

Semi-parametric approximation of Kendall's distribution function and multivariate Return Periods *

Titre: Approximation semi-paramétrique de la fonction de distribution de Kendall et périodes de retour multivariées

Gianfausto Salvadori¹, Fabrizio Durante² and Elisa Perrone³

Abstract: In this work we outline a constructive approach for the approximation of Kendall's distribution function and Kendall's Return Period in the bivariate case. First, we introduce a suitable theoretical framework, based on the Theory of Copulas, where to embed the issue. Then, we outline an original construction procedure to approximate the empirical Kendall distribution function estimated using the available data. The whole approach is semi-parametric: the empirical Kendall distribution function is approximated via a (suitable) continuous piece-wise linear function on the unit interval. A sensitivity analysis is carried out via a simulation procedure, in order to investigate the robustness of the approach proposed against several relevant factors.

Résumé : Dans le cas bivarié, une approximation de la fonction de distribution de Kendall ainsi que de la période de retour de Kendall sont proposées. Le contexte théorique est celui de la théorie des copules. L'approche semi-paramétrique suggérée consiste à approcher la fonction de distribution de Kendall empirique par une fonction linéaire par morceaux sur l'intervalle unité. La robustesse de l'approche proposée est étudiée empiriquement par le biais de simulations.

Keywords: Kendall's distribution function, multivariate return periods, copulas, risk assessment

Mots-clés : fonction de distribution de Kendall, périodes de retour multivariés, copules, évaluation de risque

AMS 2000 subject classifications: 62G05, 62H12, 86A05

* Helpful discussions with C. Sempi (Università del Salento, Italy) are acknowledged. The comments and the suggestions of two anonymous Referees are also gratefully acknowledged. [GS] The research was partially supported by the Italian M.I.U.R. via the project "Metodi stocastici in finanza matematica" (PRIN 2010). Also the support of the CMCC—Centro Mediterraneo Cambiamenti Climatici - Lecce (Italy) is acknowledged. [FD] The Author acknowledges the support of School of Economics and Management, Free University of Bozen-Bolzano (Italy), via the project "Risk and Dependence".

¹ Dipartimento di Matematica e Fisica, Università del Salento, I-73100 Lecce, Italy.

E-mail: gianfausto.salvadori@unisalento.it

² School of Economics and Management, Free University of Bozen-Bolzano, I-39100 Bolzano, Italy.

E-mail: fabrizio.durante@unibz.it

³ Institut für Angewandte Statistik, Johannes Kepler Universität, A-4040 Linz, Austria.

E-mail: elisa.perrone@jku.at

1. Introduction

The notion of Return Period (hereinafter, RP) is frequently used in environmental sciences for the identification of dangerous events, and provides a means for rational decision making and risk assessment (for a review, see [Singh et al., 2007, AghaKouchak et al., 2013], and references therein). Roughly speaking, the RP can be considered [Nappo and Spizzichino, 2009] as an analogue of the “Value-at-Risk” in Economics and Finance, since it is used to quantify and assess the risk.

The traditional definition of the RP is as “the average time elapsing between two successive realizations of a prescribed event”, which clearly has a statistical base. Equally important is the related concept of *design quantile*, usually defined as “the value of the variable(s) characterizing the event associated with a given RP ”. In engineering practice, the choice of the RP depends upon the importance of the structure, and the consequences of its failure. While in the univariate case the design quantile is usually identified without ambiguity, in the multivariate one this is not so. Indeed, the identification problem of design events in a multivariate context has recently attracted the attention of many researches (see, e.g., [Serfling, 2002, Belzunce et al., 2007, Chebana and Ouarda, 2009, Chebana and Ouarda, 2011, Chaouch and Goga, 2010, AghaKouchak et al., 2013], and references therein).

As we shall show later, the calculation of the RP is strictly related to the notion of Copula, which is the restriction of a joint distribution with Uniform margins to $\mathbf{I}^d = [0, 1]^d$, $d > 1$. The link between a multivariate distribution \mathbf{F} and the associated d -dimensional copula \mathbf{C} is given by the functional identity stated by Sklar’s Theorem [Sklar, 1959, Durante et al., 2012]:

$$\mathbf{F}(x_1, \dots, x_d) = \mathbf{C}(F_1(x_1), \dots, F_d(x_d)) \quad (1)$$

for all $\mathbf{x} \in \mathbf{R}^d$, where the F_i ’s are the univariate margins of \mathbf{F} . If all the F_i ’s are continuous, then \mathbf{C} is unique. Most importantly, the F_i ’s in Eq. (1) only play the role of (geometrically) re-mapping the probabilities induced by \mathbf{C} on the subsets of \mathbf{I}^d onto suitable subsets of \mathbf{R}^d , without changing their values: viz., the dependence structure modeled by \mathbf{C} plays a central role in tuning the probabilities of joint occurrences. For a thorough theoretical introduction to copulas see [Joe, 1997, Nelsen, 2006, Durante and Sempi, 2010]; for a practical approach see [Salvadori et al., 2007, Jaworski et al., 2010].

In order to avoid troublesome situations, hereinafter we shall assume that \mathbf{F} is continuous (but not necessarily absolutely continuous), and strictly increasing in each marginal. In turn, any point $\mathbf{x} \in \mathbf{R}^d$ can be uniquely re-mapped onto $\mathbf{u} \in \mathbf{I}^d$ (and vice-versa) via the Probability Integral Transform (PIT):

$$(u_1, \dots, u_d) = (F_1(x_1), \dots, F_d(x_d)). \quad (2)$$

Later we shall use the *Kendall’s distribution* (or *measure*) function $\mathbf{K}_{\mathbf{C}} : \mathbf{I} \rightarrow \mathbf{I}$ [Genest and Rivest, 1993, Genest and Rivest, 2001] given by

$$\mathbf{K}_{\mathbf{C}}(t) = \mathbb{P}(W \leq t) = \mathbb{P}(\mathbf{C}(U_1, \dots, U_d) \leq t), \quad (3)$$

where $t \in \mathbf{I}$ is a probability level, $W = \mathbf{C}(U_1, \dots, U_d)$ is a univariate random variable (hereinafter, *r.v.*) taking value on \mathbf{I} , and the U_i ’s are Uniform *r.v.s* on \mathbf{I} with copula \mathbf{C} . Note that $\mathbf{K}_{\mathbf{C}}(t)$ practically measures the probability that a random event will appear in the region of \mathbf{I}^d defined by

the inequality $\mathbf{C}(\mathbf{u}) \leq t$: this function was introduced as a generalization of the PIT, and stands its name by the fact that it is related to Kendall's measure of association — see [Genest and Rivest, 2001, Nappo and Spizzichino, 2009].

While in some cases (e.g., for bivariate Extreme Value copulas [Ghoudi et al., 1998], or Archimedean copulas [Barbe et al., 1996, McNeil and Nešlehová, 2009]) a close formula for \mathbf{K}_C is available, in general it is possible to resort to simulations for estimating specific values of \mathbf{K}_C (see, e.g., the procedures outlined in [Barbe et al., 1996, Salvadori et al., 2011], and references therein). As we shall see, \mathbf{K}_C turns out to be a fundamental tool for calculating a copula-based RP for multivariate events.

The paper is organized as follows. In Section 2 we recall consistent definitions of RP's and quantiles in a multivariate environment. In Section 3 we present a constructive method for the approximation of Kendall's distribution in the bivariate case. In Section 4 we carry out a simulation study, in order to test the performance of the approach outlined in the paper. Finally, in Section 5 we draw some conclusions and discuss possible future perspectives.

2. Multivariate Return period

In order to provide a consistent theory of RP's in a multivariate environment, it is first necessary to define the abstract framework where to embed the question. Preliminary studies can be found in [Salvadori, 2004, Salvadori and De Michele, 2004, Durante and Salvadori, 2010, Salvadori and De Michele, 2010], and some applications are presented in [De Michele et al., 2007, Salvadori and De Michele, 2010, Vandenberghe et al., 2010, Salvadori et al., 2011]. Hereinafter, we shall consider as the object of our investigation a sequence $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots\}$ of independent and identically distributed d -dimensional random vectors, with $d > 1$: thus, each \mathbf{X}_i has the same multivariate distribution \mathbf{F} as of the random vector $\mathbf{X} \sim \mathbf{F} = \mathbf{C}(F_1, \dots, F_d)$ describing the phenomenon under investigation, with suitable marginals F_i 's and d -copula \mathbf{C} .

In applications, usually, the event of interest is of the type $\{\mathbf{X} \in \mathcal{D}\}$, where \mathcal{D} is a non-empty Borel set in \mathbf{R}^d collecting all the values judged to be “dangerous” according to suitable criteria. Let $\mu > 0$ be the average inter-arrival time of the realizations in \mathcal{X} (viz., μ is the average time elapsing between \mathbf{X}_i and \mathbf{X}_{i+1}). According to [Salvadori et al., 2011], we may introduce a consistent notion of RP as follows.

Definition 1. The RP associated with the event $\{\mathbf{X} \in \mathcal{D}\}$ is given by $\mu_{\mathcal{D}} = \mu / \mathbb{P}(\mathbf{X} \in \mathcal{D})$.

Definition 1 is a very general one: the set \mathcal{D} may be constructed in order to satisfy broad requirements, useful in different applications. Indeed, most of the approaches already present in literature are particular cases of the one outlined above. It is important to stress that the RP is a quantity associated with a proper event. However, with a slight abuse of language, we may also speak of “the RP of a realization”, meaning in fact “the RP of the event $\{\mathbf{X} \text{ belongs to the dangerous region } \mathcal{D}_{\mathbf{x}^*} \text{ identified by the given realization } \mathbf{x}^*\}$ ”. In a univariate framework, usually the assignment of \mathbf{x}^* uniquely identifies the corresponding region $\mathcal{D}_{\mathbf{x}^*}$; instead, this is not the case in a multivariate framework. Below we show how it is possible to associate a given multi-dimensional realization $\mathbf{x}^* \in \mathbf{R}^d$ with a dangerous region $\mathcal{D}_{\mathbf{x}^*} \subset \mathbf{R}^d$. First of all we need to introduce the following notion.

Definition 2. Given a d -dimensional distribution $\mathbf{F} = \mathbf{C}(F_1, \dots, F_d)$ and $t \in (0, 1)$, the *critical layer* $\mathcal{L}_t^{\mathbf{F}}$ of level t is defined as

$$\mathcal{L}_t^{\mathbf{F}} = \{\mathbf{x} \in \mathbf{R}^d : \mathbf{F}(\mathbf{x}) = t\}. \quad (4)$$

Clearly, $\mathcal{L}_t^{\mathbf{F}}$ is the iso-hyper-surface (having dimension $d - 1$) where \mathbf{F} equals the constant value t : thus, $\mathcal{L}_t^{\mathbf{F}}$ is a (iso)line for bivariate distributions, a (iso)surface for trivariate ones, and so on. Evidently, for any given $\mathbf{x} \in \mathbf{R}^d$, there exists a *unique* critical layer $\mathcal{L}_t^{\mathbf{F}}$ supporting \mathbf{x} : namely, the one identified by the level $t = \mathbf{F}(\mathbf{x})$. Note that, thanks to Eq. (2), there exists a one-to-one correspondence between the two iso-hyper-surfaces $\mathcal{L}_t^{\mathbf{C}} = \{\mathbf{u} \in \mathbf{I}^d : \mathbf{C}(\mathbf{u}) = t\}$ (pertaining to \mathbf{C} in \mathbf{I}^d) and $\mathcal{L}_t^{\mathbf{F}}$ (pertaining to \mathbf{F} in \mathbf{R}^d).

The critical layer $\mathcal{L}_t^{\mathbf{F}}$ partitions \mathbf{R}^d into three non-overlapping and exhaustive regions:

1. $\mathcal{R}_t^{<} = \{\mathbf{x} \in \mathbf{R}^d : \mathbf{F}(\mathbf{x}) < t\}$;
2. $\mathcal{L}_t^{\mathbf{F}}$, the critical layer itself;
3. $\mathcal{R}_t^{>} = \{\mathbf{x} \in \mathbf{R}^d : \mathbf{F}(\mathbf{x}) > t\}$.

Practically, at any occurrence of the phenomenon, only three mutually exclusive things may happen: either a realization of \mathbf{X} lies in $\mathcal{R}_t^{<}$, or over $\mathcal{L}_t^{\mathbf{F}}$, or in $\mathcal{R}_t^{>}$. Note that all these three regions are Borel sets.

Thanks to the above discussion, it is now clear that the following (multivariate) notion of RP is meaningful, and coincides with the one used in the univariate framework.

Definition 3. Let \mathbf{X} be a multivariate r.v. with distribution $\mathbf{F} = \mathbf{C}(F_1, \dots, F_d)$. Also, let $\mathcal{L}_t^{\mathbf{F}}$ be the critical layer supporting a realization \mathbf{x} of \mathbf{X} (i.e., $t = \mathbf{F}(\mathbf{x})$). Then, the RP associated with \mathbf{x} is defined as

1. for the region $\mathcal{R}_t^{>}$

$$T_{\mathbf{x}}^{>} = \mu / \mathbb{P}(\mathbf{X} \in \mathcal{R}_t^{>}), \quad (5)$$

2. for the region $\mathcal{R}_t^{<}$

$$T_{\mathbf{x}}^{<} = \mu / \mathbb{P}(\mathbf{X} \in \mathcal{R}_t^{<}). \quad (6)$$

For instance, in hydrology, $T_{\mathbf{x}}^{<}$ may be of interest if droughts are investigated, while the study of floods may require the use of $T_{\mathbf{x}}^{>}$. In the sequel we shall concentrate only upon $\mathcal{R}_t^{>}$: the corresponding formulas for $\mathcal{R}_t^{<}$ could easily be derived. Now, in view of the results outlined in [Nelsen et al., 2001, Nelsen et al., 2003], it is immediate to show that

$$T_{\mathbf{x}}^{>} = \frac{\mu}{1 - \mathbf{K}_{\mathbf{C}}(t)}, \quad (7)$$

where $\mathbf{K}_{\mathbf{C}}$ is the Kendall's distribution function associated with \mathbf{C} . Clearly, $T_{\mathbf{x}}^{>}$ is a function of the critical level t identified by the relation $t = \mathbf{F}(\mathbf{x})$. It is then convenient to denote the above RP via a special notation as follows.

Definition 4. The quantity $\kappa_{\mathbf{x}} = T_{\mathbf{x}}^{>}$ is called the *Kendall's RP* of the realization \mathbf{x} belonging to $\mathcal{L}_t^{\mathbf{F}}$ (hereinafter, *KRP*).

An advantage of the approach outlined in this work is that realizations lying over the same critical layer do always identify the same “dangerous” region (i.e., $\mathcal{R}_t^>$). In particular, all the realizations \mathbf{y} having a KRP $\kappa_{\mathbf{y}} < \kappa_{\mathbf{x}}$ must lie in $\mathcal{R}_t^<$, whereas all those \mathbf{y} having a KRP $\kappa_{\mathbf{y}} > \kappa_{\mathbf{x}}$ must lie in $\mathcal{R}_t^>$, while the realizations lying over $\mathcal{L}_t^{\mathbf{F}}$ share the same KRP $\kappa_{\mathbf{x}}$.

Traditionally, in the univariate framework, once a RP (say, T) is fixed (e.g., by design or regulation constraints), the corresponding critical probability level p is calculated as $1 - p = \mathbb{P}(X > x_p) = \mu/T$, and by inverting the distribution F_X of X it is then immediate to obtain the critical quantile $x_p = F_X^{(-1)}(p)$, which is usually unique. Then, x_p is used in practice for design purposes. As shown below, the same approach can also be adopted in a multivariate environment.

Definition 5. Given a d -dimensional distribution $\mathbf{F} = \mathbf{C}(F_1, \dots, F_d)$ with d -copula \mathbf{C} , and a probability level $p \in \mathbf{I}$, the *Kendall's quantile* $q_p \in \mathbf{I}$ of order p is defined as

$$q_p = \inf\{t \in \mathbf{I} : \mathbf{K}_{\mathbf{C}}(t) = p\} = \mathbf{K}_{\mathbf{C}}^{(-1)}(p), \tag{8}$$

where $\mathbf{K}_{\mathbf{C}}^{(-1)}$ is the (generalized) inverse of $\mathbf{K}_{\mathbf{C}}$.

Definition 5 provides a close analogy with the definition of univariate quantile: indeed, recall that $\mathbf{K}_{\mathbf{C}}$ is a univariate distribution function (see Eq. (3)), and hence q_p is simply the quantile of order p of $\mathbf{K}_{\mathbf{C}}$. In turn, p is the probability measure induced by \mathbf{C} on the region $\mathcal{R}_{q_p}^<$, while $(1 - p)$ is the one of $\mathcal{R}_{q_p}^>$. From a practical point of view this means that, in a large simulation of n independent d -dimensional vectors extracted from \mathbf{F} , np realizations are expected to lie in $\mathcal{R}_{q_p}^<$, and the others in $\mathcal{R}_{q_p}^>$. A heuristic procedure for estimating q_p (if it cannot be calculated analytically) is outlined in [Salvadori et al., 2011].

Remark 1. It is worth stressing that a common error is to confuse the value of the copula \mathbf{C} with the probability induced by \mathbf{C} on \mathbf{I}^d (and, hence, on \mathbf{R}^d): on the critical layer $\mathcal{L}_{q_p}^{\mathbf{C}}$ it is $\mathbf{C} = q_p$, but the corresponding region $\mathcal{R}_{q_p}^<$ has probability $p = \mathbf{K}_{\mathbf{C}}(q_p) \geq q_p$, since $\mathbf{K}_{\mathbf{C}}$ is usually non-linear (the same rationale holds for the region $\mathcal{R}_{q_p}^>$).

3. Approximation of $\mathbf{K}_{\mathbf{C}}$ (bivariate case)

Given the discussion in the previous Sections, it is now clear the importance of Kendall's distribution function within the multivariate RP theory. In this Section we present a constructive method for the approximation of a Kendall's distribution in the bivariate case. Multivariate generalizations will be discussed later.

Here the fundamental point is as follows. Given a random sample of m multivariate observations $\{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ extracted from a common joint distribution \mathbf{F} with copula \mathbf{C} , it is possible to estimate the corresponding values of $\mathbf{K}_{\mathbf{C}}$ at any point in \mathbf{I} (e.g., via the procedure outlined in [Barbe et al., 1996]). Then, the true (but unknown!) $\mathbf{K}_{\mathbf{C}}$ can be approximated as shown below. As a consequence, the KRP's of the events of interest can be estimated, without knowing the true copula model ruling the actual multivariate statistics.

3.1. Construction and features

Let \mathcal{T}_n be a dyadic partition of order $n > 0$ of the unit interval \mathbf{I} , i.e.:

$$t_i = \frac{i}{2^n}, \quad i = 0, \dots, 2^n. \quad (9)$$

Note that we use here a dyadic partition for the sake of simplicity only: actually, the present approach can easily be generalized to an arbitrary (finite) partition of \mathbf{I} including the points $\{0\}$ and $\{1\}$. Let \mathcal{Y}_n be the set of values

$$y_i = \hat{\mathbf{K}}(t_i), \quad i = 0, \dots, 2^n, \quad (10)$$

where the $\hat{\mathbf{K}}(t_i)$'s are the values, or empirical estimates, of the (unknown) Kendall's distribution function \mathbf{K}_C associated with an (unknown) copula \mathbf{C} . Clearly, $y_0 = 0$ and $y_{2^n} = 1$; also, in order to avoid troublesome cases — which would require an *ad hoc* treatment — we only use those y_i 's such that $y_i > t_i$ (see the admissibility discussion at the beginning of Section 4.2, the constraint on \mathbf{K}_n below Eq. (16), and the note on the coefficients $c_{i,n}$'s below Eq. (19)).

Now, the idea is to approximate the true \mathbf{K}_C , whose values \mathcal{Y}_n are assumed to be known (estimated) at the points \mathcal{T}_n , with a continuous, piecewise linear, distribution function \mathbf{K}_n , which depends upon the order n of the partition \mathcal{T}_n . Such a choice is a very natural one: on the one hand, \mathbf{K}_n is the simplest continuous polynomial, and its construction is easy and requires a least amount of parameters; on the other hand, the expression of the associated Archimedean generator (see below) is both analytically and computationally tractable. *En passant* we note that, given the constraints stated above, \mathbf{K}_n turns out to be a proper Kendall's distribution function.

In turn, we simply need to find suitable parameters $a_{i,n}$'s and $b_{i,n}$'s such that

$$\mathbf{K}_n(t) = a_{i,n} + b_{i,n}t, \quad t \in [t_{i-1}, t_i], \quad (11)$$

with $i = 1, \dots, 2^n$, and to solve the following system of equations:

$$\begin{cases} a_{i,n} + b_{i,n}t_{i-1} = y_{i-1} \\ a_{i,n} + b_{i,n}t_i = y_i \end{cases}. \quad (12)$$

The solutions are

$$\begin{cases} a_{i,n} = y_i - i(y_i - y_{i-1}) \\ b_{i,n} = 2^n(y_i - y_{i-1}) \end{cases}, \quad (13)$$

with $i = 1, \dots, 2^n$. Clearly, $b_{i,n} \geq 0$ for all indices i 's: indeed, \mathbf{K}_n should represent a distribution function over \mathbf{I} — see Eq. (3), and hence it must be increasing. In addition, \mathbf{K}_n converges pointwise a.e. to \mathbf{K}_C in \mathbf{I} as the order n of the partition \mathcal{T}_n and the sample size m increase.

Now, since \mathbf{K}_n is a (Kendall's) distribution function, Theorem 4.3.4 in [Nelsen, 2006] states that

$$\mathbf{K}_n(t) = t - \frac{\gamma_n(t)}{\gamma_n'(t^+)}, \quad t \in (0, 1), \quad (14)$$

where the function γ_n is the inner generator of a suitable Archimedean bivariate copula \mathbf{C}_n .

Remark 2. We recall that a (strict) inner Archimedean generator γ_n must be positive, convex, and strictly decreasing on \mathbf{I} , with $\gamma_n(0) = +\infty$ and $\gamma_n(1) = 0$. In turn, γ_n is continuous in \mathbf{I} .

In view of Eq. (14), it is then possible to calculate γ_n , and eventually to construct an Archimedean copula \mathbf{C}_n associated with \mathbf{K}_n . Note that, in the present case, $t - \mathbf{K}_n(t)$ is continuous and negative by construction for any n , and hence so is the ratio γ_n/γ'_n in $(0, 1)$.

Remark 3. The Archimedean copula \mathbf{C}_n associated with \mathbf{K}_n could be used to generate random samples having a Kendall's distribution approximating the unknown one of the available data (i.e., \mathbf{K}_C). Clearly, the whole approach is expected to provide reasonable simulations in case the underlying observations exhibit some kind of Archimedean dependence (e.g., exchangeability, or, at least, a symmetric dependence structure). Recent tests to check the presence of these features can be found in [Jaworski, 2010, Bücher et al., 2012, Genest et al., 2012, Kojadinovic and Yan, 2012].

Practically, we need to solve the set of differential equations

$$\frac{\gamma_{i,n}(t)}{\gamma'_{i,n}(t)} = t - \mathbf{K}_n(t) = -a_{i,n} + (1 - b_{i,n})t, \quad t \in [t_{i-1}, t_i], \tag{15}$$

in each i -th interval of the partition \mathcal{T}_n , with $i = 1, \dots, 2^n$. The corresponding solutions are

$$\gamma_{i,n}(t) = \begin{cases} c_{i,n} \exp(-t/a_{i,n}), & \text{if } b_{i,n} = 1 \\ c_{i,n} (|-a_{i,n} + (1 - b_{i,n})t|)^{\frac{1}{1-b_{i,n}}}, & \text{if } b_{i,n} \neq 1 \end{cases}, \tag{16}$$

where the $c_{i,n}$'s are suitable positive constants (see below), $t \in [t_{i-1}, t_i]$, and $i = 1, \dots, 2^n$. Using Eq. (15), and the constraint $\mathbf{K}_n(t) > t$ (except for $t = 0$ and $t = 1$, where $\mathbf{K}_n(t) = t$), it follows that the argument of the absolute value is strictly negative for all $t \in [t_{i-1}, t_i]$. As a consequence, the inner Archimedean generator γ_n has the following piecewise representation:

$$\gamma_{i,n}(t) = \begin{cases} c_{i,n} \exp(-t/a_{i,n}), & \text{if } b_{i,n} = 1 \\ c_{i,n} (a_{i,n} + (b_{i,n} - 1)t)^{\frac{1}{1-b_{i,n}}}, & \text{if } b_{i,n} \neq 1 \end{cases}, \tag{17}$$

with $t \in [t_{i-1}, t_i]$ and $i = 1, \dots, 2^n$. In turn, the full expression of γ_n is as follows:

$$\gamma_n(t) = \sum_{i=1}^{2^n} \gamma_{i,n}(t) \mathbf{1}_{[t_{i-1}, t_i]}(t). \tag{18}$$

The Lemmas given below show that γ_n features all the properties of a proper (strict) inner Archimedean generator.

Lemma 1. *The function γ_n , given by the piecewise representation (18), satisfies the boundary conditions $\gamma_n(0) = +\infty$ and $\gamma_n(1) = 0$.*

Proof. Due to the constraint $\mathbf{K}_n(t) > t$ for $t \in (0, 1)$, one cannot have $b_{1,n} = 1$ or $b_{2^n,n} = 1$, otherwise the first or last segments of the piecewise linear function \mathbf{K}_n would coincide with the diagonal of the unit square: more particularly, one must have $b_{1,n} > 1$ and $b_{2^n,n} < 1$, since $\mathbf{K}_n(0) = 0$ and $\mathbf{K}_n(1) = 1$. In turn, the first and last portions of γ_n necessarily only admit the power-law representation given by Eq. (17), and hence the boundary conditions $\gamma_n(0) = +\infty$ and $\gamma_n(1) = 0$ are trivially satisfied, being $\mathbf{K}_n(0) = a_{1,n} = 0$ and $\mathbf{K}_n(1) = a_{2^n,n} + b_{2^n,n} = 1$. \square

According to Remark 2, suitable conditions must be satisfied by the coefficients $c_{i,n}$'s, for all indices i 's, in order to yield a proper inner Archimedean generator: here we shall only deal with the non-trivial case $b_{i,n} \neq 1$. First of all, we fix backwards the $c_{i,n}$'s in such a way that γ_n , given by Eq. (18), be continuous in \mathbf{I} , i.e. $\gamma_{i,n}(t_i) = \gamma_{i+1,n}(t_i)$. This yields

$$c_{i,n} = c_{i+1,n} \frac{(a_{i+1,n} + (b_{i+1,n} - 1)t_i)^{\frac{1}{1-b_{i+1,n}}}}{(a_{i,n} + (b_{i,n} - 1)t_i)^{\frac{1}{1-b_{i,n}}}} \quad (19)$$

for $i = 1, \dots, 2^n - 1$. Clearly, $c_{i,n} > 0$ for all indices i 's, since $\mathbf{K}_n(t_i) > t_i$. Note that it is possible to choose any arbitrary starting value $c_{2^n,n} > 0$, since the inner generator of an Archimedean copula is uniquely defined up to a positive multiplicative constant. The following Lemmas show that γ_n is strictly decreasing and convex over \mathbf{I} .

Lemma 2. *The function γ_n , given by the piecewise representation (18), has a continuous and strictly negative first derivative over \mathbf{I} , and hence is strictly decreasing.*

Proof. A simple calculation shows that

$$\gamma'_n(t) = \gamma'_{i,n}(t) = \begin{cases} -\frac{c_{i,n}}{a_{i,n}} \exp(-t/a_{i,n}), & \text{if } b_{i,n} = 1 \\ -c_{i,n} (a_{i,n} + (b_{i,n} - 1)t)^{\frac{b_{i,n}}{1-b_{i,n}}}, & \text{if } b_{i,n} \neq 1 \end{cases}, \quad (20)$$

with $t \in [t_{i-1}, t_i]$ and $i = 1, \dots, 2^n$. Thus, γ'_n is strictly negative, and by using Eq. (19) it is immediate to show that γ'_n is also continuous, i.e. $\gamma'_{i,n}(t_i) = \gamma'_{i+1,n}(t_i)$. \square

Lemma 3. *The function γ_n , given by the piecewise representation (18), is convex over \mathbf{I} .*

Proof. In order to show the convexity of γ_n it is enough to prove that

$$\gamma'_n(y)/\gamma'_n(x) \leq 1 \quad (21)$$

for all $x \leq y$ in \mathbf{I} , i.e. γ'_n is a non-decreasing function. We first note that γ'_n is continuous and strictly negative, as shown in Lemma 2, and that all the piecewise components $\gamma'_{i,n}$'s are convex. Then, we proceed by induction over the partition \mathcal{T}_n .

First consider the components $\gamma'_{1,n}$ and $\gamma'_{2,n}$ defined over, respectively, $I_1 = [t_0, t_1]$ and $I_2 = [t_1, t_2]$. If x and y both belong either to I_1 or I_2 , then the inequality (21) trivially holds, being $\gamma'_{1,n}$ and $\gamma'_{2,n}$ both convex. Thus, let $t_0 \leq x < t_1 < y \leq t_2$. In turn, since γ'_n is continuous in t_1 , then $\gamma'_{1,n}(t_1) = \gamma'_{2,n}(t_1)$, and hence

$$\frac{\gamma'_{2,n}(y)}{\gamma'_{1,n}(x)} = \frac{\gamma'_{2,n}(y)}{\gamma'_{2,n}(t_1)} \cdot \frac{\gamma'_{1,n}(t_1)}{\gamma'_{1,n}(x)} \leq 1,$$

being $\gamma'_{1,n}$ and $\gamma'_{2,n}$ both convex. Thus, γ_n is convex in $[t_0, t_2]$. By the same token, we can then prove that γ_n is convex in $[t_0, t_3]$, in $[t_0, t_4]$, and so on, till the end point t_{2^n} of the unit interval \mathbf{I} . \square

Remark 4. An alternative proof can be given by exploiting the (equivalent) integral representation of γ_n given in [Genest and Rivest, 1993]:

$$\gamma_n(t) = \exp \left\{ \int_{x_0}^t \frac{1}{x - \mathbf{K}_n(x)} dx \right\},$$

where $t \in \mathbf{I}$ and $x_0 \in (0, 1)$ is arbitrary. Given the particular expression of \mathbf{K}_n , it is easy to show that the inequality (21) holds for all $x \leq y$ in \mathbf{I} (however, the tedious calculations are not shown).

Now, given the fact that γ_n is a proper inner Archimedean generator, we can provide the explicit expression of the corresponding Archimedean copula \mathbf{C}_n via the standard formula

$$\mathbf{C}_n(u, v) = \gamma_n^{-1}(\gamma_n(u) + \gamma_n(v)), \tag{22}$$

where $(u, v) \in \mathbf{I}^2$. Clearly, \mathbf{C}_n has \mathbf{K}_n as its Kendall's distribution function. A further feature of \mathbf{C}_n is as follows.

Proposition 1. *The \mathbf{C}_n -measure of the level curves of \mathbf{C}_n is zero.*

Proof. For any $t \in (0, 1)$, the t -level curve of \mathbf{C}_n is defined via the relation $\gamma_n(u) + \gamma_n(v) = \gamma_n(t)$. According to Theorem 4.3.3 in [Nelsen, 2006], the corresponding \mathbf{C}_n -measure is given by

$$\frac{\gamma_n(t)}{\gamma_n'(t^-)} - \frac{\gamma_n(t)}{\gamma_n'(t^+)}.$$

Since the ratio γ_n/γ_n' is continuous in $(0, 1)$ by construction, the claim follows. □

As is well known [Nelsen et al., 2003], Kendall's distribution function induces an equivalence relation on the set of copulas as follows.

Definition 6. Let \mathbf{C}_1 and \mathbf{C}_2 be two copulas with respective Kendall's distribution functions \mathbf{K}_1 and \mathbf{K}_2 . Then, \mathbf{C}_1 is in relation with \mathbf{C}_2 , and we write $\mathbf{C}_1 \equiv_{\mathbf{K}} \mathbf{C}_2$, if and only if $\mathbf{K}_1(t) = \mathbf{K}_2(t)$ for all $t \in \mathbf{I}$.

As a consequence, there is an entire class of copulas sharing the same Kendall's distribution function. In turn, the Archimedean copula \mathbf{C}_n given by Eq. (22) may play the role as of the "representative member" of the $\equiv_{\mathbf{K}_n}$ -equivalence class. Actually, in view of the results shown in [Nelsen et al., 2009, Genest et al., 2011a, Genest et al., 2011b], there is exactly one Archimedean copula in each equivalence class (however, it should be noted that such a uniqueness result is only proven for dimensions two and three).

The last point to be investigated is the convergence of \mathbf{C}_n to a proper copula $\tilde{\mathbf{C}}$: here we use the following result shown in [Charpentier and Segers, 2008].

Proposition 2. *Let γ_n' and $\tilde{\gamma}'$ be the right-hand derivatives of, respectively, the inner Archimedean generators γ_n and $\tilde{\gamma}$. Then, set*

$$\lambda_n(t) = \frac{\gamma_n(t)}{\gamma_n'(t)}, \quad \tilde{\lambda}(t) = \frac{\tilde{\gamma}(t)}{\tilde{\gamma}'(t)}, \quad \mathbf{K}_n(t) = t - \lambda_n(t), \quad \tilde{\mathbf{K}}(t) = t - \tilde{\lambda}(t),$$

where \mathbf{K}_n and $\tilde{\mathbf{K}}$ are, respectively, Kendall's distribution functions of the Archimedean copulas \mathbf{C}_n and $\tilde{\mathbf{C}}$. Then, the following five conditions are equivalent:

1. $\lim_{n \rightarrow \infty} \mathbf{C}_n(u, v) = \tilde{\mathbf{C}}(u, v)$ for all $(u, v) \in \mathbf{I}^2$.
2. $\lim_{n \rightarrow \infty} \frac{\gamma_n(u)}{\gamma_n'(v)} = \frac{\tilde{\gamma}(u)}{\tilde{\gamma}'(v)}$ for every $u \in (0, 1]$ and $v \in (0, 1)$ such that $\tilde{\gamma}'$ is continuous in v .
3. $\lim_{n \rightarrow \infty} \lambda_n(u) = \tilde{\lambda}(u)$ for every $u \in (0, 1)$ such that $\tilde{\lambda}(u)$ is continuous in u .
4. There exist positive constants k_n such that $\lim_{n \rightarrow \infty} k_n \gamma_n(u) = \tilde{\gamma}(u)$ for all $u \in \mathbf{I}$.

5. $\lim_{n \rightarrow \infty} \mathbf{K}_n(t) = \tilde{\mathbf{K}}(t)$ for every $t \in (0, 1)$ such that $\tilde{\mathbf{K}}$ is continuous in t .

In our case, the last claim follows directly from the pointwise convergence of \mathbf{K}_n to \mathbf{K}_C . In turn, Proposition 2 guarantees the pointwise convergence of the sequence of copulas \mathbf{C}_n to the Archimedean copula $\tilde{\mathbf{C}}$, having \mathbf{K}_C as its Kendall's distribution function. As a consequence, $\tilde{\mathbf{C}}$ may play the role as of the (Archimedean) "representative member" of the equivalence class of copulas sharing the original \mathbf{K}_C Kendall's distribution function, which obviously includes the (unknown) copula \mathbf{C} .

3.2. Global and local simulation

Besides the explicit construction of the approximating copula \mathbf{C}_n , the crucial point in applications is the possibility to generate random samples having such a dependence structure, yielding the corresponding Kendall's distribution function \mathbf{K}_n of interest here. Since \mathbf{C}_n is Archimedean, several algorithms are available: here we use a well known one, providing a global simulation.

Algorithm 1. (See Exercise 4.15 in [Nelsen, 2006])

1. Generate two independent variates s and t Uniform on $(0, 1)$.
2. Set $w = \gamma_n(\mathbf{K}_n^{-1}(t))$.
3. Set $u = \gamma_n^{-1}(s w)$ and $v = \gamma_n^{-1}((1 - s) w)$.
4. The desired pair is then (u, v) .

Instead, in order to provide "local" scenarios (e.g., to simulate on a given critical layer of interest), here we adapt an algorithm based on the Conditional Inverse Method for exchangeable Archimedean copulas (see Algorithm 3.8 in [Brechmann, 2012]).

Algorithm 2. Let $p \in (0, 1)$ be a fixed probability level.

1. Calculate Kendall's quantile $q_p = \mathbf{K}_n^{-1}(p)$ (see Definition 5).
2. Set $w = \gamma_n(q_p)$.
3. Generate a variate s Uniform on $(0, 1)$.
4. Set $u = \gamma_n^{-1}(s w)$ and $v = \gamma_n^{-1}((1 - s) w)$.
5. The desired pair is then (u, v) .

Essentially, the difference between the two procedures given above is that in Algorithm 1 two independent variates s and t are needed, whereas in Algorithm 2 only one is used: this is obvious, since in the latter case the pair (u, v) is constrained to lie on the critical layer (onedimensional isoline) $\mathcal{L}_{q_p}^{\mathbf{C}_n}$. Clearly, the realization generated by Algorithm 2 has a KRP equal to $\mu/(1 - p)$.

4. A simulation study (bivariate case)

In order to test the techniques outlined in the previous Sections we adopt a simulation approach: first, we generate random samples from known copulas, and then we check how the corresponding Kendall's distribution functions (and, hence, the estimates of the return periods) are approximated by the techniques outlined above. Below we describe the testing strategy.

4.1. Design of the testing procedure

We briefly outline here the steps of the testing procedure adopted in this work.

1. A bivariate copula \mathbf{C} (with suitable parameter θ) is chosen among five well known families: Gumbel, Frank, Clayton, Gaussian, and Cuadras-Augé. This gives the possibility to test the performance of the approximating model against different structures of dependence, whose features are briefly summarized below [Nelsen, 2006, Salvadori et al., 2007]. Note that the range of the parameters may also include the limiting cases corresponding to the 2-copulas \mathbf{W}_2 (the Fréchet-Hoeffding lower bound), Π_2 (the Product copula), or \mathbf{M}_2 (the Fréchet-Hoeffding upper bound), with Kendall's τ equal to -1 , 0 , and $+1$, respectively.

Gumbel. This family is Archimedean and Extreme Value, with Kendall's τ ranging in $[0, 1]$, is absolutely continuous, and has upper tail dependence. The expressions of the copula and the corresponding generator are as follows:

$$\mathbf{C}_\theta(u, v) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right), \tag{23}$$

with $\theta \in [1, +\infty)$, and

$$\gamma_\theta(t) = (-\ln t)^\theta. \tag{24}$$

Frank. This family is Archimedean, with Kendall's τ ranging in $[-1, 1]$, is absolutely continuous, and has no tail dependence. The expressions of the copula and the corresponding generator are as follows:

$$\mathbf{C}_\theta(u, v) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right), \tag{25}$$

with $\theta \in \mathbf{R}$, and

$$\gamma_\theta(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}. \tag{26}$$

Clayton. This family is Archimedean, with Kendall's τ ranging in $[-1, 1]$, is absolutely continuous, and has lower tail dependence. The expressions of the copula and the corresponding generator are as follows:

$$\mathbf{C}_\theta(u, v) = \left[\max\left(u^{-\theta} + v^{-\theta} - 1, 0\right)\right]^{-\frac{1}{\theta}}, \tag{27}$$

with $\theta \in [-1, +\infty)$, and

$$\gamma_\theta(t) = \frac{1}{\theta} \left(t^{-\theta} - 1\right). \tag{28}$$

Gaussian. This family is not Archimedean, with Kendall's τ ranging in $[-1, 1]$, is absolutely continuous, and has no tail dependence. The expression of the copula is as follows:

$$\mathbf{C}_\theta(u, v) = \frac{1}{2\pi\sqrt{1-\theta^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left(-\frac{s^2 - 2\theta st + t^2}{2(1-\theta^2)}\right) ds dt, \tag{29}$$

with $\theta \in [-1, +1]$, where $\Phi^{-1}(\cdot)$ denotes the inverse of the Normal distribution.

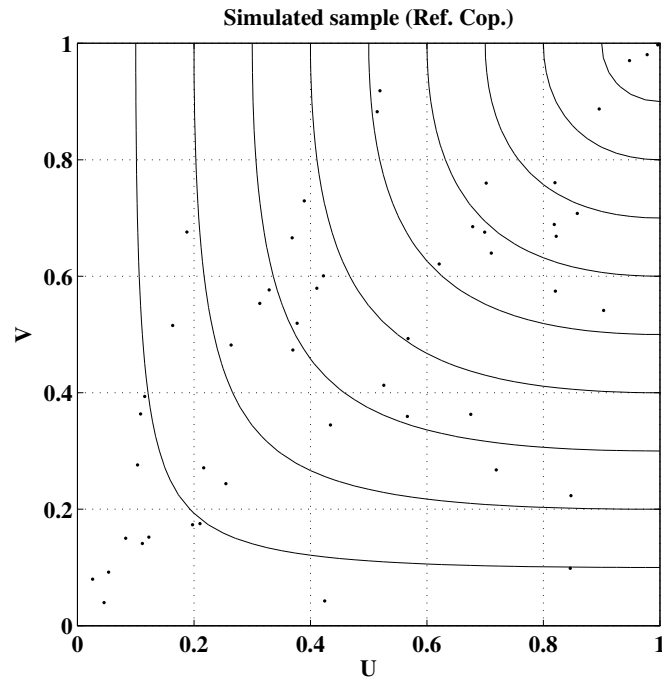


FIGURE 1. Example of a sample of size $m = 50$ extracted from the “reference” Gumbel copula — see text. Also shown are the isolines of the Gumbel copula, for the selected levels $t = 0.1, 0.2, \dots, 0.9$.

Cuadras-Augé. This family is not Archimedean, with Kendall’s τ ranging in $[0, 1]$, has both an absolutely continuous and a singular component, and has upper tail dependence. The expression of the copula is as follows:

$$\mathbf{C}_\theta(u, v) = (uv)^\theta \min\{u^{1-\theta}, v^\theta\}, \quad (30)$$

with $\theta \in [0, 1]$.

2. Three different values of the copula parameter θ are selected, in order to yield the following values of Kendall’s τ associated with \mathbf{C} : $\tau_1 = 0.25$ (for $\theta = \theta_1$), $\tau_2 = 0.5$ (for $\theta = \theta_2$), and $\tau_3 = 0.75$ (for $\theta = \theta_3$). This gives the possibility to test the robustness of the approximating procedure against different levels of association / concordance.
3. Three different orders n of the partition \mathcal{T}_n are selected: $n_1 = 3$, $n_2 = 4$, and $n_3 = 5$. Thus, \mathcal{T}_1 contains 9 points, \mathcal{T}_2 contains 17 points, and \mathcal{T}_3 contains 33 points. This gives the possibility to check the role played by the order of the partition in terms of “goodness” of approximation.
4. In order to carry out the tests, it is necessary to use a sample $\{\mathbf{X}_1, \dots, \mathbf{X}_m\}$ of bivariate observations. Three different values of the sample size m are selected: $m_1 = 50$, $m_2 = 500$, and $m_3 = 5000$. In particular, m_1 is chosen in order to represent the typical size of the samples frequently found, e.g., in hydrological or environmental applications (i.e., small sizes). Instead, m_3 should provide a statistically significant sample size, i.e. large enough

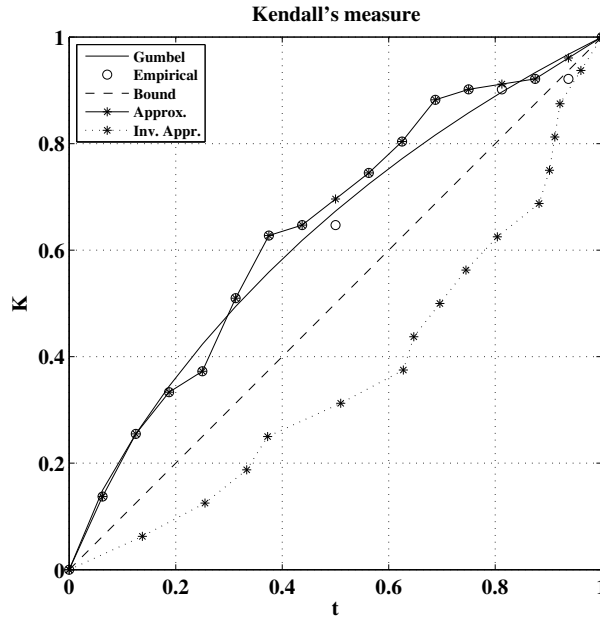


FIGURE 2. Comparison between Kendall's distribution function \mathbf{K}_C of the "reference" Gumbel copula, its empirical estimate $\hat{\mathbf{K}}$, and the corresponding approximation \mathbf{K}_n . Also shown are the lower bound \mathbf{K}_{M_2} , as well as the inverse function \mathbf{K}_n^{-1} . Here the random sample used to calculate $\hat{\mathbf{K}}$ and \mathbf{K}_n is the one shown in Figure 1, and the dyadic partition \mathcal{T}_n is of order $n = 4$.

to draw reasonable conclusions from a statistical point of view. The value chosen for m_2 represents an intermediate case between m_1 and m_3 , and is used to fully assess the behavior of the techniques proposed in this work.

5. The performance of the approach outlined in this work will be investigated via simulation techniques: thus, for each combination $(C_i, \theta_j, n_k, m_l)$, $N = 1000$ independent bivariate random samples of size m_l are extracted from the "reference" copula C_i with parameter θ_j , and analyzed by using a partition of order n_k . The value $N = 1000$ is large enough to construct reliable empirical confidence intervals for the quantities of interest here. As an illustration, in Figure 1 we show an example of (starting) sample extracted from the "reference" Gumbel copula, with parameter $\theta = 2$ and $\tau = 0.5$: hereinafter, this copula will be used to illustrate the results.

4.2. Details of the procedure

For each combination $(C_i, \theta_j, n_k, m_l)$, and for each of the N independent samples of size m_l extracted from the "reference" copula C_i with parameter θ_j , the following steps are carried out by using a partition of order n_k . Without loss of generality, here we fix the average inter-arrival time of the realizations as $\mu = 1$ year (see Section 2). Thus, the temporal unit of the Return Periods calculated in the sequel will be in years.

1. An empirical estimate $\hat{\mathbf{K}}$ of the true \mathbf{K}_C is calculated at the abscissas of the partition \mathcal{T}_n ,

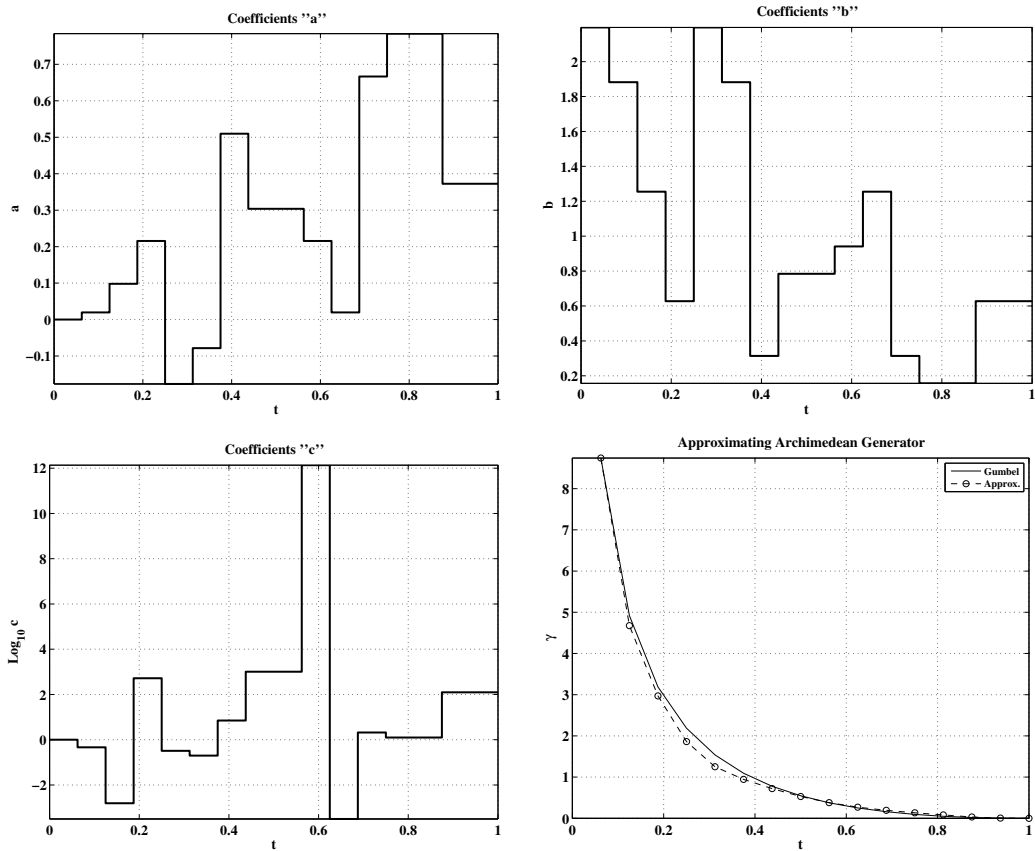


FIGURE 3. Estimates of the coefficients $a_{i,n}$'s, $b_{i,n}$'s, and $c_{i,n}$'s by using the random sample shown in Figure 1: here the order of the dyadic partition \mathcal{T}_n is $n = 4$. Also shown is the Archimedean generator γ_n associated with the approximating function \mathbf{K}_n shown in Figure 2, as well as (a version of) the one of the “reference” Gumbel copula.

generating the set of values \mathcal{Y}_n . Since the estimating procedure is not “intrinsic”, some of the estimates $\hat{\mathbf{K}}$'s may be smaller than the lower bound $\mathbf{K}_{M_2}(t) = t$ at some abscissas t_i 's. In turn, these latter are discarded, and only the consistent estimates are eventually used. As an illustration, in Figure 2 we show a comparison between Kendall's distribution function \mathbf{K}_C of the “reference” Gumbel copula, and its empirical estimate $\hat{\mathbf{K}}$. In particular, a non-admissible empirical estimate y_{15} is well visible in the right upper corner of Figure 2 at the abscissa $t_{15} = 15/16$: once discarded, the approximating function \mathbf{K}_n simply joins the closest admissible values y_{14} and y_{16} .

Note that the arbitrary dismissal of some y_i 's may yield a biased estimate of \mathbf{K}_C , whose precise features are difficult to quantify. However, on the one hand, we believe that *a priori* it makes little sense to construct a possibly non-consistent approximation \mathbf{K}_n ; on the other hand, the regularity (smoothness) of the approximating function \mathbf{K}_n generally improves by discarding the non-admissible y_i 's, as shown in Figure 2: this may yield random simulations extracted from \mathbf{C}_n less prone to show bizarre (practically unreliable) structures — see later the discussion of the features of Figure 4.

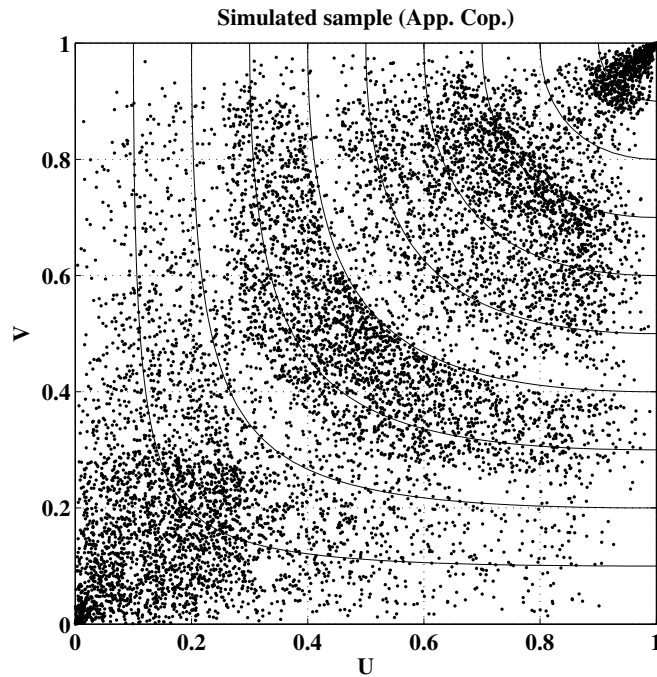


FIGURE 4. Simulated sample of size 10,000 extracted from the approximating copula \mathbf{C}_n constructed using the Archimedean generator γ_n shown in Figure 3. Also shown are the isolines of the “reference” Gumbel copula, for the selected levels $t = 0.1, 0.2, \dots, 0.9$.

2. The sequences of the relevant parameters $a_{i,n}$'s, $b_{i,n}$'s, and $c_{i,n}$'s are calculated. As an illustration, in Figure 3 we show the estimates of these coefficients, by using the sample shown in Figure 1. It is worth noting how the calculation of the $c_{i,n}$'s is a numerically ill-conditioned problem: in fact, the estimates range over quite a few orders of magnitude, with local abrupt jumps. In addition, some numerical care is needed when fixing the $b_{i,n}$'s: in fact, if $b_{i,n} \approx 1$ (within a software-dependent numerical tolerance), then the power-law term in Eq. (17) may “explode”, and the corresponding exponential expression has to be used instead — here we adopt the numerical criterion $|b_{i,n} - 1| < 0.03$. It is also interesting to note that, essentially due to a limited sample size, the empirical estimate $\hat{\mathbf{K}}$ of Kendall's distribution function \mathbf{K}_C may sometimes be constant at two (or more) successive abscissas t_i 's: this is well evident in Figure 2, at the abscissas t_7-t_8 and $t_{12}-t_{13}$, and is due to a lack of sample observations between the isolines of levels 0.4–0.5 and 0.7–0.8 shown in Figure 1. Then, in order to improve the usability of \mathbf{K}_n (see later the discussion of the features of Figure 4), these pathological values, yielding a flat approximating distribution function \mathbf{K}_n with negligible density, are discarded by simply joining the closest values y_7-y_9 and $y_{12}-y_{14}$, as shown in Figure 2.

In turn, the approximating function \mathbf{K}_n is constructed, yielding the generator γ_n and the corresponding Archimedean copula \mathbf{C}_n via Eq. (22). As an illustration, in Figure 2 we show a comparison between Kendall's distribution function \mathbf{K}_C of the “reference” Gumbel

copula, and the corresponding approximation \mathbf{K}_n : overall, the interpolation looks empirically acceptable. Also, in Figure 3 we show the approximating Archimedean generator γ_n associated with \mathbf{K}_n , as well as (a version of) the one of the “reference” Gumbel copula: a direct comparison is meaningless, since Archimedean generators are uniquely defined up to a multiplicative constant.

3. The KRP’s $T = 10, 20, 50, 100, 200, 500, 1000$ years, of interest in practical applications, are selected, and the corresponding probability levels $p = 1 - \mu/T$ and Kendall’s quantiles q_p ’s are calculated (by inverting Eq. (11)). Then, by inverting the N approximating functions \mathbf{K}_n ’s (calculated over all the N simulations mentioned above), estimates of the mean values of Kendall’s quantiles q_p ’s are computed, as well as suitable corresponding empirical Confidence Intervals. In fact, here the target is to investigate how good (or bad) the approach outlined in this work may approximate the “reference” Kendall’s distributions. In particular, we check to what extent the approximation is correct (i.e., unbiased).

As an illustration, considering the case of the “reference” Gumbel copula, in Figure 5 we plot the estimates of the mean values of Kendall’s quantiles q_p ’s, and approximate 90% Confidence Intervals, for all the values of the parameters m and n used in this work. The estimates should be compared with the exact values of Kendall’s quantiles calculated by using the “reference” Kendall’s distributions plotted in the same pictures.

4. In order to test whether the approximating dependence structure \mathbf{C}_n may generate samples having the required KRP’s, we calculate the percentages of simulated pairs “below” each critical layer $\mathcal{L}_{q_p}^{\mathbf{C}_n}$ (see Definition 2, and also the discussion ensuing Definition 5) for the KRP’s $T = 10, 20, 50, 100, 200, 500, 1000$ years mentioned above, and compare these estimates with the expected values given by $p = 1 - \mu/T$.

For this purpose, a large simulation of size 10,000 is generated from the approximating copula \mathbf{C}_n via Algorithm 1: see Figure 4 for an illustration. It is interesting to note the very particular structure of the simulated pairs, as well as the link with the corresponding approximating function \mathbf{K}_n shown in Figure 2. Evidently, the density of points in the unit square is proportional to the density of \mathbf{K}_n given by $\mathbf{K}'_n(t) = d\mathbf{K}_n(t)/dt = b_{i,n}$, for some suitable index i depending upon t . The plot of the coefficients $b_{i,n}$ ’s shown in Figure 3 provides the graph of \mathbf{K}'_n : in fact, the $b_{i,n}$ ’s are simply the slopes of the linear segments forming \mathbf{K}_n . In turn, intervals in \mathbf{I} where \mathbf{K}_n is particularly flat (or, equivalently, \mathbf{K}'_n is small) correspond to “circular strips” in \mathbf{I}^2 where only a few pairs can be generated, coherently with the probabilistic meaning of Kendall’s distribution function (see Eq. (3)): a small value of $b_{i,n}$ in the i -th interval means that it is unlikely that random realizations could be simulated in the corresponding circular strip.

4.3. Analysis of the results

In this Section we briefly discuss the results of the tests mentioned in the previous Sections. For three of the families of copulas considered here (namely, the Gumbel, the Gaussian, and the Cuadras-Augé), in Tables 1–3 below we report the quantities

$$\Delta = 100 \cdot \frac{\bar{q}_p - q_p}{q_p}, \quad (31)$$

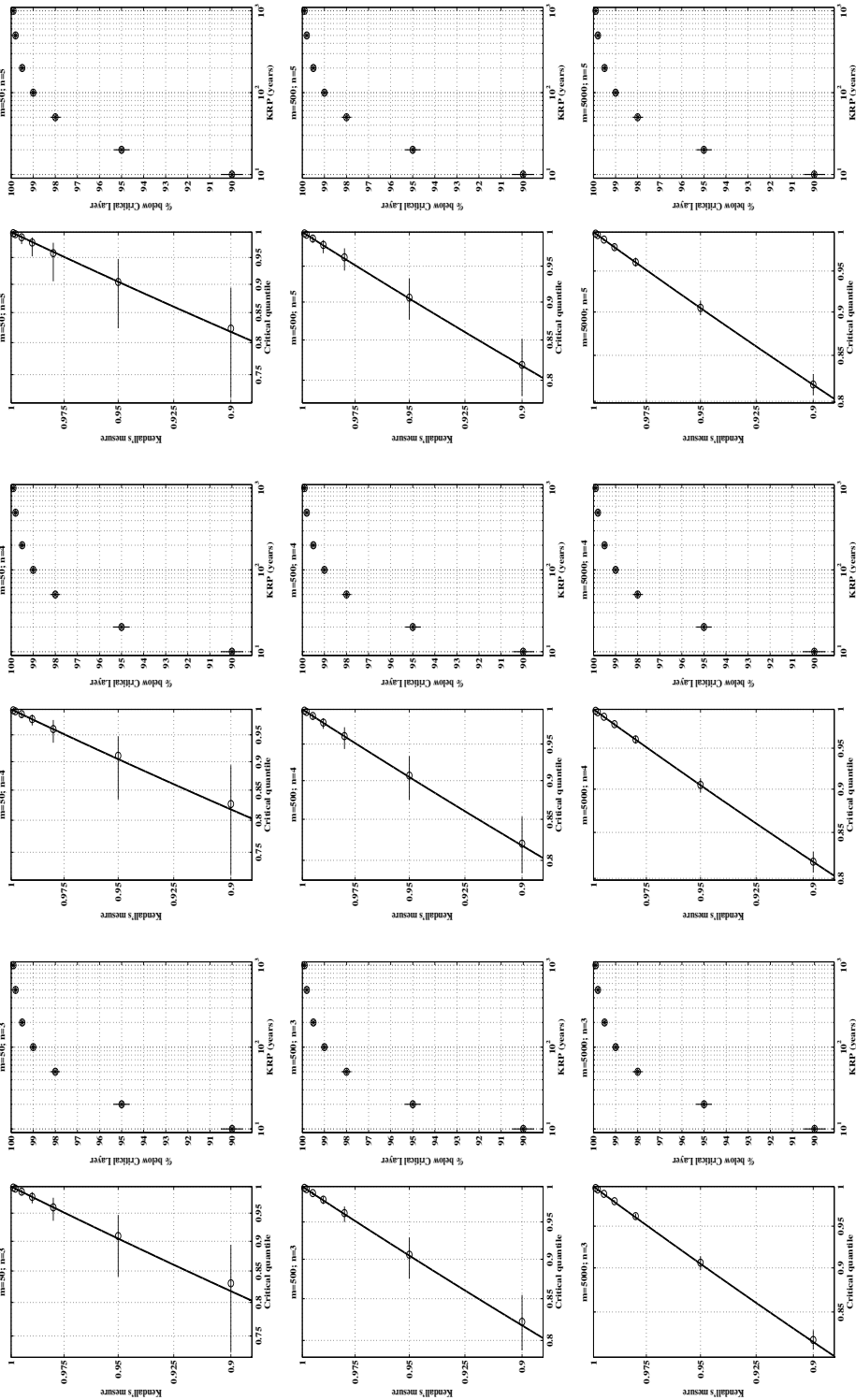


FIGURE 5. Estimates of the mean values of Kendall's quantiles q_p 's, and probabilities of the sub-critical regions, for all the values of the parameters m and n used in this work: here the case of the "reference" Gumbel copula is considered. (Odd columns) Mean values of Kendall's quantiles q_p 's (markers) — see also Table 1: the approximate horizontal Confidence Intervals have level 90%. (Even columns) Exact probabilities of the sub-critical regions, for the KRP's $T = 10, 20, 50, 100, 200, 500, 1000$ years (circles), and corresponding estimates (asterisks): the approximate vertical Confidence Intervals have level 90%.

TABLE 1. Estimates of the relative percent biases Δ 's for the Gumbel family: shown are the percentages “%”.

KRP	θ	m	n			τ
			3	4	5	
10 y	$\frac{4}{3}$	50	4.147	2.695	2.477	$\frac{1}{4}$
		500	0.597	0.407	0.256	
		5000	0.298	0.105	0.044	
	2	50	1.523	1.093	0.683	$\frac{1}{2}$
		500	0.529	0.289	0.149	
		5000	0.276	0.038	0.0523	
	4	50	-0.225	-0.755	-1.236	$\frac{3}{4}$
		500	0.082	0.098	-0.119	
		5000	0.069	0.021	0.020	
100 y	$\frac{4}{3}$	50	0.659	0.702	0.583	$\frac{1}{4}$
		500	0.404	0.255	0.269	
		5000	0.457	0.141	0.083	
	2	50	0.012	0.030	-0.109	$\frac{1}{2}$
		500	0.095	0.038	0.071	
		5000	0.103	0.035	0.011	
	4	50	-0.095	-0.208	-0.297	$\frac{3}{4}$
		500	0.014	-0.003	-0.064	
		5000	0.025	0.005	-0.002	
1000 y	$\frac{4}{3}$	50	0.085	0.089	0.078	$\frac{1}{4}$
		500	0.060	0.041	0.015	
		5000	0.066	0.035	0.019	
	2	50	0.003	0.005	-0.009	$\frac{1}{2}$
		500	0.011	0.006	0.004	
		5000	0.012	0.006	0.003	
	4	50	-0.006	-0.017	-0.026	$\frac{3}{4}$
		500	0.005	0.003	-0.003	
		5000	0.006	0.004	0.003	

where q_p is the exact Kendall's quantile of order p , and \bar{q}_p the (average) estimated one. Thus, the Δ 's represent the percent relative biases, which provide sensible measures to study the goodness of the approximating approach proposed in this work. For the sake of shortness, the results concerning the Frank and the Clayton Archimedean families are not reported, being similar to the Gumbel case. Also, only three KRP's are shown: namely, 10, 100, and 1000 years (for instance, in hydrology, these are the return periods frequently used to design structures like, respectively, sewers, harbors, and dams).

The analysis of Tables 1–3 yields the following considerations. In general, the percent relative biases Δ 's are quite small: these are always smaller than 5%, and smaller than 1% (if not negligible) in almost all cases. In turn, the approximating procedure outlined in this work seems to provide valuable results concerning the interpolation of the true (but unknown) \mathbf{K}_C . Below we discuss how the parameters of interest may affect the quality of the approximations.

The partition order. Apparently, the Δ 's get smaller and smaller by increasing n when the bias is positive (i.e., in case of over-estimation), whereas they increase, in an absolute sense, when the bias is negative (i.e., in case of under-estimation). Shortly, finer partitions generally reduce the amount of over-estimation of Kendall's quantiles.

TABLE 2. Estimates of the relative percent biases Δ 's for the Gaussian family: shown are the percentages “%”.

KRP	θ	m	n			τ
			3	4	5	
10 y	0.38268	50	3.807	2.619	2.204	$\frac{1}{4}$
		500	1.046	0.542	0.519	
		5000	0.713	0.145	0.108	
	0.70711	50	2.593	1.952	0.955	$\frac{1}{2}$
		500	0.647	0.388	0.108	
		5000	0.453	0.140	0.047	
	0.92388	50	0.373	0.014	-0.495	$\frac{3}{4}$
		500	0.271	0.212	0.011	
		5000	0.189	0.118	0.082	
100 y	0.38268	50	3.342	2.829	2.565	$\frac{1}{4}$
		500	2.110	0.970	0.736	
		5000	2.068	0.264	0.179	
	0.70711	50	1.185	1.073	0.805	$\frac{1}{2}$
		500	1.109	0.533	0.334	
		5000	1.102	0.501	0.085	
	0.92388	50	0.274	0.135	0.051	$\frac{3}{4}$
		500	0.352	0.223	0.118	
		5000	0.365	0.231	0.101	
1000 y	0.38268	50	1.182	1.133	1.109	$\frac{1}{4}$
		500	1.065	0.840	0.707	
		5000	1.062	0.818	0.476	
	0.70711	50	0.323	0.312	0.286	$\frac{1}{2}$
		500	0.315	0.258	0.190	
		5000	0.314	0.256	0.194	
	0.92388	50	0.069	0.056	0.046	$\frac{3}{4}$
		500	0.077	0.065	0.052	
		5000	0.079	0.065	0.052	

The sample size. Apparently, the Δ 's get smaller and smaller by increasing m . Indeed, this has to be expected: a larger sample size usually provides a better estimate of the true (but unknown) \mathbf{K}_C .

The Kendall's τ . Apparently, the Δ 's always get smaller and smaller by increasing τ . Indeed, this has to be expected: in fact, intuitively speaking, a larger concordance means that the dependence structure described by \mathbf{C} is “closer” to the one of the co-monotone copula \mathbf{M}_2 , independently of the family of \mathbf{C} . In turn, samples extracted from \mathbf{C} tend to arrange along the main diagonal in the unit square, and the spread decreases by increasing τ . Then, both \mathbf{K}_C and \mathbf{K}_n tend to the common lower bound given by the main diagonal in the unit square, thus reducing the biases of the estimates.

The copula family. As long as both the sample size m and the partition order n are large enough, the role played by the copula family seems to be more and more irrelevant. In fact, apparently, in the limiting case, the magnitudes of the Δ 's are practically the same in all the cases illustrated here.

Finally, considering the case of the “reference” Gumbel copula, in Figure 5 we show the comparisons between the exact probabilities of the sub-critical regions (i.e., the ones “below” the

TABLE 3. Estimates of the relative percent biases Δ 's for the Cuadras-Augé family: shown are the percentages “%”.

KRP	θ	m	n			τ
			3	4	5	
10 y	$\frac{2}{5}$	50	3.845	3.080	0.991	$\frac{1}{4}$
		500	0.864	0.465	0.591	
		5000	0.362	0.173	0.068	
	$\frac{2}{3}$	50	1.891	0.876	0.246	$\frac{1}{2}$
		500	0.536	0.405	0.182	
		5000	0.291	0.046	0.074	
	$\frac{6}{7}$	50	-0.421	-0.693	-1.009	$\frac{3}{4}$
		500	0.048	0.068	-0.071	
		5000	0.035	0.033	0.037	
100 y	$\frac{2}{5}$	50	0.650	0.618	0.274	$\frac{1}{4}$
		500	0.483	0.159	0.474	
		5000	0.459	0.138	0.084	
	$\frac{2}{3}$	50	0.080	0.006	-0.147	$\frac{1}{2}$
		500	0.112	0.048	0.045	
		5000	0.108	0.046	0.015	
	$\frac{6}{7}$	50	-0.094	-0.164	-0.255	$\frac{3}{4}$
		500	0.012	-0.001	-0.054	
		5000	0.027	0.013	0.006	
1000 y	$\frac{2}{5}$	50	0.080	0.077	0.043	$\frac{1}{4}$
		500	0.064	0.027	0.033	
		5000	0.061	0.030	0.014	
	$\frac{2}{3}$	50	0.014	0.007	-0.009	$\frac{1}{2}$
		500	0.017	0.011	0.007	
		5000	0.017	0.011	0.008	
	$\frac{6}{7}$	50	-0.007	-0.013	-0.022	$\frac{3}{4}$
		500	0.004	0.003	-0.003	
		5000	0.006	0.004	0.003	

critical layers), for the KRP's $T = 10, 20, 50, 100, 200, 500, 1000$ years, and the corresponding empirical estimates. Evidently, the agreement is valuable in all cases: in particular, the estimated probabilities are practically the same as of the theoretical expected ones. Similar results are found by using the other four families of copulas investigated here (namely, the Frank, Clayton, Gaussian, and Cuadras-Augé families): however, some of these structures exhibit some (or all) kinds of Archimedean dependence (e.g., exchangeability, or, at least, a symmetric dependence structure).

5. Conclusions and perspectives

In this work we outline a constructive approach for the approximation of Kendall's measure and Kendall's Return Period in the bivariate case. First, we introduce a suitable theoretical framework, based on the Theory of Copulas, where to embed the issue. Then, we outline the original construction procedure. The whole approach is semi-parametric, since the approximation of Kendall's measure is only based on the available data, and is simply given by a (suitable) continuous piece-wise linear function on the unit interval. In turn, the Archimedean copula associated with the approximating Kendall's measure is worked out, and used for the simulation

of random samples featuring suitable properties. Finally, a sensitivity analysis is carried out via a simulation study, in order to investigate the robustness of the approach proposed against several relevant factors: namely, the parent family of copulas, the level of association / concordance, the sampling density, and the size of the available sample. As a conclusion, the approximation strategy outlined in this work may yield reasonable and valuable results in all the cases investigated here.

Below, we discuss future perspectives, as well as possible improvements and generalizations, of the approach sketched in the previous Sections, which may find fruitful applications to practical problems.

Generalization to the multivariate case. In this work we limited our attention to the bivariate case. It would be interesting to check whether a similar approach can be adopted in a more general d -variate case, with $d > 2$. The starting point is the general result given by Proposition 4.5 in [McNeil and Nešlehová, 2009], which provides the expression of the Kendall's distribution function of an Archimedean d -copula \mathbf{C} . Now, as in Section 3.1, the idea is to approximate the true $\mathbf{K}_{\mathbf{C}}$ with a continuous, piece-wise linear, distribution function \mathbf{K}_n , and then solve a suitable differential equation in order to calculate the approximating Archimedean generator. However, for dimensions $d > 2$, numerical estimates of the high-order derivatives involved may be needed, and this can make the procedure ill-conditioned (to be compared with [Hofert et al., 2012]).

A “better” partition. As already mentioned in Section 4.3, the dyadic partition may not always be the best choice: for instance, a non-uniform partition could provide better information about the left (or right) tail of the Kendall's distribution, which may be of interest in practical applications. The construction of suitable “optimal” partitions, devised for specific purposes and targets, is in progress.

Comparison of parametric vs. non-parametric approaches. A further test of interest would be the comparison between the results achieved by the present approach with those obtained exploiting known parametric models. For instance, in view of a possible better assessment of hydrological structures' safety (e.g., dams, harbors, dikes, and so on), the re-analysis of several publications already present in Literature might yield an improvement of the design values previously estimated (see, e.g., [Salvadori et al., 2011, Gräler et al., 2013]).

References

- [AghaKouchak et al., 2013] AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., and Sorooshian, S., editors (2013). *Extremes in a Changing Climate*, volume 65 of *Water Science and Technology Library*. Springer Science & Business Media, Dordrecht. ISBN 978-94-007-4479-0.
- [Barbe et al., 1996] Barbe, P., Genest, C., Ghouli, K., and Rémillard, B. (1996). On Kendall's process. *J. Multivariate Anal.*, 58(2):197–229.
- [Belzunce et al., 2007] 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(10):5112—5129.
- [Brechmann, 2012] Brechmann, E. C. (2012). Hierarchical Kendall copulas: Properties and inference. *arXiv:1202.1998v2*.
- [Bücher et al., 2012] Bücher, A., Dette, H., and Volgushev, S. (2012). A test for Archimedeanity in bivariate copula models. *J. Multivariate Anal.*, 110:121–132.
- [Chaouch and Goga, 2010] Chaouch, M. and Goga, C. (2010). Design-based estimation for geometric quantiles with application to outlier detection. *Computational Statistics & Data Analysis*, 54(10):2214–2229.

- [Charpentier and Segers, 2008] Charpentier, A. and Segers, J. (2008). Convergence of Archimedean copulas. *Statist. Probab. Lett.*, 78(412–419).
- [Chebana and Ouarda, 2009] Chebana, F. and Ouarda, T. B. M. J. (2009). Index flood-based multivariate regional frequency analysis. *Water Resour. Res.*, 45:W10435.
- [Chebana and Ouarda, 2011] Chebana, F. and Ouarda, T. B. M. J. (2011). Multivariate quantiles in hydrological frequency analysis. *Environmetrics*, 22(1):63–78.
- [De Michele et al., 2007] De Michele, C., Salvadori, G., Passoni, G., and Vezzoli, R. (2007). A multivariate model of sea storms using copulas. *Coastal Engineering*, 54:734–751.
- [Durante et al., 2012] Durante, F., Fernández-Sánchez, J., and Sempi, C. (2012). Sklar’s Theorem obtained via regularization techniques. *Nonlinear Anal.*, 75(2):769–774.
- [Durante and Salvadori, 2010] Durante, F. and Salvadori, G. (2010). On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21:143–161.
- [Durante and Sempi, 2010] Durante, F. and Sempi, C. (2010). Copula theory: an introduction. In Jaworki, P., Durante, F., Härdle, W., and Rychlik, T., editors, *Copula Theory and its Applications*, volume 198 of *Lecture Notes in Statistics - Proceedings*, pages 3–31. Springer, Berlin Heidelberg.
- [Genest et al., 2012] Genest, C., Nešlehová, J., and Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. *Ann. Inst. Statist. Math.*, 4:811–834.
- [Genest et al., 2011a] Genest, C., Nešlehová, J., and Ben Ghorbal, N. (2011a). Estimators based on Kendall’s tau in multivariate copula models. *Australian and New Zealand Journal of Statistics*, 53:157–177.
- [Genest et al., 2011b] Genest, C., Nešlehová, J., and Ziegel, J. (2011b). Inference in multivariate Archimedean copula models (with discussion). *Test*, 20:223–292.
- [Genest and Rivest, 1993] Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.*, 88(423):1034–1043.
- [Genest and Rivest, 2001] Genest, C. and Rivest, L.-P. (2001). On the multivariate probability integral transformation. *Statist. Probab. Lett.*, 53(4):391–399.
- [Ghoudi et al., 1998] Ghoudi, K., Khoudraji, A., and Rivest, L. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.*, 26:187–197.
- [Gräler et al., 2013] 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. *Hydrol. Earth Syst. Sci.*, 17:1281–1296.
- [Hofert et al., 2012] Hofert, M., Mächler, M., and McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *J. Multivariate Anal.*, 110:133–150.
- [Jaworski, 2010] Jaworski, P. (2010). Testing Archimedeanity. In Borgelt, C., González-Rodríguez, G., Trutschnig, W., Lubiano, M., Gil, M., Grzegorzewski, P., and Hryniewicz, O., editors, *Combining Soft Computing and Statistical Methods in Data Analysis*, volume 77 of *Advances in Intelligent and Soft Computing*, pages 353–360. Springer, Berlin.
- [Jaworski et al., 2010] Jaworski, P., Durante, F., Härdle, W., and Rychlik, T., editors (2010). *Copula Theory and its Applications*, volume 198 of *Lecture Notes in Statistics - Proceedings*. Springer, Berlin Heidelberg.
- [Joe, 1997] Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall, London.
- [Kojadinovic and Yan, 2012] Kojadinovic, I. and Yan, J. (2012). A nonparametric test of exchangeability for extreme-value and left-tail decreasing bivariate copulas. *The Scandinavian Journal of Statistics*, 39(3):480–496.
- [McNeil and Nešlehová, 2009] McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d -monotone functions and ℓ^1 -norm symmetric distributions. *The Annals of Statistics*, 37:3059–3097.
- [Nappo and Spizzichino, 2009] Nappo, G. and Spizzichino, F. (2009). Kendall distributions and level sets in bivariate exchangeable survival models. *Information Sciences*, 179(17):2878–2890.
- [Nelsen, 2006] Nelsen, R. (2006). *An introduction to copulas*. Springer-Verlag, New York, second edition.
- [Nelsen et al., 2001] Nelsen, R. B., Quesada-Molina, J. J., Rodríguez-Lallena, J. A., and Úbeda-Flores, M. (2001). Distribution functions of copulas: a class of bivariate probability integral transforms. *Statist. Probab. Lett.*, 54(3):277–282.
- [Nelsen et al., 2003] Nelsen, R. B., Quesada-Molina, J. J., Rodríguez-Lallena, J. A., and Úbeda-Flores, M. (2003). Kendall distribution functions. *Statist. Probab. Lett.*, 65(3):263–268.

- [Nelsen et al., 2009] Nelsen, R. B., Quesada-Molina, J. J., Rodríguez-Lallena, J. A., and Úbeda-Flores, M. (2009). Kendall distribution functions and associative copulas. *Fuzzy Sets and Systems*, 160(1):52–57.
- [Salvadori, 2004] Salvadori, G. (2004). Bivariate return periods via 2-copulas. *Statist. Methodol.*, 1:129–144.
- [Salvadori and De Michele, 2004] Salvadori, G. and De Michele, C. (2004). Frequency analysis via Copulas: theoretical aspects and applications to hydrological events. *Water Resour. Res.*, 40:W12511.
- [Salvadori and De Michele, 2010] Salvadori, G. and De Michele, C. (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resour. Res.*, 46:W10501.
- [Salvadori et al., 2011] Salvadori, G., De Michele, C., and Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrol. Earth Syst. Sci.*, 15:3293–3305.
- [Salvadori et al., 2007] Salvadori, G., De Michele, C., Kottegoda, N., and Rosso, R. (2007). *Extremes in nature. An approach using copulas*, volume 56 of *Water Science and Technology Library*. Springer, Dordrecht.
- [Serfling, 2002] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statist. Neerlandica*, 56(2):214–232.
- [Singh et al., 2007] Singh, V., Jain, S., and Tyagi, A. (2007). *Risk and Reliability Analysis*. ASCE Press, Reston, Virginia.
- [Sklar, 1959] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.
- [Vandenberghe et al., 2010] Vandenberghe, S., Verhoest, N. E. C., Buyse, E., and Baets, B. D. (2010). A stochastic design rainfall generator based on copulas and mass curves. *Hydrol. Earth Syst. Sci.*, 14:2429–2442.