenberghe, S., van den Berg, M. J., Gräler, B., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. E. C. (2012). Joint return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences Discussions*, 9:6781–6828.