HAL
open science

A flexible and tractable class of one-factor copulas

Gildas Mazo, Stéphane Girard, Florence Forbes

To cite this version:

Gildas Mazo, Stéphane Girard, Florence Forbes. A flexible and tractable class of one-factor copulas. Statistics and Computing, 2016, 26 (5), pp.965-979. 10.1007/s11222-015-9580-7 . hal-00979147v3

HAL Id: hal-00979147

https://hal.science/hal-00979147v3

Submitted on 26 May 2015

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.

A flexible and tractable class of one-factor copulas

Gildas Mazo, Stéphane Girard and Florence Forbes

MISTIS, Inria - Laboratoire Jean Kuntzmann, France

Abstract

Copulas are a useful tool to model multivariate distributions. While there exist various families of bivariate copulas, the construction of flexible and yet tractable copulas suitable for high-dimensional applications is much more challenging. This is even more true if one is concerned with the analysis of extreme values. In this paper, we construct a class of one-factor copulas and a family of extreme-value copulas well suited for high-dimensional applications and exhibiting a good balance between tractability and flexibility. The inference for these copulas is performed by using a least-squares estimator based on dependence coefficients. The modeling capabilities of the copulas are illustrated on simulated and real datasets.

Keywords: extreme-value copula, factor model, multivariate copula, high-dimension.

1 Introduction

The modeling of random multivariate events (i.e., of dimension greater than 2) is a central problem in various scientific domains and the construction of multivariate distributions able to properly model the variables at play is challenging. The challenge is even more difficult if the data provide evidence of tail dependencies or non Gaussian behaviors. To address this problem, the concept of copulas is a useful tool as it permits to impose a dependence structure on pre-determined marginal distributions. Standard books covering this subject include [23, 32]. See also [18] for an introduction to this topic. The most common copula models used in high dimensional applications are discussed below.

The popular Archimedean copulas are tractable and allow to model a different behavior in the lower and upper tails. For instance, the Gumbel copula is upper, but not lower, tail dependent; the opposite holds for the Clayton copula. Nevertheless, the dependence structure of Archimedean copulas is severely restricted because they are exchangeable, implying that all the pairs of variables have the same distribution. More details about these copulas can be found in the above mentioned books.

Nested Archimedean copulas are a class of hierarchical copulas generalizing the class of Archimedean copulas. They allow to introduce asymmetry in the dependence structure but only between groups of variables. This hierarchical structure is not desirable when no prior knowledge on the random phenomenon under consideration is available. Furthermore, constraints on the parameters

restrict the tractability of these copulas. These copulas first appeared in [23] Section 4.2.

The class of elliptical copulas arises from the class of elliptical distributions. These copulas are interesting in many ways but they are tail symmetric, meaning that the lower tail dependence coefficient is equal to the upper tail dependence coefficient (these coefficients are defined in Section 2.2). This may not be the case in applications. See, e.g., [31] Section 5 or [16] for an introduction to these copulas.

Pair copula constructions and Vines are flexible copula models based on the decomposition of the density as a product of conditional bivariate copulas. However, these models are difficult to handle. Furthermore, the conditional bivariate copulas are typically assumed not to depend on the conditioning variables. This so called *simplifying assumption* can be misleading, as remarked in [2]. Pair-copula constructions first appeared in [23] Section 4.5. See also [4, 5, 28] for theoretical developments and [1] for a practical introduction to modeling with Vines.

As mentioned above, most copula models are either tractable or flexible, but rarely both. In this paper, we propose a tractable and yet flexible class of one-factor copulas well suited for high-dimensional applications. This class is nonparametric, and, therefore, encompasses many distributions with different features. Unlike elliptical copulas, the members of this class allow for tail asymmetry. Furthermore, we derive the associated extreme-value copulas, and, therefore, the analysis of extreme values can be carried out with the presented models. Finally, we show how to perform theoretically well-grounded, and practically fast and accurate, inference of these copulas, thanks to the ability of calculating explicitly the dependence coefficients.

The remaining of this paper is organized as follows. Section 2 presents the proposed class of one-factor copulas, Section 3 deals with inference, and, in Section 4, the proposed copulas are applied to simulated and real datasets. The proofs are postponed to the Appendix.

2 A tractable and flexible class of one-factor copulas

The class of copulas proposed in this paper, referred to as the FDG class (see Section 2.2 for an explanation of this acronym), can be embedded in the framework of one-factor models. We therefore introduce the later in Section 2.1. The construction and properties of FDG copulas are given in Section 2.2. Parametric examples are proposed in Section 2.3. The extreme-value copulas associated to the FDG class are derived in Section 2.4.

2.1 One-factor copulas

By definition, the coordinates of a random vector distributed according to a one-factor copula [27] are independent given a latent factor. More specifically, let $U_0, U_1, \ldots, U_d$ ($d \geq 2$) be standard uniform random variables such that the coordinates of $(U_1, \ldots, U_d)$ are conditionally independent given U_0. The variable U_0 plays the role of a latent, or unobserved, factor. Let us write C_{0i} the distribution of (U_0, U_i) and $C_{i|0}(\cdot|u_0)$ the conditional distribution of U_i given

$U_0 = u_0$ for $i = 1, \ldots, d$. It is easy to see that the distribution of $(U_1, \ldots, U_d)$, called a one-factor copula, is given by

$$C(u_1, \ldots, u_d) = \int_0^1 C_{1|0}(u_1|u_0) \ldots C_{d|0}(u_d|u_0)\, du_0. \tag{1}$$

The copulas C_{0i} are called the *linking copulas* because they link the factor U_0 to the variables of interest U_i. The one-factor model has many advantages to address high dimensional problems. We recall and briefly discuss them below.

Nonexchangeability. The one-factor model is nonexchangeable. Recall that a copula C is said to be *exchangeable* if $C(u_1, \ldots, u_d) = C(u_{\pi(1)}, \ldots, u_{\pi(d)})$ for any permutation π of $(1, \ldots, d)$. This means in particular that all the bivariate marginal distributions are equal to each other. For example, Archimedean copulas are exchangeable copulas. Needless to say, this assumption may be too strong in practice.

Parsimony. The one-factor model is parsimonious. Indeed, only d linking copulas are involved in the construction of the one-factor model, and since they are typically governed by one or two parameters, the number of parameters in total increases only linearly with the dimension. Parsimony is more and more desirable as the dimension increases.

Random generation. The conditional independence property of the one factor model allows to easily generate data $(U_1, \ldots, U_d)$ from this copula.

1 Generate $U_0, V_1, \ldots, V_d$ independent standard uniform random variables.

2 For $i = 1, \ldots, d$, put $U_i = C_{i|0}^{-1}(V_i|U_0)$ where $C_{i|0}^{-1}(\ .\ |U_0)$ denotes the inverse of $v \mapsto C_{i|0}(v|U_0)$.

Dependence properties of the one-factor model have been studied in [27]. The authors investigated how positive dependence properties of the linking copulas extend to the bivariate margins

$$C_{ij}(u_i, u_j) := C(1, \ldots, 1, u_i, 1, \ldots, 1, u_j, 1, \ldots, 1).$$

These properties included positive quadrant dependence, increasing in the concordance ordering, stochastic increasing, and tail dependence. For details about these dependence concepts, see [23] Section 2. The copulas proposed in this paper, presented in Section 2.2 and Section 2.4, admit simple expressions and therefore the properties mentioned above can be made more specific.

2.2 Construction and properties of FDG copulas

The class of FDG copulas is constructed by choosing appropriate linking copulas for the one-factor copula model (1). The class of linking copulas which served to build the FDG copulas is referred to as the *Durante* class [11] of bivariate copulas, which can also be viewed as part of the framework of [3]. The Durante class consists of the copulas C of the form

$$C(u, v) = \min(u, v) f(\max(u, v)), \tag{2}$$

where $f : [0,1] \to [0,1]$, called the *generator* of C, is a differentiable and increasing function such that $f(1) = 1$ and $t \mapsto f(t)/t$ is decreasing. The FDG acronym thus stands for "one-Factor copula with Durante Generators". The advantages of taking Durante linking copulas are twofold: the integral (1) can be calculated and the resulting multivariate copula is nonparametric.

Theorem 1. *Let C be defined by* (1) *and assume that C_{0i} belongs to the Durante class* (2) *with given generator f_i. Then*

$$C(u_1,\dots,u_d) = u_{(1)} \left[\left(\prod_{j=2}^{d} u_{(j)} \right) \int_{u_{(d)}}^{1} \prod_{j=1}^{d} f_j'(x)dx + f_{(1)}(u_{(2)}) \left(\prod_{j=2}^{d} f_{(j)}(u_{(j)}) \right) \right. \tag{3}$$
$$\left. + \sum_{k=3}^{d} \left(\prod_{j=2}^{k-1} u_{(j)} \right) \left(\prod_{j=k}^{d} f_{(j)}(u_{(j)}) \right) \int_{u_{(k-1)}}^{u_{(k)}} \prod_{j=1}^{k-1} f_{(j)}'(x)dx \right],$$

where $u_{(i)} := u_{\sigma(i)}$, $f_{(i)} := f_{\sigma(i)}$ and σ is the permutation of $(1,\dots,d)$ such that $u_{\sigma(1)} \le \cdots \le u_{\sigma(d)}$.

In expression (3), we use the convention that the $\sum_{k=3}^{d}$ is zero when $d = 2$. The particularity of the copula expression (3) is that it depends on the generators through their reordering underlain by the permutation σ. For instance, with $d = 3$ and $u_1 < u_3 < u_2$ we have $u_{(1)} = u_1$, $u_{(2)} = u_3$, $u_{(3)} = u_2$, $\sigma = \{1,3,2\}$ and $f_{(1)} = f_{\sigma(1)} = f_1$, $f_{(2)} = f_{\sigma(2)} = f_3$, $f_{(3)} = f_{\sigma(3)} = f_2$. This feature gives its flexibility to the model. Observe also that $C(u_1,\dots,u_d)$ writes as $u_{(1)}$ multiplied by a functional of $u_{(2)},\dots,u_{(d)}$, form that is similar to (2). Although the expression of a FDG copula has the merit to be explicit, it is rather cumbersome. Hence, we shall continue its analysis through the prism of its bivariate margins.

Proposition 1. *Let C_{ij} be a bivariate margin of the FDG copula* (3)*. Then C_{ij} belongs to the Durante class* (2) *with generator*

$$f_{ij}(t) = f_i(t)f_j(t) + t\int_t^1 f_i'(x)f_j'(x)dx.$$

In other words,

$$C_{ij}(u_i,u_j) = C_{f_{ij}}(u_i,u_j) = \min(u_i,u_j)f_{ij}(\max(u_i,u_j)).$$

In view of Proposition 1, the FDG copula can be regarded as a multivariate generalization of the Durante class of bivariate copulas. In fact, such a generalization was already proposed in the literature [13]:

$$C_f(u_1,\dots,u_d) = u_{(1)} \prod_{i=2}^{d} f(u_{(i)}),$$

where f is a generator in the usual sense of the Durante class of bivariate copulas. Nonetheless, since there is only one generator to determine the whole copula in arbitrary dimension, this generalization lacks flexibility to be used

in applications. This issue is overcome by the FDG copula. To illustrate this further, its pairwise dependence coefficients are given next. Note that, since the bivariate margins of the FDG copula belong to the Durante class of bivariate copulas, a more detailed account of their properties can be found in the original paper [11]. Recall that the Spearman's rho ρ, the Kendall's tau τ, the lower $\lambda^{(L)}$ and upper $\lambda^{(U)}$ tail dependence coefficients of a general bivariate copula C are respectively given by

$$\rho = 12\int_{[0,1]^2} C(u,v)dudv - 3, \quad \tau = 4\int_{[0,1]^2} C(u,v)dC(u,v) - 1,$$
$$\lambda^{(L)} = \lim_{u\downarrow 0}\frac{C(u,u)}{u}, \text{ and} \qquad \lambda^{(U)} = \lim_{u\uparrow 1}\frac{1-2u+C(u,u)}{1-u}.$$

In the case where C belongs to the Durante class with generator f, these coefficients are respectively given by

$$\rho = 12\int_0^1 x^2 f(x)dx - 3, \quad \tau = 4\int_0^1 xf(x)^2 dx - 1,$$
$$\lambda^{(L)} = f(0), \text{ and} \qquad \lambda^{(U)} = 1 - f'(1).$$

Hence, to get the dependence coefficients of the FDG bivariate margins, it is enough to apply the above formulas and Proposition 1. The obtained coefficient expressions are given in Proposition 2 below.

Proposition 2. *The Spearman's rho, the Kendall's tau, the lower and upper tail dependence coefficients of the FDG bivariate margins C_{ij} are respectively given by*

$$\begin{aligned}
\rho_{ij} &= 12\int_0^1 x^2 f_i(x)f_j(x)dx + 3\int_0^1 x^4 f_i'(x)f_j'(x)dx - 3,\\
\tau_{ij} &= 4\int_0^1 x\left(f_i(x)f_j(x) + x\int_x^1 f_i'(t)f_j'(t)dt\right)^2 dx - 1,\\
\lambda_{ij}^{(L)} &= \lambda_i^{(L)}\lambda_j^{(L)} \text{ and}\\
\lambda_{ij}^{(U)} &= \lambda_i^{(U)}\lambda_j^{(U)},
\end{aligned}$$

where $\lambda_i^{(L)} := f_i(0)$, $\lambda_i^{(U)} := 1 - f_i'(1)$, $i = 1,\ldots,d$, are the lower and upper tail dependence coefficients of the bivariate linking copulas respectively.

2.3 Examples of parametric families

Four examples of families indexed by a real parameter for the generators $f_1,\ldots,f_d$ are given below.

Example 1 (Cuadras-Augé generators)**.** *In* (3)*, let*

$$f_i(t) = t^{1-\theta_i}, \quad \theta_i \in [0,1]. \tag{4}$$

A copula belonging to the Durante class with generator (4) *gives rise to the well known Cuadras-Augé copula with parameter θ_i [7]. By Proposition 1, the*

generator for the bivariate margin C_{ij} of the FDG copula is given by

$$f_{ij}(t) = \begin{cases} t^{2-\theta_i-\theta_j}\left(1 - \frac{(1-\theta_i)(1-\theta_j)}{1-\theta_i-\theta_j}\right) + t\frac{(1-\theta_i)(1-\theta_j)}{1-\theta_i-\theta_j} & \text{if } \theta_i + \theta_j \neq 1 \\ t(1-(1-\theta)\theta\log t) & \text{if } \theta := \theta_j = 1-\theta_i. \end{cases}$$

The Spearman's rho, the lower and upper tail dependence coefficients are respectively given by

$$\rho_{ij} = \frac{3\theta_i\theta_j}{5-\theta_i-\theta_j},\ \lambda_{ij}^{(L)} = 0 \text{ and } \lambda_{ij}^{(U)} = \theta_i\theta_j.$$

The Kendall's tau is given by

$$\tau_{ij} = \begin{cases} \frac{\theta_i\theta_j(\theta_i\theta_j+6-2(\theta_i+\theta_j))}{(\theta_i+\theta_j)^2-8(\theta_i+\theta_j)+15} & \text{if } \theta_i + \theta_j \neq 1 \\ \frac{\theta(\theta-1)(\theta^2-\theta-4)}{8} & \text{if } \theta := \theta_i = 1-\theta_j. \end{cases}$$

Example 2 (Fréchet generators)**.** *In* (3)*, let*

$$f_i(t) = (1-\theta_i)t + \theta_i, \quad \theta_i \in [0,1]. \tag{5}$$

A copula belonging to the Durante class with generator (5) *gives rise to the well known Fréchet copula with parameter θ_i [17]. By Proposition 1, the generator for the bivariate margin C_{ij} of the FDG copula is given by*

$$f_{ij}(t) = (1-\theta_i\theta_j)\,t + \theta_i\theta_j.$$

By noting that f_{ij} is of the form (5) *with parameter $\theta_i\theta_j$, one can see that the bivariate margins of the FDG copula based on Fréchet generators are still Fréchet copulas. The Spearman's rho, the lower and upper tail dependence coefficients are respectively given by*

$$\rho_{ij} = \lambda_{ij}^{(L)} = \lambda_{ij}^{(U)} = \theta_i\theta_j,$$

the Kendall's tau is given by

$$\tau_{ij} = \frac{\theta_i\theta_j(\theta_i\theta_j+2)}{3}.$$

Example 3 (Durante-sinus generators)**.** *In* (3)*, let*

$$f_i(t) = \frac{\sin(\theta_i t)}{\sin(\theta_i)}, \quad \theta_i \in (0, \pi/2]. \tag{6}$$

This generator was proposed in [11]. By Proposition 1, the generator for the bivariate margin C_{ij} of the FDG copula is given by

$$f_{ij}(t) = \frac{\sin(\theta_i t)\sin(\theta_j t)}{\sin(\theta_i)\sin(\theta_j)} + \frac{t\theta_i\theta_j}{2(\theta_j^2-\theta_i^2)\sin(\theta_i)\sin(\theta_j)} \times$$

$$\Big\{(\theta_i+\theta_j)\left[\sin\left((\theta_i-\theta_j)t\right) + \sin\left(\theta_j-\theta_i\right)\right]$$

$$+ (\theta_j-\theta_i)\left[\sin(\theta_i+\theta_j) - \sin\left((\theta_i+\theta_j)t\right)\right]\Big\} \text{ if } \theta_i \neq \theta_j, \text{ and}$$

$$f_{ij}(t) = \frac{4\sin(t\theta)^2 + t\theta\left(2(1-t)\theta + \sin(2\theta) - \sin(2t\theta)\right)}{4\sin(\theta)^2} \text{ if } \theta_i = \theta_j =: \theta.$$

The Spearman's rho, the lower and upper tail dependence coefficients are respectively given by

$$\rho_{ij} = 12(\sin\theta_i \sin\theta_j)^{-1} \int_0^1 x^2 \sin(\theta_i x)\sin(\theta_j x) + \frac{1}{4}\theta_i\theta_j x^4 \cos(\theta_i x)\cos(\theta_j x)dx - 3, \tag{7}$$

$$\lambda_{ij}^{(L)} = 0 \text{ and } \lambda_{ij}^{(U)} = \left(1 - \frac{\theta_i}{\tan(\theta_i)}\right)\left(1 - \frac{\theta_j}{\tan(\theta_j)}\right). \tag{8}$$

Example 4 (Durante-exponential generators)**.** *In* (3)*, let*

$$f_i(t) = \exp\left(\frac{t^{\theta_i} - 1}{\theta_i}\right), \quad \theta_i > 0 \tag{9}$$

This generator was proposed in [11]. By Proposition 1, the generator for the bivariate margin C_{ij} of the FDG copula is given by

$$f_{ij}(t) = \exp\left(\frac{t^{\theta_i} - 1}{\theta_i} + \frac{t^{\theta_j} - 1}{\theta_j}\right) + t\int_t^1 \exp\left(\frac{x^{\theta_i} - 1}{\theta_i} + \frac{x^{\theta_j} - 1}{\theta_j}\right) x^{\theta_i+\theta_j-2}dx.$$

The Spearman's rho, the lower and upper tail dependence coefficients are respectively given by

$$\rho_{ij} = 12\int_0^1 \exp\left(\frac{x^{\theta_i} - 1}{\theta_i} + \frac{x^{\theta_j} - 1}{\theta_j}\right)\left(x^2 + \frac{1}{4}x^{2+\theta_i+\theta_j}\right)dx - 3,$$

$$\lambda_{ij}^{(L)} = \exp\left(-\frac{1}{\theta_i} - \frac{1}{\theta_j}\right), \text{ and } \lambda_{ij}^{(U)} = 0.$$

Remark 1. *The calculation of the integral in* (7) *with $\theta_i = \theta_j = \pi/2$ shows that for the FDG copula with Durante-sinus generators, the Spearman's rho is such that*

$$0 \le \rho_{ij} \le \frac{3\pi^4 - 100\pi^2 + 840}{40\pi^2} \simeq 0.37.$$

The Spearman's rho values for all the other models in the examples above spread the entire interval $[0, 1]$.

The four above examples allow to get all possible types of tail dependencies, as shown in Table 1. The Cuadras-Augé and Durante-sinus families allow for upper but no lower tail dependence, the Durante-exponential family allows for lower but no upper tail dependence, and the Fréchet family allows for both. In the Fréchet case, furthermore, the lower and upper tail dependence coefficients are equal: this is called tail symmetry, a property of elliptical copulas.

2.4 Extreme-value attractors associated to FDG copulas

Extreme-value copulas are theoretically well-grounded copulas to perform a statistical analysis of extreme values such as maxima of random samples. Recall that a copula $C_\#$ is an extreme-value copula if there exists a copula $\tilde{C}$ such that

$$C_\#(u_1, \ldots, u_d) = \lim_{n\uparrow\infty} \widetilde{C}^n(u_1^{1/n}, \ldots, u_d^{1/n}), \quad (u_1, \ldots, u_d) \in [0, 1]^d, \tag{10}$$

family of generators	$\lambda_{ij}^{(L)}$	$\lambda_{ij}^{(U)}$
Cuadras-Augé	0	$\theta_i\theta_j$
Fréchet	$\theta_i\theta_j$	$\theta_i\theta_j$
Durante-sinus	0	$(1-\frac{\theta_i}{\tan\theta_i})(1-\frac{\theta_j}{\tan\theta_j})$
Durante-exponential	$\exp(-\frac{1}{\theta_i}-\frac{1}{\theta_j})$	0

Table 1: Lower $\lambda_{ij}^{(L)}$ and upper $\lambda_{ij}^{(U)}$ tail dependence coefficients for the four families presented in Section 2.3.

see, e.g. [21]. The extreme-value copula $C_\#$ is called the *attractor* of $\widetilde{C}$ and $\widetilde{C}$ is said to belong to the *domain of attraction* of $C_\#$. The class of extreme-value copulas corresponds exactly to the class of max-stable copulas, that is, the copulas $C_\#$ such that

$$C_\#^n(u_1^{1/n},\ldots,u_d^{1/n}) = C_\#(u_1,\ldots,u_d), \quad n\geq 1,\ (u_1,\ldots,u_d)\in[0,1]^d.$$

The upper tail dependence coefficient of a (bivariate) extreme-value copula $C_\#$ has the particular form

$$\lambda^{(U)} = 2+\log C_\#(e^{-1},e^{-1}). \tag{11}$$

This coefficient is a natural dependence coefficient for extreme-value copulas because of the following representation on the diagonal of the unit square:

$$C_\#(u,u) = u^{2-\lambda}, \tag{12}$$

where $\lambda := \lambda^{(U)}$. If $\lambda = 0$ then $C_\#(u,u) = \Pi(u,u) = u^2$, where Π stands for the independence copula. If $\lambda = 1$ then $C_\#(u,u) = M(u,u) = \min(u,u) = u$, where M stands for the Fréchet-Hoeffding upper bound for copulas, that is, the case of perfect dependence. In the case of extreme-value copulas, this interpolation between Π and M allows to interpret λ as a coefficient that measures general dependence, not only dependence in the tails. In order to emphasize this interpretation, λ will be referred to as the *extremal dependence coefficient* of an extreme-value copula. See [6] for more about extreme-value statistics, and, see, e.g. [21] for an account about extreme-value copulas.

In the case of FDG copulas, the limit (10) can be calculated. This leads to a new family of extreme-value copulas, referred to as the EV-FDG family. The bivariate margins $C_{\#,ij}$ of this new family are Cuadras-Augé copulas. These results are specified in Theorem 2 and Proposition 3 below.

Theorem 2. *Assume that the generators f_i of the FDG copula are twice continuously differentiable on $[0,1]$. Then, the attractor $C_\#$ of the FDG copula exists and is given by*

$$C_\#(u_1,\ldots,u_d) = \prod_{i=1}^d u_{(i)}^{\chi_i}, \tag{13}$$

where

$$\chi_i = \left(\prod_{j=1}^{i-1}(1-\lambda_{(j)})\right)\lambda_{(i)} + 1 - \lambda_{(i)},$$

with the convention that $\prod_{j=1}^{0}(1-\lambda_{(j)}) = 1$ *and where* $\lambda_i = 1 - f_i'(1)$. *As in* (3), $u_{(i)} = u_{\sigma(i)}$ *and* $f'_{(i)}(1) = f'_{\sigma(i)}(1)$ *where* σ *is the permutation of* $(1, \ldots, d)$ *such that* $u_{(1)} \leq \cdots \leq u_{(d)}$.

Proposition 3. *Let* $C_{\#,ij}$ *be a bivariate margin of the EV-FDG copula* (13). *Then* $C_{\#,ij}$ *is a Cuadras-Augé copula with parameter (and therefore extremal dependence coefficient)* $\lambda_i\lambda_j$. *In other words,*

$$C_{\#,ij}(u_i, u_j) = \min(u_i, u_j)\max(u_i, u_j)^{1-\lambda_i\lambda_j}. \tag{14}$$

Remark 2. *In view of Table 1, the FDG copulas with Cuadras-Augé and Fréchet generators both lead to the same EV-FDG copula.*

Multivariate generalizations of the bivariate Cuadras-Augé copula were already proposed in the literature [14, 29], but they are less flexible than EV-FDG. Indeed, let us consider the following exchangeable copula proposed in [29],

$$A(u_1, \ldots, u_d) = u_{(1)} \prod_{i=2}^{d} u_{(i)}^{a_i},$$

where $(a_1 = 1, a_2, a_3, \ldots, a_d)$ is a d-monotone sequence of real numbers, that is, a sequence which satisfies $\nabla^{j-1}a_k \geq 0$, $k = 1, \ldots, d$, $j = 1, \ldots, d-k+1$ where $\nabla^j a_k = \sum_{i=0}^{j}(-1)^i\binom{j}{i}a_{k+i}$, $j, k \geq 1$ and $\nabla^0 a_k = a_k$. In particular, the bivariate margins are Cuadras-Augé copulas

$$A_{ij}(u_i, u_j) = \min(u_i, u_j)\max(u_i, u_j)^{a_2}$$

with the same parameter $1 - a_2$. This means that all of them exhibit the same statistical behavior. For instance, all the upper tail dependence coefficients are equal and are given by $1 - a_2$. This is far too restrictive for most applications. Then let us consider the generalization proposed in [14],

$$B(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i^{1-\sum_{j=1, j\neq i}^{d}\lambda_{ij}} \prod_{i<j} \min(u_i, u_j)^{\lambda_{ij}},$$

where $\lambda_{ij} \in [0, 1], \lambda_{ij} = \lambda_{ji}$ and

$$\sum_{j=1,\ldots,d;\, j\neq i} \lambda_{ij} \leq 1, \quad i = 1, \ldots, d. \tag{15}$$

The bivariate margins B_{ij} are Cuadras-Augé copulas

$$B_{ij}(u_i, u_j) = \min(u_i, u_j)\max(u_i, u_j)^{1-\lambda_{ij}}$$

with parameters λ_{ij}. Unlike the copula A, the tail dependence coefficients can take distinct values. Unfortunately, the constraints (15) are quite restrictive, as it was already stressed by the original authors in [14]. The EV-FDG class achieves greater flexibility than these competitors. In particular, one can obtain different bivariate marginal distributions with *no* conditions on the parameters.

3 Parametric inference

Let $(X_1, \ldots, X_d)$ be a random vector following a distribution F with continuous margins $F_1, \ldots, F_d$. Suppose that its copula, C, is a FDG copula defined by (3). Denote by $(X_1^{(k)}, \ldots, X_d^{(k)})$, $k = 1, \ldots, n$ independent and identically distributed observations obtained from F. Suppose that all the generators f_i of the FDG copula belong to the same parametric family $\{f_\theta, \theta \in \Theta \subset \mathbb{R}\}$, that is, there exists $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0d}) \in \Theta^d$ such that $f_{\theta_{0i}} = f_i$. The generators f_i are regarded as functions defined over the product space $[0,1] \times \Theta$ and we write $f_i(t) = f(t, \theta_i)$ for all t in $[0,1]$. The nonparametric inference problem has turned into a parametric one where the parameter vector $\boldsymbol{\theta}_0 \in \Theta^d$ has to be estimated.

In order to estimate the parameters of the FDG and EV-FDG copulas, a least-squares estimator based on dependence coefficients is adopted. This estimation strategy was considered in various articles, see e.g. [14, 18–20, 24, 30, 33]. Its construction is recalled below. Let us choose a type of dependence coefficient (Spearman's rho, Kendall's tau, tail dependence coefficient, etc) and let us denote by $r(\theta_i, \theta_j)$ the chosen dependence coefficient between the variables X_i and X_j. Suppose that the map r is continuous and symmetric in its arguments. Let $p = d(d-1)/2$ be the number of variable pairs (X_i, X_j), $i < j$. Denote by $\mathbf{r}$ be the p-variate map defined on Θ^d such that $\mathbf{r}(\theta_1, \ldots, \theta_d) = (r(\theta_1, \theta_2), \ldots, r(\theta_{d-1}, \theta_d))$. The least-squares estimator based on dependence coefficients is defined as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta^d}{\arg\min} \left\| \hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\theta}) \right\|^2, \tag{16}$$

where the quantity $\hat{\mathbf{r}} = (\hat{r}_{1,2}, \ldots, \hat{r}_{d-1,d})$ is an empirical estimator of $\mathbf{r}(\boldsymbol{\theta_0})$. To be more specific, $r(\theta_i, \theta_j)$ may be the Spearman's rho (4) of (X_i, X_j) and

$$\hat{r}_{i,j} = \frac{\sum_{k=1}^n \left(\hat{U}_i^{(k)} - \overline{\hat{U}}_i\right)\left(\hat{U}_j^{(k)} - \overline{\hat{U}}_j\right)}{\left[\sum_{k=1}^n \left(\hat{U}_i^{(k)} - \overline{\hat{U}}_i\right)^2 \sum_{k=1}^n \left(\hat{U}_j^{(k)} - \overline{\hat{U}}_j\right)^2\right]^{1/2}}, \tag{17}$$

where $\overline{\hat{U}}_i = \sum_{k=1}^n U_i^{(k)}/n$ and $\hat{U}_i^{(k)} = \sum_{l=1}^n \mathbf{1}(X_i^{(l)} \leq X_i^{(k)})/(n+1)$. In the case where $r(\theta_i, \theta_j)$ is the Kendall's tau (4) of (X_i, X_j), then

$$\hat{r}_{i,j} = \binom{n}{2}^{-1} \sum_{k<l} \operatorname{sign}\left((X_i^{(k)} - X_i^{(l)})(X_j^{(k)} - X_j^{(l)})\right), \tag{18}$$

where $\operatorname{sign}(x) = 1$ if $x > 0$, -1 if $x < 0$ and 0 if $x = 0$. Equation (12) suggests that the extremal dependence coefficient can also be used to estimate the parameters of an extreme-value copula. If the margins F_i are known, various empirical estimators of the extremal dependence exist. For instance, assuming that C is an extreme-value copulas, [15] introduced

$$\hat{r}_{i,j} = 3 - \frac{1}{1 - \sum_{k=1}^n \max(U_i^{(k)}, U_j^{(k)})/n}, \tag{19}$$

where $U_i^{(k)} = F_i(X_i^{(k)})$.

In the context of non-absolutely continuous copulas with respect to the Lebesgue measure – as FDG copulas –, asymptotic properties of (16) were derived in [30], see the proposition below.

Proposition 4. *Suppose that the following assumptions hold.*

(A1) $\sqrt{n}(\hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\theta_0})) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$ *as* $n \to \infty$ *for some symmetric and positive definite matrix* $\boldsymbol{\Sigma}$.

(A2) The map $\mathbf{r}$ *is a twice continuously differentiable homeomorphism from* Θ^d *to its image* $\mathbf{r}(\Theta^d)$.

(A3) The Jacobian matrix $\mathbf{J}$ *of the map* $\mathbf{r}$ *at* $\boldsymbol{\theta_0}$ *is of full rank.*

Then, as $n \to \infty$*, the estimator* $\hat{\boldsymbol{\theta}}$ *defined in* (16) *is unique with probability tending to one, consistent for* $\boldsymbol{\theta_0}$*, and asymptotically normal*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Xi}), \quad \textit{where } \boldsymbol{\Xi} = \left(\mathbf{J}^T\mathbf{J}\right)^{-1} \mathbf{J}^T\boldsymbol{\Sigma}\mathbf{J} \left(\mathbf{J}^T\mathbf{J}\right)^{-1}.$$

There exist various situations where (A1) holds, as, for instance, in the context of Spearman's rho (estimator (17)) and Kendall's tau dependence coefficients (estimator (18)), see [22] for a proof. In the context of extreme-value copulas with known margins, (A1) also holds for the extremal dependence coefficient (estimator (19)), see [30] for a proof.

The remaining of this section is devoted to show that, for certain (EV-)FDG copulas, we are able to show that the assumptions of Proposition 4 hold.

Lemma 1. *(i) Define the univariate function* $r_{\theta_j}(\theta_i) := r(\theta_i, \theta_j)$ *and assume that it is a twice continuously differentiable homeomorphism. Let* $r_{1,2}, \ldots, r_{d-1,d}$ *be* p *elements of* $\mathbf{r}(\Theta^d)$*. Define* $s_{i,j}(\theta) := r_\theta^{-1}(r_{i,j})$ *for* $\theta \in \Theta$*. Then, the function* $s_{1,3} \circ s_{1,2} \circ s_{2,3}$ *has at least one fix point, that is, the equation*

$$s_{1,3} \circ s_{1,2} \circ s_{2,3}(\theta) = \theta, \quad \theta \in \Theta \tag{20}$$

has at least one solution.

(ii) If, moreover, the function $s_{1,3} \circ s_{1,2} \circ s_{2,3}$ *has exactly one fix point, that is, equation* (20) *has exactly one solution, then assumptions (A2) and (A3) of Proposition 4 hold.*

Remarking that the extremal dependence coefficients of the EV-FDG bivariate margins write $\lambda_{i,j}(\boldsymbol{\theta}) = \lambda(\theta_i)\lambda(\theta_j)$, with $\lambda(\theta) = 1 - \frac{\partial f(t,\theta)}{\partial t}\Big|_{t=1}$, $\theta \in \Theta$, allows to apply Lemma 1 and therefore to satisfy the assumptions of Proposition 4. The next theorem thus provides some situations where the least-squares estimator is unique with probability tending to one, consistent, and asymptotically Gaussian.

Theorem 3. *(i) Assume that* $(X_1, \ldots, X_d)$ *has EV-FDG* (13) *as copula and let* $r(\theta_i, \theta_j)$ *be the extremal dependence coefficient of* (X_i, X_j)*. Consider one of the following cases:*

- *The generators are Cuadras-Augé* (4) *and* $\Theta = (0, 1)$.
- *The generators are Fréchet* (5) *and* $\Theta = (0, 1)$.
- *The generators are Durante-sinus* (6) *and* $\Theta = (0, \pi/2)$.

Then, assumptions (A1), (A2) and (A3) of Proposition 4 hold with $\hat{\mathbf{r}}$ *being as in* (19).

(ii) *Assume that the copula of* $(X_1, \ldots, X_d)$ *belongs to the class of FDG copulas given in Example 2 and suppose that* $r(\theta_i, \theta_j)$ *is Spearman's rho coefficient of* (X_i, X_j) *and* $\hat{\mathbf{r}}$ *is as in* (17). *Then, assumptions (A1), (A2) and (A3) of Proposition 4 hold.*

4 Applications to simulated and real datasets

The modeling of data with (EV-)FDG copulas is illustrated through numerical experiments in Section 4.1 and a real dataset application in Section 4.2. In the numerical experiments, we first provide empirical evidence that the proposed FDG copulas are well suited for high-dimensional applications. In the real dataset application, critical levels of potentially dangerous hydrological events are estimated. Throughout this section, the four copulas of Example 1–4, are respectively referred to as FDG-CA, FDG-F, FDG-sinus, and FDG-exponential.

The minimization of the loss function (16) was carried out with standard gradient descent algorithms whose implementations can be found in the function `optim` from the R software [34]. In principle, several runs with different starting points should be tested to ensure that the global minimizer is reached. Also, formal statistical procedures can be performed to test wether the given minimizer is indeed the global minimizer; see [8, 9, 35]. We found that a single run was enough to find what appeared to be the global optimum. Thus, the loss functions one encounters when dealing with FDG copulas seem to be easy to minimize in practice. All the experiments were conducted using the `FDG` package that we developed and which is freely available on the CRAN archive at `http://cran.r-project.org/web/packages/FDG`.

4.1 Numerical experiments

Our first goal is to investigate the numerical behavior of the estimation of FDG copulas as the sample size varies from $n = 10$ to $n = 500$ with fixed $d = 9$ (as in the real dataset application). To this aim, 200 datasets are simulated from each of the copulas FDG-CA, FDG-F, FDG-sinus, and FDG-exponential. The coordinates of the parameter vector were chosen to be regularly spaced within $[0.3, 0.9], [0.3, 0.9], [1, 1.55]$ and $[3, 20]$ for FDG-CA, FDG-F, FDG-sinus, and FDG-exponential respectively. For each replication and each model, the samples are simulated from the copula using the principle described in Section 2. The model parameters are estimated by the least squares estimator (16) combined with Spearman's rho dependence coefficient as in (17). The accuracy of the method is assessed using the mean absolute error and the relative mean absolute error, respectively defined as

$$\mathrm{MAE}_r = \frac{1}{p} \sum_{i<j} |r_{i,j} - r(\hat{\theta}_i, \hat{\theta}_j)| \text{ and } \mathrm{RMAE} = \frac{1}{d} \sum_{i=1}^{d} \frac{|\hat{\theta}_i - \theta_{0i}|}{\theta_{0i}},$$

computed and averaged over the replications. It appears on Figure 1 that the MAE_ρs are very stable whatever the sample size. FDG-CA, FDG-F and FDG-sinus models also provide good results in terms of RMAE provided that the

sample size is larger than $n = 30$. At the opposite, it seems that the estimation of FDG-exponential model requires large sample sizes.

The second experiment examines the scalability of the FDG copulas when the dimension increases from $d = 10$ to $d = 50$ for a fixed sample size $n = 500$. Similarly to the previous experiment, the coordinates of the parameter vector were chosen to be regularly spaced within $[0.3, 0.9]$, $[0.3, 0.9]$, $[1, 1.55]$ and $[3, 20]$ for FDG-CA, FDG-F, FDG-sinus, and FDG-exponential respectively. It appears on Figure 2 that the MAE_ρs and RMAEs are very stable whatever the dimension. The inference for these models seems not to be sensitive to the dimension. FDG-exponential has a larger RMAE, but it is still below 17%, and its MAE_ρ is as good as that of the other models.

Figure 3 displays the associated computed times for one sample on a 8 GiB memory and 3.20 GHz processor computer. The simulation time increases linearly with the dimension whereas the estimation time increases exponentially. Simulating all the models even in high dimension is instantaneous because of the conditional independence property seen in (1). Less than two minutes are necessary to fit all the models. In particular, simulating or fitting FDG-F is instantaneous. The computational costs for performing the inference of FDG-exponential and FDG-sinus are larger because their dependence coefficient expressions, given in Example 3 and 4, involve integrals which have to be computed numerically. To summarize, FDG copulas seem to scale up well.

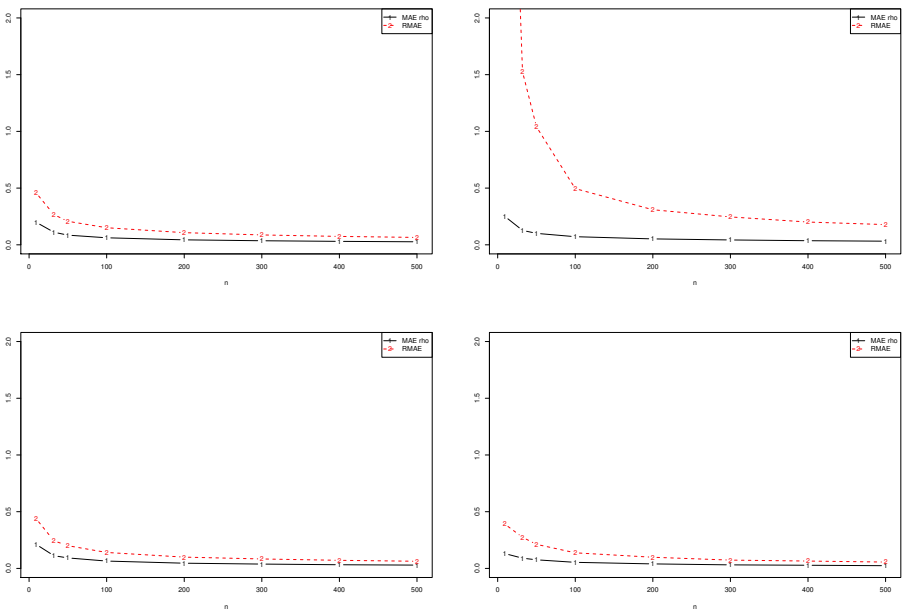

Figure 1: Fitting errors as a function of the sample size n for a dimension $d = 9$. Continuous line (1): MAE_ρ, Dashed line (2): RMAE. Four copulas were tested (FDG-CA: upper left, FDG-exponential: upper right, FDG-F: lower left, FDG-Sinus: lower right).

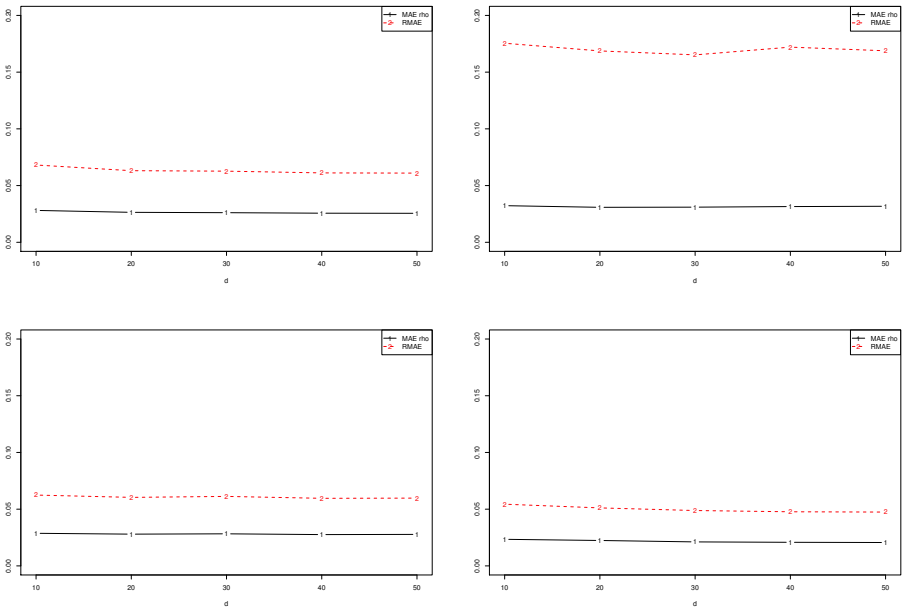

Figure 2: Fitting errors as a function of the dimension d for a sample size $n = 500$. Continous line (1): MAE_ρ, Dashed line (2): RMAE. Four copulas were tested (FDG-CA: upper left, FDG-exponential: upper right, FDG-F: lower left, FDG-Sinus: lower right).

4.2 Application to a hydrological dataset

4.2.1 Data and Context

The dataset consists of $n = 32$ observations $(X_1^{(k)}, \ldots, X_d^{(k)})$, $k = 1, \ldots, n$, of annual maxima river flow rates located at $d = 9$ sites across south-east France between 1969 and 2007 (some records are missing). Let us denote by F the distribution with continuous margins $F_1, \ldots, F_d$ of the random vector $(X_1, \ldots, X_d)$ whose realizations provide the observed dataset. The location of the sites are shown in Figure 4. The number of variable pairs is $p = 36$. Due to the heterogeneous dispersion of the sites, the span of positive dependence is almost maximum; for instance, Spearman's rho dependence coefficients range from about 0 to 0.9.

In hydrology, it is of interest to get information about the statistical distribution of a potentially dangerous event, such as $\{F_1(X_1) > q, \ldots, F_d(X_d) > q\}$, or, equivalently, $\{\min(F_1(X_1), \ldots, F_d(X_d)) > q\}$, where q is the *critical level* associated to that event. The *return period* T is defined as

$$T = \frac{1}{1 - M(q)}, \quad \text{where } M(q) = P(\min(F_1(X_1), \ldots, F_d(X_d)) \leq q). \tag{21}$$

For instance, a return period of $T = 30$ years and a critical level of $q = 0.7$ means that each X_i exceeds its quantile of order 70% once every 30 years in average. A common question in the study of extreme events is the following. Given a return period T, how dangerous is the corresponding event? In other words,

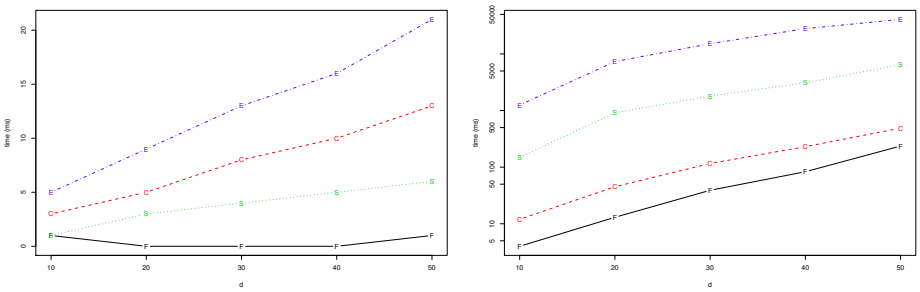

Figure 3: Computation times (in milliseconds) as a function of the dimension d for a sample of size $n = 500$. Left panel: Simulation of the copula, right panel: Estimation of the parameters. Four copulas were tested (C: FDG-CA, F: FDG-F, S: FDG-Sinus, E: FDG-exponential).

what is the associated critical level q? The answer is obtained by inverting (21)

$$q = M^{-1}\left(1 - \frac{1}{T}\right). \tag{22}$$

Thus, the answer q is the quantile of order $1 - 1/T$ of the distribution M. This quantile can be estimated empirically from the data and parametrically by fitting a model to the data.

Potentially dangerous events happen with the co-occurrence of extremely high flow rates at several locations. Thus, it is clear that the models to describe this dataset should be upper tail dependent. Hence, good candidates are the copulas of Example 1–3, referred to as FDG-CA, FDG-F, and FDG-sinus respectively, and all the extreme-value copulas. However, as it was shown in Remark 1, Spearman's rho of FDG-sinus cannot take values greater than 0.37. Hence, this copula is removed from the candidate models. So the considered models are FDG-CA, FDG-F, and their extreme-value attractor EV-FDG-CAF given in (13) (recall that FDG-CA and FDG-F lead to the same extreme-value copula). Two other popular copula models, the Gumbel and Student copulas, are also fitted to the data. The Gumbel copula is famous among hydrologists [36] and the Student copula is well known in risk management [31]. They serve as a benchmark for our models. A factor structure is assumed for the Student copula, that is, its (i, j)-th element $(i \neq j)$ of its correlation matrix writes $\theta_i\theta_j$, where $\theta_1, \ldots, \theta_d$ belong to $[-1, 1]$. Recall that a Gumbel copula is an extreme-value copula. More details about the Gumbel and the Student copula can be found respectively in, e.g., [23, 32] and [10].

4.2.2 Method and Results

A practically convenient approach dictated the estimation of the copula parameters. For each copula model, the dependence coefficients with the simplest mathematical forms were chosen to build the loss function (16). In other words, the parameters of FDG-F, FDG-CA and EV-FDG-CAF were estimated with Spearman's rho as in (17). The parameters of the Gumbel and Student copulas were estimated with Kendall's tau as in (18). Finally, the degree of freedom of

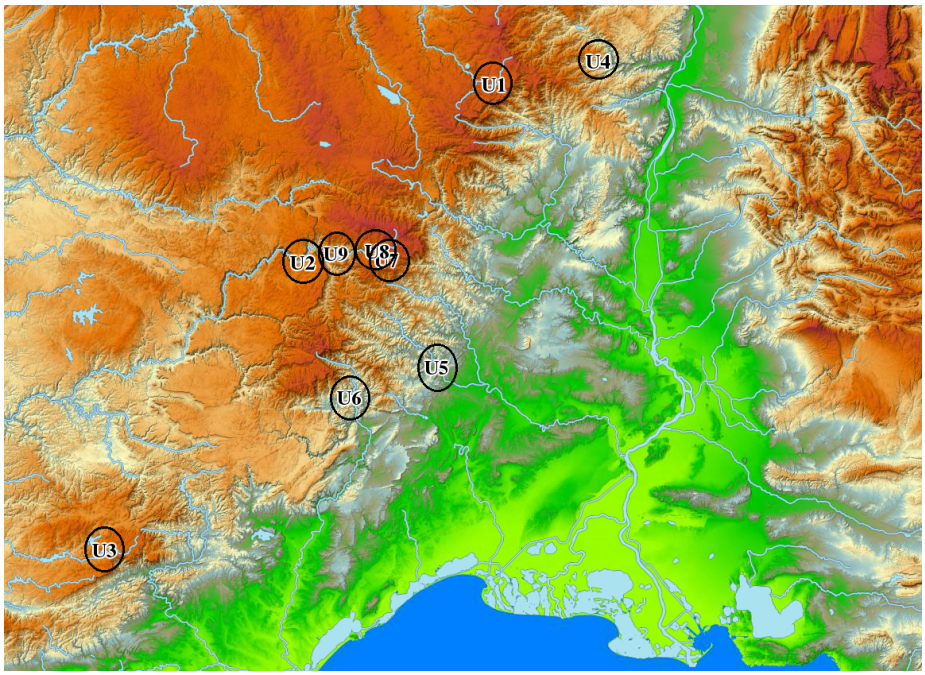

Figure 4: Location of the 9 sites for the flow rate dataset. The sea in dark blue at the bottom (south) is the Mediterranean sea. The rivers are shown in light blue. The river flowing from north to south in the green area is the Rhône. Green indicates low altitude, and orange high altitude. The map of this figure was drawn with Géoportail (`www.geoportail.gouv.fr`).

the Student copula was estimated by maximizing its likelihood but with all the parameters of the correlation matrix held fixed. This approach improves the speed, tractability, and chances of success of the minimization procedure.

The fit of the tested copulas was assessed by comparing the pairwise dependence coefficients and the critical levels.

Pairwise dependence coefficients. The mean absolute error (MAE), defined as

$$MAE_r = \frac{1}{p} \sum_{i<j} \left| \hat{r}_{i,j} - r(\hat{\theta}_i, \hat{\theta}_j) \right|$$

was computed for Spearman's rho (MAE_ρ) and Kendall's tau (MAE_τ) dependence coefficients. They are reported in Table 2. The Gumbel copula has the largest errors (more than 0.17) and does not seem to fit the data well. This was expected, because this model has only one parameter to account for a $d = 9$ dimensional phenomenon. All the remaining errors are smaller. Thus, according to these criteria, the Gumbel copula is not appropriate.

Critical levels. The critical levels obtained from the empirical data and the models were calculated by making use of (22). In statistical terms, this amounts to compare the quantiles of the distribution M under the empirical data and

	FDG-F	FDG-CA	EV-FDG-CAF	Gumbel	Student
MAE_ρ	0.12	0.12	0.12	0.22	0.12
MAE_τ	0.12	0.11	0.11	0.17	0.10

Table 2: Mean absolute errors for Spearman's rho (MAE_ρ) and Kendall's tau (MAE_τ) dependence coefficients of the models.

under the different models. The results are presented in Figure 5, where the independence copula $C(u_1, \ldots, u_d) = u_1 \ldots u_d$ was added to emphasize the need for a joint model on such a dataset. The Gumbel model is confirmed to perform poorly. FDG-F, FDG-CA and EV-FDG-CAF seem to fit the data quite well. In particular, FDG-F and FDG-CA are as close as the Student copula to the empirical curve.

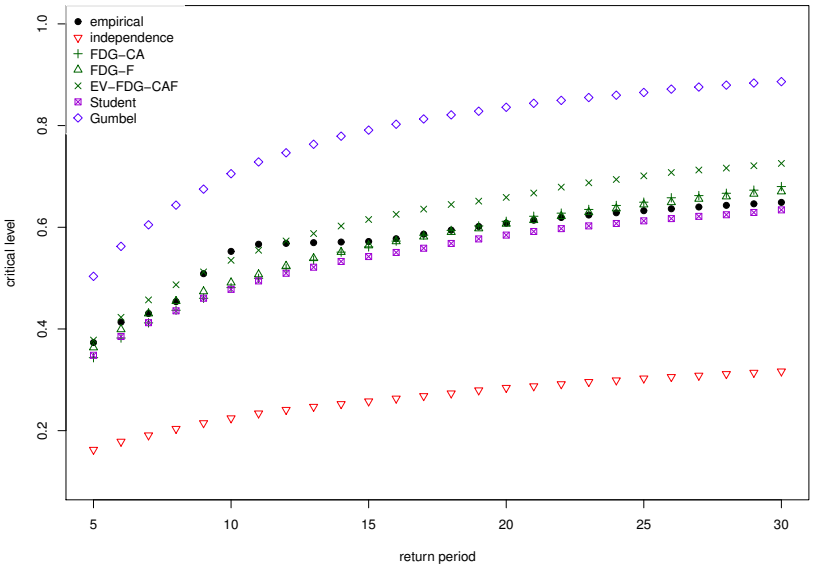

Figure 5: Critical level q as a function of the return period T. "empirical" stands for the empirical critical levels, and "independence" for the independence copula $C(u_1, \ldots, u_d) = \prod_{i=1}^d u_i$.

With such a small sample size $n = 32$, one must be extremely careful when looking at empirical data, because one is likely to observe a large deviation from the true underlying statistical distribution. In view of this remark, one should select a statistical model based not only on empirical data, but also on the model properties. The class of FDG copulas is very interesting in this respect. Indeed, the practitioner has with this class three models that fit well the data and with different features: FDG-F is upper, lower, and symmetric tail dependent, FDG-CA is upper tail dependent but no lower tail dependent, and EV-FDG-CAF is an extreme-value copula. The user is then free to choose the model that most suits his expert knowledge about the underlying phenomenon at play. The test

of extreme-value dependence [25] gave a p-value of 0.21, which means that one does not reject extreme-value copulas at the 5% level. Of course, as before, one must be extremely careful when looking at the p-value because of the small data sample size. Finally, goodness-of-fit tests for a given parametric family can be found in [26].

5 Discussion

In this article, we have constructed a new class of copulas by combining one-factor copulas, that is, a conditional independent property, together with a class of bivariate copulas called the Durante class of bivariate copulas. This combination led to many advantageous properties. The copulas within the proposed class, referred to as FDG copulas, are tractable, flexible, and cover all types of tail dependencies. The theoretically well grounded least-squares inference estimator is particularly well suited for FDG copulas because their dependence coefficients are easy to compute, if not in closed form. This allows to perform fast and reliable inference in the parametric case. We have demonstrated, furthermore, that FDG copulas work well in practice and are able to model both high-dimensional and real datasets. Finally, we have derived the extreme-value copulas (EV-FDG) associated to FDG copulas, yielding a new extreme-value copula, which can be viewed as a generalisation of the well-known Cuadras-Augé copula. This copula benefits from almost all the many advantageous properties of FDG copulas, and therefore opens the door for statistical analyses of extreme data in high dimension.

One may argue that a model with a singular component, as a FDG copula, is not natural nor realistic to model hydrological data. While this may be true in the bivariate case, this argument becomes weaker when the dimension increases. Indeed, in high-dimensional applications, the focus is less on the distribution itself than on a feature of interest of the data, such as, for instance, the critical levels defined in (22). If so-called "unrealistic" models are able to better estimate these features than "realistic" models – compare the fit of the Gumbel copula to the fit of FDG copulas in Section 4.2 – then one should consider using them.

This work raises several research questions. First, how to estimate the generators nonparametrically? The generator of a bivariate Durante copula was estimated nonparametrically in [12], but the matter is more complicated in our case because this bivariate relationship occurs between the variable of interest and the unobserved latent factor. Second, one may add more factors when building an FDG copula. Nonetheless, the model might not be as tractable as it is and therefore it may be less appealing in practice. Finally, FDG copulas possess the conditional independence property, but the extreme-value EV-FDG copulas were not shown to do so. If this property held, this would be of great interest for the simulation of datasets from this model.

Acknowledgments. The authors thank "Banque HYDRO du Ministère de l'Écologie, du Développement durable et de l'Énergie" for providing the data and Benjamin Renard for fruitful discussions about statistical issues in hydrological science. They also thank two anonymous referees and the associate editor for helpful suggestions and comments.

A Appendix

Proof of Theorem 1 Let $C_{j|0}(\cdot|u_0)$ be the conditional distribution of U_j given $U_0 = u_0$. The U_j's are conditionally independent given U_0, hence,

$$\begin{aligned}
C(u) &= \int_0^1 \prod_{j=1}^d C_{j|0}(u_j|u_0)du_0 \qquad (23)\\
&= \int_0^1 \prod_{j=1}^d \frac{\partial C_{0j}(u_0,u_j)}{\partial u_0} du_0\\
&= \int_0^1 \prod_{j=1}^d \frac{\partial C_{0(j)}(u_0,u_{(j)})}{\partial u_0} du_0\\
&= \int_0^{u_{(1)}} \prod_{j=1}^d \frac{\partial C_{0(j)}(u_0,u_{(j)})}{\partial u_0} du_0\\
&+ \sum_{k=2}^d \int_{u_{(k-1)}}^{u_{(k)}} \prod_{j=1}^{k-1} \frac{\partial C_{0(j)}(u_0,u_{(j)})}{\partial u_0} \prod_{j=k}^d \frac{\partial C_{0(j)}(u_0,u_{(j)})}{\partial u_0}\\
&+ \int_{u_{(d)}}^1 \prod_{j=1}^d \frac{\partial C_{0(j)}(u_0,u_(j))}{\partial u_0} du_0.
\end{aligned}$$

Since

$$\frac{\partial C_{0j}(u_0,u_j)}{\partial u_0} = \begin{cases} f_j(u_j) & \text{if } u_0 < u_j \\ u_j f_j'(u_0) & \text{if } u_0 > u_j, \end{cases}$$

equation (23) yields

$$\begin{aligned}
C(u) =& u_{(1)} \prod_{j=1}^d f_j(u_j) + \sum_{k=2}^d \prod_{j=k}^d f_{(j)}(u_{(j)}) \int_{u_{(k-1)}}^{u_{(k)}} \prod_{j=1}^{k-1} u_{(j)} f'_{(j)}(u_0)du_0\\
&+ \prod_{j=1}^d u_j \int_{u_{(d)}}^1 f_j'(u_0)du_0.
\end{aligned}$$

Putting $u_{(1)}$ in factor and noting that $\int_{u_{(1)}}^{u_{(2)}} f'_{(1)}(x)dx = f_{(1)}(u_{(2)}) - f_{(1)}(u_{(1)})$ finishes the proof.

Proof of Proposition 1 It suffices to set all u_k equal to one but u_i and u_j in the formula (3).

Proof of Proposition 2 It suffices to apply the formulas (4) with f_{ij} given in Proposition 1. To compute the Spearman's rho, note that

$$\int_0^1 x^2 f_{ij}(x)dx = \int_0^1 x^2 f_i(x)f_j(x)dx + \int_0^1 x^3 \int_x^1 f_i'(z)f_j'(z)dzdx.$$

An integration by parts yields $\int_0^1 x^3 \int_x^1 f_i'(z)f_j'(z)dzdx = (1/4)\int_0^1 x^4 f_i'(x)f_j'(x)dx$ and the result follows.

Proof of Theorem 2 Fix $(u_1, \ldots, u_d) \in [0,1]^d$ and let $n \geq 1$ be an integer. Let us introduce

$$\alpha_n := u_{(1)}^{1/n} \prod_{j=1}^d \frac{f_j(u_j^{1/n})}{u_j^{1/n}}, \quad \beta_n := \int_{u_{(d)}^{1/n}}^1 \prod_{j=1}^d f_j'(u_0) du_0,$$

$$\gamma_n := \prod_{j=k}^d \frac{f_{(j)}(u_{(j)}^{1/n})}{u_{(j)}^{1/n}}, \qquad \delta_{n,k} := \int_{u_{(k-1)}^{1/n}}^{u_{(k)}^{1/n}} \prod_{j=1}^{k-1} f_{(j)}'(u_0) du_0,$$

and define

$$A_n := \alpha_n + \beta_n + \sum_{k=2}^d \gamma_{n,k} \delta_{n,k}.$$

Our goal is to derive asymptotic equivalent sequences for $\alpha_n, \beta_n \gamma_n$ and δ_n. Let $\sim$ denote the equivalent symbol at infinity (i.e., $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$). By using the well known formulas $e^x \sim 1 + x$ (when $x \to 0$), $\log x \sim x - 1$ (when $x \to 1$) and $f_j(x) \sim 1 + (x-1) f_j'(1)$ (when $x \to 1$), we get

$$\alpha_n \sim \left(1 + \frac{1}{n} \log u_{(1)}\right) \left(1 - \frac{1}{n} \log(u_1 \ldots u_d)\right) \left(1 + \frac{1}{n} \sum_{j=1}^d \log u_{(j)} f_{(j)}'(1)\right) \text{ and}$$

$$\gamma_{n,k} \sim \left(1 - \frac{1}{n} \sum_{j=k}^d \log u_{(j)}\right) \left(1 + \frac{1}{n} \sum_{j=k}^d \log u_{(j)} f_{(j)}'(1)\right).$$

For β_n the equivalence is obtained as follows. Let $F(x)$ be a primitive of $\prod_{j=1}^d f_j'(x)$. It follows that $\beta_n = F(1) - F(u_{(d)}^{1/n})$. A Taylor expansion yields

$$F(u_{(d)}^{1/n}) = F(1) + (u_{(d)}^{1/n} - 1) F'(1) + \frac{(u_{(d)}^{1/n} - 1)^2}{2} F''(x_n)$$

where x_n is between $u_{(d)}^{1/n}$ and 1. Since F'' is assumed to be continuous on $[0,1]$, it is uniformly bounded on this set and therefore $(u_{(d)}^{1/n} - 1)^2 F''(x_n)/2 = o(1/n)$ where $o(1/n)$ is a quantity such that $no(1/n) \to 0$ as $n \to \infty$. Hence, since $u_{(d)}^{1/n} = \exp(\log(u_{(d)})/n) \sim 1 + \log(u_{(d)})/n$, we have as $n \to \infty$

$$F(1) - F(u_{(d)}^{1/n}) \sim -\frac{1}{n} \log(u_{(d)}) F'(1).$$

The same arguments apply to get

$$\beta_n \sim -\frac{1}{n} \log u_{(d)} \prod_{j=1}^d f_j'(1)$$

$$\delta_{n,k} \sim \frac{1}{n} \log\left(\frac{u_{(k)}}{u_{(k-1)}}\right) \prod_{j=1}^{k-1} f_{(j)}'(1).$$

The quantity A_n is a polynomial with respect to n^{-1} of order at most three. In (24), the coefficients of order 0, 2, and 3 vanish at infinity. Only remain the terms of order 1, hence,

$$\lim_{n\uparrow\infty} n(A_n - 1) = \log u_{(1)} - \log(u_1 \dots u_d) + \sum_{j=1}^{d} \log u_{(j)} f'_{(j)}(1) \tag{24}$$
$$- \log u_{(d)} \prod_{j=1}^{d} f'_{(j)}(1) + \sum_{k=2}^{d} \prod_{j=1}^{k-1} f'_{(j)}(1) \log\left(\frac{u_{(k)}}{u_{(k-1)}}\right).$$

From Abel's identity for two sequences (a_i) and (b_i) of real numbers, that is,

$$\sum_{i=1}^{d-1} a_i(b_{i+1} - b_i) = \sum_{i=1}^{d-1} b_i(a_{i-1} - a_i) + a_{d-1}b_d - a_1 b_1$$

it follows

$$\lim_{n\uparrow\infty} n(A_n - 1) = \sum_{k=1}^{d} \left(\underbrace{\left(\prod_{j=1}^{k-1} f'_{(j)}(1)\right)(1 - f'_{(k)}(1)) + f'_{(k)}(1) - 1}_{=:\chi_k} \right) \log u_{(k)},$$

with the convention that $\prod_{j=1}^{0} f'_{(j)}(1) = 1$. Then (3) entails that

$$\begin{aligned} C^n(u_1^{1/n}, \dots, u_d^{1/n}) =& u_1 \dots u_d \exp\left[n \log A_n\right] \\ =& u_1 \dots u_d \exp\left[n(A_n - 1)(1 + o(1))\right] \\ \to& \prod_{i=k}^{d} u_{(k)}^{\chi_k} \end{aligned}$$

as $n \to \infty$.

Proof of Lemma 1 (i) Since the p-uple $(r_{1,2}, \dots, r_{d-1,d})$ belongs to the image space $\mathbf{r}(\Theta^d)$, the system

$$\left\{ \begin{array}{ccc} r(\theta_1, \theta_2) & = & r_{1,2} \\ \vdots & & \vdots \\ r(\theta_{d-1}, \theta_d) & = & r_{d-1,d} \end{array} \right.$$

has at least one solution. In particular, there exists $(\theta_1, \theta_2, \theta_3)$ in Θ^3 such that

$$\left\{ \begin{array}{ccc} r(\theta_1, \theta_2) & = & r_{1,2} \\ r(\theta_1, \theta_3) & = & r_{1,3} \\ r(\theta_2, \theta_3) & = & r_{2,3}. \end{array} \right. \tag{25}$$

The system (25) rewrites

$$\left\{ \begin{array}{ccc} r_{\theta_2}(\theta_1) & = & r_{1,2} \\ r_{\theta_3}(\theta_1) & = & r_{1,3} \\ r_{\theta_3}(\theta_2) & = & r_{2,3}, \end{array} \right.$$

or equivalently,

$$\begin{cases} \theta_1 &= s_{1,2}(\theta_2) \\ \theta_1 &= s_{1,3}(\theta_3) \\ \theta_2 &= s_{2,3}(\theta_3). \end{cases}$$

This yields

$$s_{1,3}(\theta_3) = s_{1,2} \circ s_{2,3}(\theta_3). \tag{26}$$

Let us note that $s_{1,3}$ is involutive at θ_3, that is, $s_{1,3} \circ s_{1,3}(\theta_3) = \theta_3$. Indeed, $r(\theta_1, \theta_3) = r_{\theta_3}(\theta_1) = r_{1,3}$ is equivalent to $\theta_1 = r_{\theta_3}^{-1}(r_{1,3}) = s_{1,3}(\theta_3)$. This implies $r(s_{1,3}(\theta_3), \theta_3) = r_{1,3}$, and, composing by $r_{s_{1,3}(\theta_3)}^{-1}$ in both sides, we get $r_{s_{1,3}(\theta_3)}^{-1}(r_{1,3}) = s_{1,3}(s_{1,3}(\theta_3)) = \theta_3$. Therefore, one can compose both sides of (26) by $s_{1,3}$ to get

$$\theta_3 = s_{1,3} \circ s_{1,2} \circ s_{2,3}(\theta_3).$$

Hence (20) has at least one solution.

(ii) If (20) admits exactly one solution θ_3, then θ_2 and θ_1 are also unique. Furthermore, for all $j \geq 3$,

$$\theta_{j+1} = s_{j,j+1}(\theta_j)$$

which concludes the proof that assumption (A2) holds. It is now shown that assumption (A3) holds as well. Define $\partial_1 r$, respectively, $\partial_2 r$, the derivative of r with respect to the first, respectively, second, variable of r. Hence for all θ_i and θ_j in Θ, the quantities $\partial_1 r(\theta_i, \theta_j)$ and $r_{\theta_j}(\theta_i)$ only differ in the notation. The first step in the proof is to consider the case $d = 3$. The Jacobian matrix of $\mathbf{r}$ at $\boldsymbol{\theta_0}$ is given by

$$\mathbf{J} = \begin{pmatrix} \partial_1 r(\theta_{01}, \theta_{02}) & \partial_2 r(\theta_{01}, \theta_{02}) & 0 \\ \partial_1 r(\theta_{01}, \theta_{03}) & 0 & \partial_2 r(\theta_{01}, \theta_{03}) \\ 0 & \partial_1 r(\theta_{02}, \theta_{03}) & \partial_2 r(\theta_{02}, \theta_{03}) \end{pmatrix}.$$

To show that $\mathbf{J}$ has full rank, we show that its determinant

$$\partial_1 r(\theta_{01}, \theta_{02}) \partial_2 r(\theta_{01}, \theta_{03}) \partial_1 r(\theta_{02}, \theta_{03}) + \partial_2 r(\theta_{02}, \theta_{03}) \partial_1 r(\theta_{01}, \theta_{03}) \partial_2 r(\theta_{01}, \theta_{02})$$

is nonzero. Indeed, note that for all θ in Θ, the map $r_\theta : \Theta \to r_\theta(\Theta)$ is a twice continuously differentiable homeomorphism. Furthermore, by assumption, the true parameter vector $\boldsymbol{\theta_0}$ lies in the interior of Θ that is open. Finally, by symmetry, for all $i < j$,

$$\partial r_1(\theta_{0i}, \theta_{0j}) > 0 \quad (\text{respectively } \partial r_1(\theta_{0i}, \theta_{0j}) < 0)$$

is equivalent to

$$\partial r_2(\theta_{0j}, \theta_{0i}) > 0 \quad (\text{respectively } \partial r_2(\theta_{0j}, \theta_{0i}) < 0).$$

For the general case, we proceed by mathematical induction. When the dimension is d, write $\mathbf{J}(\boldsymbol{\theta}) = \mathbf{J}^{(d)}(\boldsymbol{\theta})$ to emphasize the dependence on the dimension. Notice that it was already shown above that $\mathbf{J}^{(3)}(\boldsymbol{\theta})$ has full rank. Now suppose that the kernel of $\mathbf{J}^{(d-1)}(\boldsymbol{\theta})$ is null when the dimension is $d-1$. Let $\mathbf{A} = \mathbf{J}^{(d)}(\boldsymbol{\theta})$. Each row of $\mathbf{A}$ writes

$$(0, \ldots, 0, \partial_1 r(\theta_i, \theta_j), 0 \ldots, 0, \partial_2 r(\theta_i, \theta_j), 0, \ldots, 0)$$

where $\partial_1 r(\theta_i, \theta_j)$ is at the i-th position and $\partial_2 r(\theta_i, \theta_j)$ at the j-th position. There are $d-1$ rows of A which depend on θ_d and $p-d+1$ which do not (recall $p = d(d-1)/2$ is the number of pairs). Since the kernel of a matrix is invariant by permutation, we can without loss of generality put all the rows which do not depend on θ_d on the top. More precisely, decompose $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

such that $\mathbf{A}_{11}$ is a $(p-d+1) \times (d-1)$ matrix containing all the rows which do not depend on θ_d and $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$ are $(p-d+1) \times 1$ and $(d-1) \times 1$ matrices respectively. Note that $\mathbf{A}_{12}$ is the null vector of size $p-d+1 \times 1$. Let $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x} = (\mathbf{x}_1^T, x_2)^T$ where $\mathbf{x}_1 \in \mathbb{R}^{d-1}$, $x_2 \in \mathbb{R}$. It follows that $\mathbf{A}\mathbf{x} = \mathbf{0}$ is equivalent to

$$\begin{cases} \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}x_2 = \mathbf{0} \\ \mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}x_2 = 0. \end{cases}$$

But $\mathbf{A}_{12} = \mathbf{0}$ and since $\mathbf{A}_{11} = \mathbf{J}^{(d-1)}(\boldsymbol{\theta})$ whose kernel is null, $\mathbf{x}_1 = \mathbf{0}$. Then $\mathbf{A}_{22}x_2 = 0$ and the assumptions imply $x_2 = 0$, which concludes the proof.

Proof of Theorem 3 To prove (i), it suffices to apply Lemma 1. Since $r(\theta_i, \theta_j)$ denotes the extremal dependence coefficient of the $\mathcal{E}$ copula bivariate marginal $C_{\#,ij}$ defined in (14), we have

$$r(\theta_i, \theta_j) = \lambda(\theta_i)\lambda(\theta_j), \text{ where } \lambda(\theta) := 1 - \frac{\partial f(t,\theta)}{\partial t}\Big|_{t=1}. \tag{27}$$

In the Cuadras-Augé and the Fréchet cases, (27) is given by $\lambda(\theta) = \theta$, and in the sinus case, $\lambda(\theta) = 1 - \theta/\tan(\theta)$. In all these situations, it is easy to see that the map $r_{\theta_j}(\cdot)$ is a twice continuously differentiable homeomorphism. Therefore, Lemma 1 (i) applies. To apply the second part of Lemma 1, note that equation (20) translates into

$$\lambda(\theta)^2 = \frac{r_{1,3}r_{2,3}}{r_{1,2}}.$$

Since it has a unique solution, Lemma 1 (ii) applies, and the result is proved.

(ii) The proof is straightforward because under the assumptions of (ii), $r(\theta_i, \theta_j) = \theta_i\theta_j$. But this is the precise form of $r(\theta_i, \theta_j)$ in (i); hence, one can also apply Lemma 1.

References

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

[2] E. F. Acar, C. Genest, and J. Nešlehová. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90, 2012.

[3] C. Amblard and S. Girard. A new extension of bivariate FGM copulas. *Metrika*, 70(1):1–17, 2009.

[4] T. Bedford and R.M. Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4):245–268, 2001.

[5] T. Bedford and R.M. Cooke. Vines–a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

[6] S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.

[7] C.M. Cuadras and J. Augé. A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4):339–353, 1981.

[8] M. De Carvalho. Confidence intervals for the minimum of a function using extreme value statistics. *International Journal of Mathematical Modelling and Numerical Optimisation*, 2(3):288–296, 2011.

[9] M. de Carvalho. A generalization of the solis–wets method. *Journal of Statistical Planning and Inference*, 142(3):633–644, 2012.

[10] S. Demarta and A. J. McNeil. The t copula and related copulas. *International statistical review*, 73(1):111–129, 2005.

[11] F. Durante. A new class of symmetric bivariate copulas. *Nonparametric Statistics*, 18(7-8):499–510, 2006.

[12] F. Durante and O. Okhrin. Estimation procedures for exchangeable marshall copulas with hydrological application. *Stochastic Environmental Research and Risk Assessment*, published online, 2014.

[13] F. Durante, J.J. Quesada-Molina, and M. Úbeda Flores. On a family of multivariate copulas for aggregation processes. *Information Sciences*, 177(24):5715–5724, 2007.

[14] F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161, 2010.

[15] M. Ferreira. Nonparametric estimation of the tail-dependence coefficient. *REVSTAT–Statistical Journal*, 11(1):1–16, 2013.

[16] G. Frahm, M. Junker, and A. Szimayer. Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286, 2003.

[17] M. Fréchet. Remarques au sujet de la note précédente. *CR Acad. Sci. Paris Sér. I Math*, 246:2719–2720, 1958.

[18] C. Genest and A. C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.

[19] C. Genest, J. Nešlehová, and N. Ben Ghorbal. Estimators based on Kendall's tau in multivariate copula models. *Australian & New Zealand Journal of Statistics*, 53(2):157–177, 2011.

[20] C. Genest and L.-P. Rivest. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043, 1993.

[21] G. Gudendorf and J. Segers. Extreme-value copulas. In P. Jaworski, F. Durante, W.K. Härdle, and T. Rychlik, editors, *Copula Theory and Its Applications*, page 127–145. Springer, 2010.

[22] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

[23] H. Joe. *Multivariate models and dependence concepts.* Chapman & Hall/CRC, Boca Raton, FL, 2001.

[24] C. Klüppelberg and G. Kuhn. Copula structure analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):737–753, 2009.

[25] I. Kojadinovic, J. Segers, and J. Yan. Large-sample tests of extreme-value dependence for multivariate copulas. *Canadian Journal of Statistics*, 39(4):703–720, 2011.

[26] I. Kojadinovic and J. Yan. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21(1):17–30, 2011.

[27] P. Krupskii and H. Joe. Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101, 2013.

[28] D. Kurowicka and R.M. Cooke. Distribution-free continuous Bayesian belief nets. In *Proceedings of Mathematical methods in Reliability Conference, Santa Fe, New Mexico, USA*, 2004.

[29] J.F. Mai and M. Scherer. Lévy-frailty copulas. *Journal of Multivariate Analysis*, 100(7):1567–1585, 2009.

[30] G. Mazo, S. Girard, and F. Forbes. Weighted least-squares inference based on dependence coefficients for multivariate copulas. *http://hal.archives-ouvertes.fr/hal-00979151*, 2014.

[31] A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques, and tools.* Princeton university press, 2010.

[32] R.B. Nelsen. *An introduction to copulas.* Springer, New York, 2006.

[33] D. Hwan Oh and A. J. Patton. Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, 108(502):689–700, 2013.

[34] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013.

[35] M. R. Veall. Testing for a global maximum in an econometric context. *Econometrica*, pages 1459–1465, 1990.

[36] L. Zhang and V. P. Singh. Gumbel–Hougaard copula for trivariate rainfall frequency analysis. *Journal of Hydrologic Engineering*, 12(4):409–419, 2007.