

# On the return period and design in a multivariate framework

G. Salvadori<sup>1</sup>, C. De Michele<sup>2</sup>, and F. Durante<sup>3</sup>

<sup>1</sup>Dipartimento di Matematica, Università del Salento, Provinciale Lecce-Arnesano, P.O. Box 193, 73100 Lecce, Italy

<sup>2</sup>DIAR (Sezione CIMI), Politecnico di Milano, Piazza Leonardo da Vinci 32, 20132 Milano, Italy

<sup>3</sup>School of Economics and Management, Free University of Bozen-Bolzano, 39100 Bolzano, Italy

Received: 18 May 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 10 June 2011

Revised: 5 October 2011 – Accepted: 24 October 2011 – Published: 4 November 2011

**Abstract.** Calculating return periods and design quantiles in a multivariate environment is a difficult problem: this paper tries to make the issue clear. First, we outline a possible way to introduce a consistent theoretical framework for the calculation of the return period in a multi-dimensional environment, based on Copulas and the Kendall's measure. Secondly, we introduce several approaches for the identification of suitable design events: these latter quantities are of utmost importance in practical applications, but their calculation is yet limited, due to the lack of an adequate theoretical environment where to embed the problem. Throughout the paper, a case study involving the behavior of a dam is used to illustrate the new concepts outlined in this work.

## 1 Introduction

The notion of Return Period (hereinafter, RP) is frequently used in hydrology (as well as in water resources and civil engineering, and more generally in geophysical and environmental sciences) for the identification of dangerous events, and provides a means for rational decision making (for a review, see Singh et al., 2007, and references therein).

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. For example, the RP of a dam design quantile is usually 1000 years or more (Midttomme et al., 2001), while for a sewer it is about 5–10 years (Briere, 1999).

While in the univariate case the design quantile is usually identified without ambiguity – and widely used in the engineering practice (Chow et al., 1988) – in the multivariate one this is not so. Indeed, the identification problem of design events in a multivariate context is of fundamental importance, but of troublesome nature. Recently, several efforts have been spent on the issues of multivariate design and quantiles (see, e.g. Serfling, 2002; Belzunce et al., 2007; Chebana and Ouarda, 2009, 2011b; Chaouch and Goga, 2010, and references therein; for a methodology to identify multivariate extremes by using depth functions see Chebana and Ouarda, 2011a). Here we address the following crucial question: “How is it possible to calculate the critical design event(s) in the multivariate case?” Below, we outline a suitable approach in order to provide consistent answers.

As we shall show later, the calculation of the RP is strictly related to the notion of Copula. The use of copulas in environmental sciences is recent and rapidly growing. Shortly, a multivariate copula  $C$  is a joint distribution on  $I^d = [0, 1]^d$  with Uniform margins. The link between a multivariate distribution  $F$  and the associated  $d$ -dimensional copula  $C$  is given by the functional identity stated by Sklar's Theorem (Sklar, 1959):

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

for all  $\mathbf{x} \in R^d$ , where the  $F_i$ 's are the univariate margins of  $F$ . If all the  $F_i$ 's are continuous, then  $C$  is unique. Most importantly, the  $F_i$ 's in Eq. (1) only play the role of (geometrically) re-mapping the probabilities induced by  $C$  on the subsets of  $I^d$  onto suitable subsets of  $R^d$ , without changing their values: viz., the dependence structure modeled by  $C$  plays a central role in tuning the probabilities of joint occurrences. In fact, under weak regularity conditions, any point  $\mathbf{x} \in R^d$  can be uniquely re-mapped onto  $\mathbf{u} \in I^d$  (and vice-versa) via the Probability Integral Transform:

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



Correspondence to: G. Salvadori  
(gianfausto.salvadori@unisalento.it)

For a thorough theoretical introduction to copulas see Joe (1997); Nelsen (2006); for a practical approach see Salvadori et al. (2007); Jaworski et al. (2010). In order to avoid troublesome situations, hereinafter we shall assume that  $F$  is continuous (but not necessarily absolutely continuous), and strictly increasing in each marginal: these regularity constraints are rather weak, and satisfied by the majority of the distributions used in applications. Clearly, also pathological cases can be carried out, but they require suitable techniques that go beyond the scope of this work.

Later we shall use the *Kendall's distribution* (or *measure*) function  $K_C : I \rightarrow I$  (Genest and Rivest, 1993, 2001) given by

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

where  $t \in I$  is a probability level,  $W = \mathbf{C}(U_1, \dots, U_d)$  is a univariate random variable (hereinafter, *r.v.*) taking value on  $I$ , and the  $U_i$ 's are Uniform *r.v.s* on  $I$  with copula  $\mathbf{C}$ . Note that Eq. (3) practically measures the probability that a random event will appear in the region of  $I^d$  defined by the inequality  $\mathbf{C}(\mathbf{u}) \leq t$  – see also Genest and Rivest (2001); Nappo and Spizzichino (2009). Thus, as we shall see,  $K_C$  turns out to be a fundamental tool for calculating a copula-based RP for multivariate events.

Unfortunately, at present no general analytical expressions of  $K_C$  are known – except for special cases, like the one of bivariate Extreme Value copulas (Ghoudi et al., 1998), and some Archimedean copulas (McNeil and Nešlehová, 2009) – and it is necessary to resort to simulations (see, e.g. Algorithm 1 outlined later).

The paper is organized as follows. In Sect. 2 we first illustrate the case study. In Sect. 3 we reconsider a previously introduced notion of RP in a multivariate environment, and compare it with other approaches. In Sect. 4 we show how to calculate the corresponding quantile. Then, in Sect. 5 we present two strategies to calculate critical design events in a multivariate context. Finally, in Sect. 6 we discuss the results outlined in the paper, and draw some conclusions.

## 2 The case study

Although this work is of methodological nature, we feel important to illustrate with practical examples the new concepts introduced. For this reason, we first present the case study that will be used throughout the paper.

The data are collected at the Ceppo Morelli dam, and are essentially the same as those investigated in De Michele et al. (2005), to which we make reference for further details. The dam, completed in 1929, is located in the valley of Anza catchment, a sub-basin of the Toce river (Northern Italy), and was built to produce hydroelectric energy. The dam is characterized by a small water storage of about  $0.47 \times 10^6 \text{ m}^3$ . The minimum level of regulation is 774.75 m a.s.l., while the maximum is 780.75 m a.s.l. The

maximum water level is at 782.5 m a.s.l., and the dam crest level is at 784 m a.s.l. The dam has an uncontrolled spillway (84 m long) at 780.75 m a.s.l., and also intermediate and bottom outlets (the latter ones are obstructed by river sediments).

In De Michele et al. (2005), “undisturbed” flood hydrographs incoming the reservoir were fixed by using the inverse reservoir routing, the water levels in the reservoir, and the operations on the controlled outlets. Then, maximum annual flood peaks  $Q$  and volumes  $V$  were identified and selected for 49 years, from 1937 to 1994. As a result of a thorough investigation, almost all of the occurrence dates of the  $Q$ 's and the  $V$ 's were the same: i.e. they happened during the same flood event.

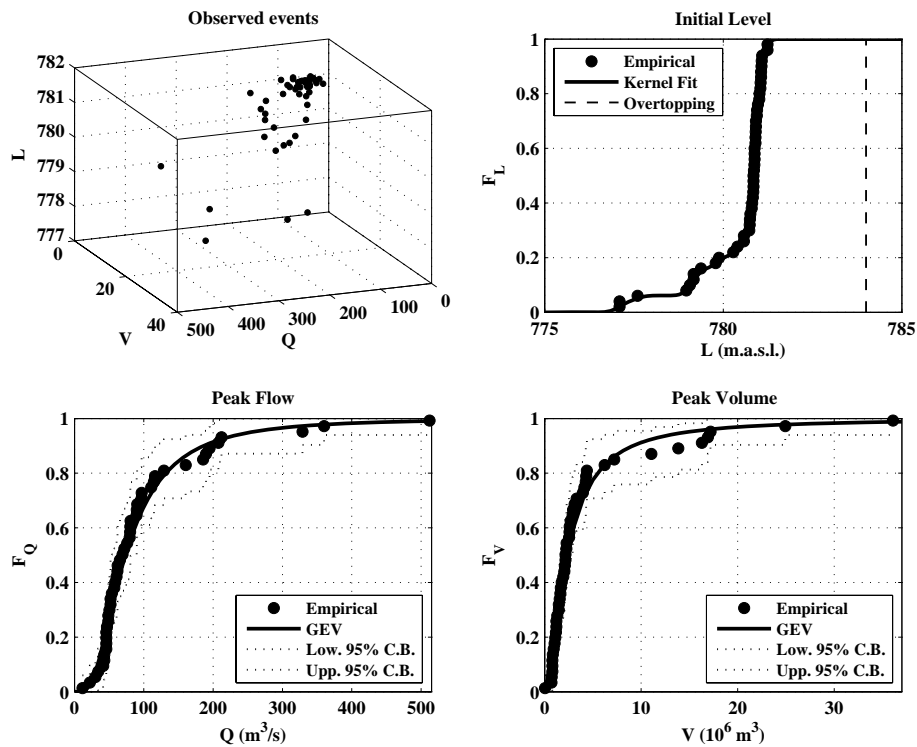
As an improvement over De Michele et al. (2005), beyond the pair  $(Q, V)$ , also the initial water level  $L$  in the reservoir before the flood event  $(Q, V)$  is considered in this work, in order to analyse the triplet  $(Q, V, L)$  of practical interest: in fact, on the one hand  $L$  represents the starting state of the dam; on the other hand,  $(Q, V)$  is the hydrologic “forcing” to the structure. Clearly, there are physical reasons to assume that  $L$  is independent of  $(Q, V)$  – see also below. The sample mean of  $L$  is about 780.44 m a.s.l., with a sample standard deviation of about 1 m. The small variability of  $L$  with respect to its range (here [774.75, 780.75] is the regulation range), is mainly due to the management policy of the reservoir: the target of the dam manager is to keep a high water level, in order to get the maximum benefit concerning the production of electric energy.

Using the pair  $(Q, V)$ , it is possible to calculate the associated flood hydrograph with peak  $Q$  and volume  $V$ , once the shape of the hydrograph has been chosen. As first approximation, it is possible to consider a triangular shape, where the base time is equal to  $T_b = 2V/Q$ , the time of rise equals  $T_r = T_b/2.67$ , and the time of recession is equal to  $1.67 T_r$ , (see Soil Conservation Service, 1972 and Chow et al., 1988, p.229 – for a different approach see Serinaldi and Grimaldi, 2011). Consequently, the flood hydrograph  $q$  is given by

$$q(t) = \begin{cases} 1.335 t Q^2/V, & 0 \leq t \leq T_r \\ 1.6 Q - 0.8 t Q^2/V, & T_r \leq t \leq T_b \end{cases}$$

Later, in Sect. 5, we shall test the behavior of the dam subject to selected hydrographs. More particularly, we shall first operate the reservoir routing of the flood hydrograph (see, e.g. Bras, 1990, p.475–478 and Zoppou, 1999) considering as outlet only the uncontrolled spillway, and then we shall check whether or not the reservoir level exceeds the crest level of the dam.

In Fig. 1 we show the trivariate plot of the available observations, as well as the fits of the marginal distributions. However, we shall not insist on this point, being of secondary importance with respect to the actual methodological target of the paper. The GEV law is used to model the statistics of both  $Q$  and  $V$ , since these are annual maxima:



**Fig. 1.** Trivariate plot of the available ( $Q, V, L$ ) observations, and fits of the marginal distributions – see text.

the estimates of the parameters are reported in Table 1. The fits are valuable, as they passed standard goodness-of-fit tests (namely, Kolmogorov-Smirnov and Anderson-Darling – see, e.g. Kottegoda and Rosso, 1997) at all usual levels (viz., 1 %, 5 %, and 10 %). Instead, the behavior of the variable  $L$  is quite tricky (as explained above, the water level is arbitrarily fixed by the dam manager): for this reason, its law is calculated via a non-parametric Normal Kernel estimation (Bowman and Azzalini, 1997). As a result, also in this case the Kolmogorov-Smirnov test is passed at all usual levels.

The trivariate plot of the observations, as shown in Fig. 1, is the first step usually carried out by practitioners to investigate the multivariate behavior of the phenomenon. However, we want to stress that this type of graph only provides partial information, and should not be used to draw rough conclusions about the dependence structure of ( $Q, V, L$ ) – see below, and also Genest and Favre (2007) for a thorough review.

In order to investigate the joint behavior of the variables ( $Q, V, L$ ), as is typical in copula analysis, we shall use the normalized ranks to carry out a non-parametric study. The trivariate rank-plot shown in Fig. 2 provides some rough indications about the global dependence structure (i.e. the copula) linking the three variables ( $Q, V, L$ ).

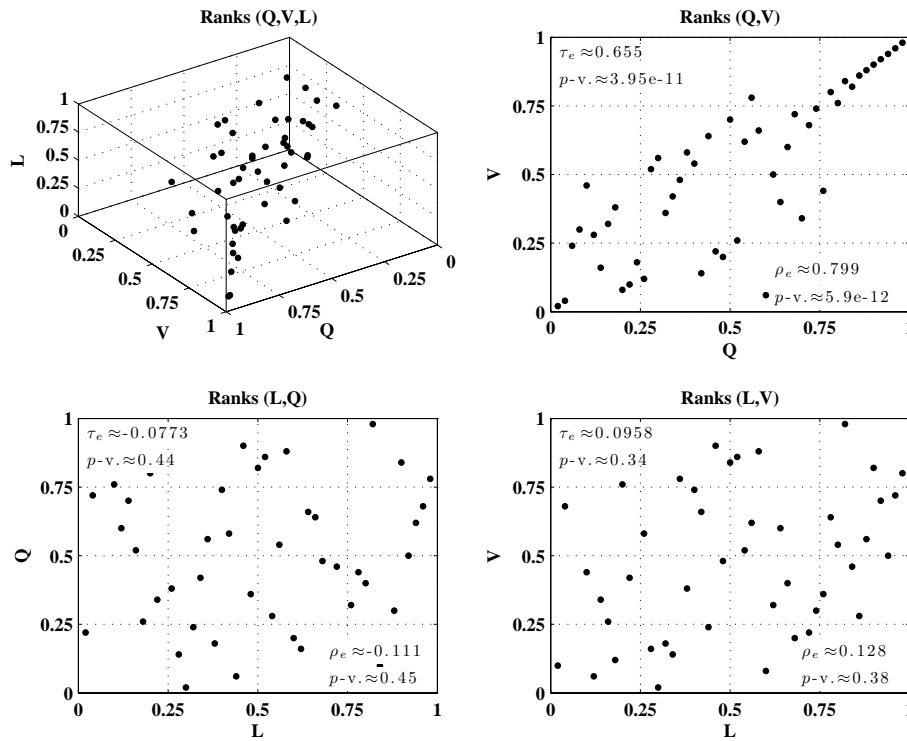
As already mentioned above, there are physical reasons to assume that  $L$  is independent of ( $Q, V$ ): the rank-plots shown in Fig. 2 support this fact. Indeed, the sample is rather uniformly sparse in both the ( $Q, L$ ) and ( $V, L$ ) planes. Also,

**Table 1.** Maximum-Likelihood estimates of the GEV parameters for  $Q$  and  $V$ , and corresponding 95% Confidence Intervals.

Variable	Shape	Scale	Position
$Q$ ( $\text{m}^3 \text{s}^{-1}$ )	0.3677	36.2031 ( $\text{m}^3 \text{s}^{-1}$ )	59.3507 ( $\text{m}^3 \text{s}^{-1}$ )
95% C. I.	[0.15, 0.58]	[27.57, 47.55]	[48.15, 70.55]
$V$ ( $10^6 \text{ m}^3$ )	0.6149	1.5246 ( $10^6 \text{ m}^3$ )	1.7231 ( $10^6 \text{ m}^3$ )
95% C. I.	[0.37, 0.86]	[1.10, 2.11]	[1.26, 2.19]

the estimates of the Kendall’s  $\tau$  and the Spearman’s  $\rho$  are not statistically significant (as confirmed by the corresponding  $p$ -values), and formal tests of independence suggest to accept the hypothesis that  $L$  is independent of ( $Q, V$ ). On the contrary, the variables ( $Q, V$ ) are significantly positively associated, and thus  $Q$  and  $V$  are not independent: the estimates of both the Kendall’s  $\tau$  and the Spearman’s  $\rho$  are large, and the corresponding  $p$ -values are negligible (see the values reported in Fig. 2).

As in De Michele et al. (2005), a Gumbel copula was used to model the dependence between  $Q$  and  $V$ , with parameter  $\theta \approx 3.1378$ , calculated via the inversion of the Kendall’s  $\tau$ . The ability of this copula to model the available bivariate data is checked via multivariate goodness-of-fit tests (Berg,



**Fig. 2.** Trivariate rank-plot of the available  $(Q, V, L)$  observations, and bivariate rank-plots of the marginals – see text. Also shown are the estimates of the Kendall’s  $\tau$  and the Spearman’s  $\rho$ , as well as the corresponding  $p$ -values (derived from non-parametric tests of independence based on rank statistics).

2009; Genest et al., 2009; Kojadinovic et al., 2011): the resulting large  $p$ -values indicate that the Gumbel copula  $C_{QV}$  cannot be rejected at all standard levels. As a matter of facts, the analysis of the  $(Q, V)$  rank-plot in Fig. 2 shows a significant association between these two variables in the upper-right corner of the unit square: indeed, the extreme pairs practically lie on the main diagonal. Thus, it is not a surprise that the fitted Gumbel copula, having a large upper tail dependence coefficient  $\lambda_{Upp} \approx 0.75$  (Nelsen, 2006; Salvadori et al., 2007) is suitable for modeling the dependence structure of the pair  $(Q, V)$ . In passing, note that  $C_{QV}$  is an Extreme Value copula (Nelsen, 2006): since both  $F_Q$  and  $F_V$  are GEV distributions, it turns out that  $F_{QV} = C_{QV}(F_Q, F_V)$  is a bivariate Extreme Value law (after all,  $Q$  and  $V$  are annual maxima).

Given the previous results, since  $L$  can be assumed to be independent of  $(Q, V)$ , it is immediate to construct a suitable trivariate copula  $C_{QVL}$  to model the dependence structure of the triplet  $(Q, V, L)$ :

$$C_{QVL}(u, v, w) = C_{QV}(u, v) w, \tag{4}$$

where  $(u, v, w) \in I^3$ . As above, the ability of this copula to model the trivariate data is properly checked, and the resulting large  $p$ -values indicate that it cannot be rejected at all standard levels. In passing, note that also  $C_{QVL}$  is an Extreme Value copula. In addition, since  $C_{QV}$

is Archimedean (Nelsen, 2006), then  $C_{QVL}$  is a particular case of a “nested” Archimedean copula (Joe, 1997; Grimaldi and Serinaldi, 2006; Serinaldi and Grimaldi, 2007; Härdle and Okhrin, 2010; Hering et al., 2010). However,  $F_{QVL} = C_{QVL}(F_Q, F_V, F_L)$  is not a trivariate Extreme Value law, since  $F_L$  is not a GEV distribution.

### 3 Return period in a multivariate framework

In order to provide a consistent theory of RP’s in a multivariate environment, it is first necessary to precisely 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). Hereinafter, we shall consider as the object of our investigation a sequence  $\mathcal{X} = \{X_1, X_2, \dots\}$  of independent and identically distributed  $d$ -dimensional random vectors, with  $d > 1$ : thus, each  $X_i$  has the same multivariate distribution  $F$  as of the random vector  $X \sim F = C(F_1, \dots, F_d)$  describing the phenomenon under investigation, with suitable marginals  $F_i$ ’s and  $d$ -copula  $C$ . For example, we may think of a set of flood observations given by the pairs of non-independent r.v.’s Flood Peak – Flood

Volume, joined by the copula  $C$ . The case of a non-stationary sequence  $\mathcal{X}$  is rather tricky, and will be discussed in a future work.

In applications, usually, the event of interest is of the type  $\{X \in \mathcal{D}\}$ , where  $\mathcal{D}$  is a non-empty Borel set in  $R^d$  collecting all the values judged to be “dangerous” according to some suitable criterion. Note that the Borel family includes all the sets of interest in practice (like, e.g. the intervals  $(-\infty, x_1)$ ,  $(x_1, x_2)$ ,  $(x_2, \infty)$ , as well as the corresponding multivariate versions). Let  $\mu > 0$  be the average inter-arrival time of the realizations in  $\mathcal{X}$  (viz.,  $\mu$  is the average time elapsing between  $X_i$  and  $X_{i+1}$ ). Following, e.g. Embrechts et al. (2003), and given the fact that the sequence  $\mathcal{X}$  is i.i.d. (and, thus, stationary), the univariate r.v.’s  $\{B_i = I_{\mathcal{D}}(X_i)\}$  form a Bernoulli process (where  $I$  is an indicator set function), with positive probability of “success”  $p_{\mathcal{D}}$  given by

$$p_{\mathcal{D}} = \mathbb{P}(X \in \mathcal{D}), \tag{5}$$

where we assume that  $0 < p_{\mathcal{D}} < 1$ . Then, it makes sense to calculate the first random time  $A_{\mathcal{D}}$  that the sequence  $\mathcal{B} = \{B_1, B_2, \dots\}$ , generated by  $\mathcal{X}$ , takes on the value 1 (viz., the first random time that  $\mathcal{X}$  enters  $\mathcal{D}$ ):

$$A_{\mathcal{D}} = \mu \cdot \min \{i : X_i \in \mathcal{D}\}. \tag{6}$$

Clearly, the r.v.  $A_{\mathcal{D}}/\mu$  follows a Geometric distribution with parameter  $p_{\mathcal{D}}$ , and therefore the expected value of  $A_{\mathcal{D}}$  is

$$E(A_{\mathcal{D}}) = \mu_{\mathcal{D}} = \mu/p_{\mathcal{D}}. \tag{7}$$

Given the well known “memoryless property” of the Geometric distribution, and the features of the Bernoulli process (see, e.g. Feller, 1971), it is clear that  $\mu_{\mathcal{D}}$  also corresponds to the average inter-arrival time between two successive realizations of the event  $\{X \in \mathcal{D}\}$ . Evidently,  $\mu_{\mathcal{D}}$  ranges in  $[\mu, +\infty)$ : for example, if annual maxima are investigated, then  $\mu = 1$  year, and hence  $\mu_{\mathcal{D}} = 1/p_{\mathcal{D}} \geq \mu$ . We are now ready to introduce a consistent notion of RP.

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

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.

As a univariate example, let  $X$  be a r.v. with distribution  $F_X$ . In order to identify a dangerous region, traditionally a prescribed critical design value  $x^*$  is used. Then,  $\mathcal{D}$  (or, better,  $\mathcal{D}_{x^*}$ ) contains all the realizations that are judged to be “more dangerous” than  $x^*$ . For instance, in hydrology, if droughts are of concern,  $x^*$  may represent a small value of river flow, and the dangerous realizations of interest are those for which  $X \leq x^*$  (viz.,  $\mathcal{D}_{x^*} = [0, x^*]$ ). Instead, if floods are of concern,  $x^*$  may indicate a large value of river flow, and the dangerous realizations of interest are those for which

$X \geq x^*$  (viz.,  $\mathcal{D}_{x^*} = [x^*, \infty)$ ). According to Definition 1, the corresponding RP’s are  $\mu/F_X(x^*)$  in the former case, and  $\mu/(1 - F_X(x^*))$  in the latter one.

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” (viz.,  $x^*$  in the example given above), meaning in fact “the RP of the event  $\{X \text{ belongs to the dangerous region } \mathcal{D}_{x^*} \text{ identified by the given realization } x^*\}$ ”. Indeed, in a univariate framework, the assignment of  $x^*$  uniquely specifies the corresponding region  $\mathcal{D}_{x^*}$ .

Actually, also in a multivariate framework it is possible to associate a given multi-dimensional realization  $\mathbf{x}^* \in R^d$  with a dangerous region  $\mathcal{D}_{\mathbf{x}^*} \subset R^d$ . As an illustration, consider the two different bivariate dangerous regions constructed in Salvadori (2004); Salvadori and De Michele (2004). In these papers the joint behavior of the vector  $(X, Y) \sim F = C(F_X, F_Y)$  was analysed: for instance, in terms of variables of hydrological interest, think of the pairs flood peak-volume, or storm intensity-duration. In particular, great attention was paid to the following two sets:

1. (“OR” case)  $\mathcal{D}_{\mathbf{z}^*}^{\vee} = \{(x, y) \in R^2 : x > x^* \vee y > y^*\}$ , where at least one of the components exceeds a prescribed threshold (roughly, it is enough that one of the variables is too large);
2. (“AND” case)  $\mathcal{D}_{\mathbf{z}^*}^{\wedge} = \{(x, y) \in R^2 : x > x^* \wedge y > y^*\}$ , where both the components exceed a prescribed threshold (roughly, it is necessary that both variables are too large).

Here  $\mathbf{z}^* = (x^*, y^*)$  is a prescribed vector of thresholds, and  $\vee, \wedge$  are the “(inclusive) OR” and “AND” operators.

In this work we follow a different approach. The idea stems from the possibility to write, in the univariate case, the dangerous region  $\mathcal{D}_{x^*}$  in two equivalent ways: either as  $\mathcal{D}_{x^*} = \{x : x \geq x^*\}$ , or  $\mathcal{D}_{x^*} = \{x : F_X(x) \geq F_X(x^*)\}$ . Clearly, the same rationale holds by considering as a dangerous region the set  $\mathcal{D}_{x^*} = \{x : x \leq x^*\}$ , which may be of interest, e.g. for the study of droughts. Then, by considering the above formulation as given in terms of the distribution function  $F_X$ , it is clear how it can be extended in a natural way to the multi-dimensional case, as we shall illustrate below. First of all we need to introduce the following notion.

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

$$\mathcal{L}_t^F = \{\mathbf{x} \in R^d : F(\mathbf{x}) = t\}. \tag{8}$$

Clearly,  $\mathcal{L}_t^F$  is the iso-hyper-surface (having dimension  $d - 1$ ) where  $F$  equals the constant value  $t$ : thus,  $\mathcal{L}_t^F$  is a (iso)line for bivariate distributions, a (iso)surface for trivariate ones, and so on. Evidently, for any given  $\mathbf{x} \in R^d$ , there exists a *unique* critical layer  $\mathcal{L}_t^F$  supporting  $\mathbf{x}$ : namely, the

one identified by the level  $t = F(\mathbf{x})$ . Note that, thanks to Eq. (2), there exists a one-to-one correspondence between the two iso-hyper-surfaces  $\mathcal{L}_t^C = \{\mathbf{u} \in I^d : C(\mathbf{u}) = t\}$  (pertaining to  $C$  in  $I^d$ ) and  $\mathcal{L}_t^F$  (pertaining to  $F$  in  $R^d$ ).

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

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

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

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

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

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

$$T_x^> = \mu / \mathbb{P}(X \in \mathcal{R}_t^>), \tag{9}$$

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

$$T_x^< = \mu / \mathbb{P}(X \in \mathcal{R}_t^<). \tag{10}$$

In the sequel we shall concentrate only upon  $\mathcal{R}_t^>$ : the corresponding formulas for  $\mathcal{R}_t^<$  could easily be derived. Note that  $\mathcal{R}_t^>$  may be of interest, e.g. when floods are investigated, while  $\mathcal{R}_t^<$  may be appropriate if droughts are of concern.

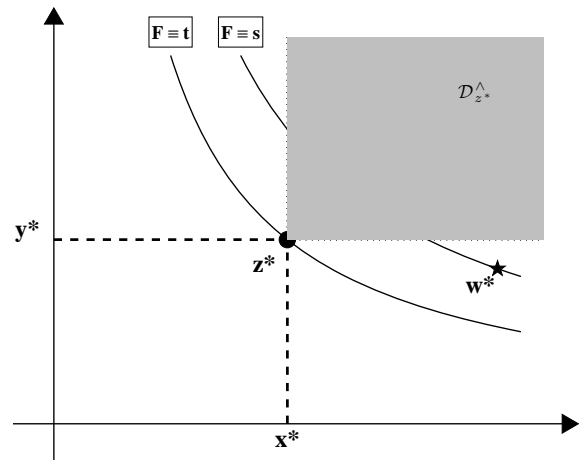
Now, in view of the results outlined in Nelsen et al. (2001, 2003), it is immediate to show that

$$\begin{aligned} T_x^> &= \frac{\mu}{\nu_F(\{\mathbf{x} \in R^d : F(\mathbf{x}) > t\})} \\ &= \frac{\mu}{1 - \nu_F(\{\mathbf{x} \in R^d : F(\mathbf{x}) \leq t\})} \\ &= \frac{\mu}{1 - K_C(t)}, \end{aligned} \tag{11}$$

where  $\nu_F$  is the probability measure induced by  $F$  over  $R^d$ , and  $K_C$  is the Kendall's distribution function associated with  $C$  (see Eq. 3 and the ensuing discussion). Clearly,  $T_x^>$  is a function of the critical level  $t$  identified by the relation  $t = F(\mathbf{x})$ . It is then convenient to denote the above RP via a special notation as follows.

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

"AND" case



**Fig. 3.** Graphical illustration of the dangerous region  $\mathcal{D}_{z^*}^\wedge$  (shaded) in the “AND” case – see text.

An advantage of the approach outlined in this work is that realizations lying over the same critical layer do always generate the same dangerous region. Evidently, this is not the case considering the “OR–AND” approach discussed above.

Furthermore, in the “AND” case, it may happen that realizations not lying in the dangerous region  $\mathcal{D}_{z^*}^\wedge$  of interest have a RP larger than the one of  $z^*$ . More specifically, as graphically illustrated in Fig. 3, for a given realization  $z^*$  lying on the isoline of level  $t \in (0, 1)$  (where  $F \equiv t$ ), the dangerous region  $\mathcal{D}_{z^*}^\wedge$  is given by the shaded area. However, given another realization  $w^*$ , lying on the isoline of level  $s > t$ , the corresponding RP may be larger than the one of  $z^*$ , but  $w^*$  does not belong to  $\mathcal{D}_{z^*}^\wedge$ . A similar rationale also holds for the “OR” case. Instead, in the approach outlined in this work, all the realizations  $\mathbf{y}$  having a KRP  $\kappa_y < \kappa_x$  must lie in  $\mathcal{R}_t^<$ , whereas all those  $\mathbf{y}$  having a KRP  $\kappa_y > \kappa_x$  must lie in  $\mathcal{R}_t^>$  – clearly, all the realizations lying over  $\mathcal{L}_t^F$  share the same KRP  $\kappa_x$ .

For the sake of convenience, we report below the algorithm explained in Salvadori and De Michele (2010) for the calculation of  $K_C$  (see also Genest and Rivest, 1993; Barbe et al., 1996), which yields a consistent Maximum-Likelihood estimator of  $K_C$ . Here we assume that the copula model is well specified, i.e. it is available in a parametric form.

**ALGORITHM 1. Calculation of the Kendall's measure function  $K_C$ .**

1. Simulate a sample  $\mathbf{u}_1, \dots, \mathbf{u}_m$  from the  $d$ -copula  $C$ .
2. For  $i = 1, \dots, m$  calculate  $v_i = C(\mathbf{u}_i)$ .
3. For  $t \in I$  calculate  $\hat{K}_C(t) = \frac{1}{m} \sum_{i=1}^m 1(v_i \leq t)$ .

As an illustration, in Fig. 4a we plot an estimate of the function  $K_C$  associated with the copula  $C_{QVL}$ : here Algorithm 1 is used, running a simulation of size  $m = 5 \times 10^7$ . Also shown is the empirical estimate of  $K_C$  calculated by using the available observations: the horizontal patterns are simply due to the small sample size.

#### 4 Quantiles associated with the KRP

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  $F_X$  it is then immediate to obtain the quantile  $x_p = F_X^{(-1)}(p)$ , which is usually unique. Then,  $x_p$  is used in practice for design purposes and rational decision making. As shown below, the same approach can also be adopted in a multivariate environment (to be compared with Belzunce et al., 2007).

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

$$q_p = \inf\{t \in I : K_C(t) = p\} = K_C^{(-1)}(p), \quad (12)$$

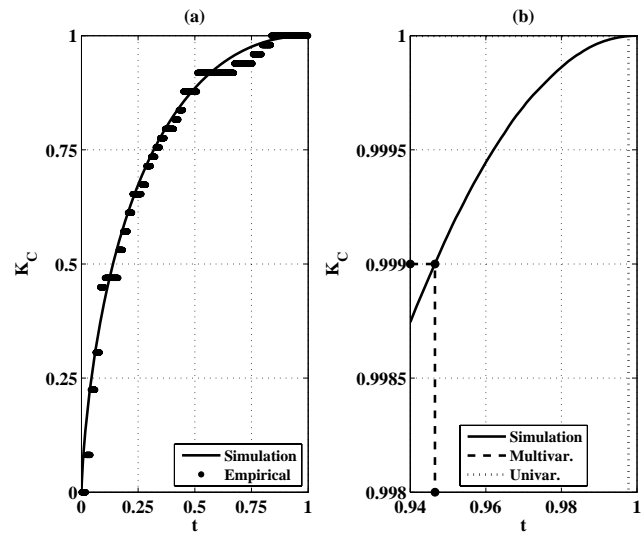
where  $K_C^{(-1)}$  is the inverse of  $K_C$ .

Definition 5 provides a close analogy with the definition of univariate quantile: indeed, recall that  $K_C$  is a univariate distribution function (see Eq. 3), and hence  $q_p$  is simply the quantile of order  $p$  of  $K_C$ . Thanks to Eq. (2), it is clear that the critical layer  $\mathcal{L}_{q_p}^F$  is the iso-hyper-surface in  $R^d$  where  $F$  takes on the value  $q_p$ , while  $\mathcal{L}_{q_p}^C$  is the corresponding one in  $I^d$  where the related copula  $C$  equals  $q_p$ .

Now, let  $\mathcal{L}_{q_p}^F$  be fixed. Then, according to Eq. (3),  $p = K_C(q_p) = \mathbb{P}(C(F_1(X_1), \dots, F_d(X_d)) \leq q_p)$ . Therefore,  $p$  is the probability measure induced by  $C$  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  $F$ ,  $np$  realizations are expected to lie in  $\mathcal{R}_{q_p}^<$ , and the others in  $\mathcal{R}_{q_p}^>$ .

**REMARK 1.** It is worth stressing that a common error is to confuse the value of the copula  $C$  with the probability induced by  $C$  on  $I^d$  (and, hence, on  $R^d$ ): on the critical layer  $\mathcal{L}_{q_p}^C$  it is  $C = q_p$ , but the corresponding region  $\mathcal{R}_{q_p}^<$  has probability  $p = K_C(q_p) \neq q_p$ , since  $K_C$  is usually non-linear (the same rationale holds for the region  $\mathcal{R}_{q_p}^>$ ). In other words, while in the univariate case the value  $p = F_X(x_p)$  corresponds to the probability induced on the region  $\mathcal{R}_p^<$ , where  $x_p$  is the quantile of  $X$  of order  $p$ , this is not so in the multivariate case.

Since  $K_C$  is a probability distribution, and  $q_p$  is the corresponding quantile of order  $p$ , we could use a standard bootstrap technique (see, e.g. Davison and Hinkley, 1997) to estimate  $q_p$  if it cannot be calculated analytically. The idea is simple, and stems directly from the very definition of  $q_p$ :



**Fig. 4.** (a) Simulation-based estimate of the function  $K_C$  (continuous line) associated with the copula  $C_{QVL}$ ; also shown is its empirical estimate (markers) calculated by using the available observations – see text. (b) Plot of the (millenary KRP) quantile  $t^* \approx 0.946537$  (thick-dashed line) associated with the critical probability level  $p = 0.999$ ; also shown (thin-dashed line) is the estimate of the value  $K_C \approx 0.999998$  associated with the critical level  $t_{ID}^* \approx 0.997754$  – see text.

viz., to look for the value  $q_p$  of  $C$  such that, in a simulation of size  $n$ ,  $np$  realizations show a copula value less than  $q_p$ . Then, by performing a large number of independent simulations of size  $n$ , the sample average of the estimated  $q_p$ 's is expected to converge to the true value of  $q_p$ . A possible algorithm is given below, most suitable for vectorial software. Here we assume that the copula model is well specified, i.e. it is available in a parametric form.

**ALGORITHM 2. Calculation of  $q_p$ .** First of all, choose a sample size  $n$ , a critical probability level  $p$ , the total number of simulations  $N$ , and fix the critical index  $k = \lfloor np \rfloor$ .

```

for i = 1:N
    S = sim(C; n); % simulate n d-vectors
                    from copula C
    C = C(S); % calculate C for simulated
              vectors
    C = sort(C); % sort-ascending
              simulated C values
    E(i) = C(k); % store new estimate of q_p
              into vector E
end
q = Mean(E); % calculate the estimate of
              q_p
    
```

Then, once the loop is completed,  $q$  provides an estimate of  $q_p$ . Practically, Algorithm 2 does the “inverse” task of Algorithm 1. The bootstrap method may also

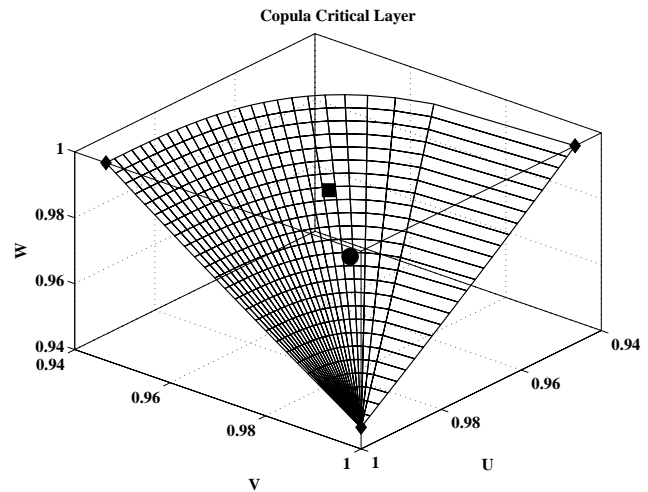
yield an approximate confidence interval for  $q_p$  (see DiCiccio and Efron, 1996 for more refined solutions): for instance, at a 10 % level, the random interval  $(q_{0.05}, q_{0.95})$  can be used, where  $q_{0.05}$  and  $q_{0.95}$  are, respectively, the quantiles of order 5 % and 95 % extracted from the vector  $\mathbf{E}$ . Using Algorithm 2 (and setting  $n = 10^4$ ,  $p = 0.999$ , and  $N = 10^7$ , for a total of  $10^{11}$  simulated triplets), we estimated  $q_{0.999} \approx 0.946537$  for the copula  $C_{QVL}$  of interest here, and a 10 % confidence interval  $(0.946110, 0.946973)$ , a process that took about 48 h of CPU time on a iMac equipped with a 3.06 GHz Intel Core 2 Duo processor and 8 GB RAM. As an illustration, in Fig. 4b we show an estimate of the function  $K_C$ , associated with the copula  $C_{QVL}$ , at  $t^* \approx 0.946537$  (corresponding to a millenary KRP): as expected, the value  $K_C(t^*)$  is almost exactly equal to 99.9 %.

As a further illustration, in Fig. 5 we plot the critical iso-surface  $\mathcal{L}_{t^*}^C$  of the trivariate copula  $C_{QVL}$  for the critical level  $t^* \approx 0.946537$ , corresponding to a regulation return period of 1000 years (viz., all the realizations on  $\mathcal{L}_{t^*}^C$  have a KRP equal to 1000 years). Then,  $C_{QVL} = t^*$  for all points belonging to  $\mathcal{L}_{t^*}^C$ . Instead,  $C_{QVL} < t^*$  (and  $\kappa_x < 1000$  years) in the region  $\mathcal{R}_{t^*}^<$  “below”  $\mathcal{L}_{t^*}^C$ , the one containing the origin  $\mathbf{0} = (0, 0, 0)$ , whereas  $C_{QVL} > t^*$  (and  $\kappa_x > 1000$  years) in the region  $\mathcal{R}_{t^*}^>$  “above”  $\mathcal{L}_{t^*}^C$ , the one containing the upper corner  $\mathbf{1} = (1, 1, 1)$ . On average, only 0.1 % of the realizations extracted from a simulation of  $C_{QVL}$  are expected to lie in  $\mathcal{R}_{t^*}^>$ . However, the level of the critical layer is  $t^* = q_{0.999} \approx 0.946537 < p = 0.999$ , as indicated by the diamonds markers in the plot.

As a further example, consider that the region  $\mathcal{R}_{0.999}^>$  identified by the critical layer  $\mathcal{L}_{0.999}^F$  (where the multivariate distribution  $F$ , or, equivalently, the copula  $C$ , takes on the value 0.999) has an estimated probability smaller than  $10^{-6}$ , and a corresponding KRP of about  $3 \times 10^6$  years: practically, only one realization of  $C_{QVL}$  out of  $3 \times 10^6$  simulations is expected to lie in  $\mathcal{R}_{0.999}^>$  (instead of 1 out of 1000). Evidently, if  $F$  (or  $C$ ) were substituted for  $K_C$  in Eq. (11) during the design phase, then the structure to be constructed would result over-sized (being expected to withstand stunning extreme events).

### 5 Design in a multivariate framework

The situation outlined in the previous section is generally similar to the one found in the study of univariate phenomena, where a single r.v.  $X$  with distribution  $F_X$  is used to model the stochastic dynamics. However, as already mentioned, the multivariate case generally fails to provide a natural solution to the problem of identifying a unique design realization. In fact, even if also the layer  $\mathcal{L}_t^F$  acts as a (multi-dimensional) critical threshold, there is no natural criterion to select which realization lying on  $\mathcal{L}_t^F$  (among the  $\infty^{d-1}$  possibilities) should be used for design purposes. In other words, in a multivariate environment, the sole tool provided by the



**Fig. 5.** Critical iso-surface  $\mathcal{L}_{t^*}^C$  of the copula  $C_{QVL}$  corresponding to the (millenary KRP) critical level  $t^* \approx 0.946537$ , indicated by the diamond markers on the axes. The circle and the square markers indicate, respectively, the Component-Wise and the Most-Likely design realizations – see text.

RP may not be sufficient to identify a design realization, and additional considerations may be required in order to pick out a “characteristic” realization over the critical layer of interest. In the following, we outline possible ways to carry out such a selection. Clearly, several approaches can be proposed, each one possibly yielding a different solution: below, we show two possible elementary strategies to deal with the problem.

The basic idea is simply to introduce a suitable function (say,  $w$ ) that “weighs” the realizations lying on the critical layer of interest. Following this approach, the practitioner can then freely choose the criterion (i.e. the function  $w$ ) that best fits the practical needs. Clearly, without loss of generality,  $w$  can be assumed to be non-negative. In turn, a “design realization” can be defined as follows.

**DEFINITION 6.** Let  $w : \mathcal{L}_t^F \rightarrow [0, \infty)$  be a weight function. The design realization  $\delta_w \in \mathcal{L}_t^F$  is defined as

$$\delta_w(t) = \operatorname{argmax}_{\mathbf{x} \in \mathcal{L}_t^F} w(\mathbf{x}), \tag{13}$$

provided that the argmax exists and is finite.

Definition 6 deserves some comments.

- In general, the unicity of the maximum may not be guaranteed. When this happens, a recourse to physical/phenomenological considerations, or to additional procedures (like, e.g. Maximum Information/Entropy schemes; Jaynes, 2003), may help solving the problem.
- Different copulas may share the same Kendall’s measure  $K_C$ , and hence the same KRP (e.g. all the bivariate Extreme Value copulas with the same Kendall’s  $\tau$ ;



Ghoudi et al., 1998). However, in general, the critical layers of such copulas will have a different geometry, and, in turn, will provide different design realizations.

- The search of the point of maximum in Eq. (13) can be subjected to additional constraints, in order to take into account the possible sensitivity of the structure under design to the behavior of specific marginals (see also the discussion in Remark 2): for instance, a Bayesian approach might be advisable.
- Sometimes it could be more appropriate to select a set of possible design realizations (i.e. an ensemble, rather than a single one) that should be used, together with experts’ opinions, in order to better evaluate the features of the phenomenon affecting the structure under design. This procedure can be carried out by using a suitable step function in Eq. (13).

In passing, we note that, in the present case study, the distribution  $F_{QVL}$  and the copula  $C_{QVL}$  are trivariate, and hence the corresponding critical layers are simply two-dimensional surfaces in  $R^3$  and  $I^3$ , respectively. Figure 5 shows the critical layer of level  $t^*$  pertaining to  $C_{QVL}$ , and the corresponding one pertaining to  $F_{QVL}$  can be drawn by exploiting Eq. (2). Now, for the sake of graphical illustration, it is possible to parametrize  $\mathcal{L}_{t^*}^F$  in polar coordinates (say,  $(\alpha, r)$ ) via a one-to-one transformation, and thus re-map and plot any function  $w$  defined over  $\mathcal{L}_{t^*}^F$  onto the rectangle  $(0, \pi/4) \times (0, \tilde{r})$ , for a suitable maximum ray  $\tilde{r}$ . In turn, it is rather easy to have a peek of the behavior of any weight function  $w$  over  $\mathcal{L}_{t^*}^F$ .

**REMARK 2.** A delicate problem may arise when adopting the approach outlined above: to make the point clear, consider the following example. Suppose that we use the duration of a storm and the storm intensity as the two variables of interest. In a fast responding system (e.g. a sewer structure), a storm having short duration but high intensity may cause a failure, whereas the same storm may not cause any problem at a catchment level. In the catchment, however, a storm with long duration and intermediate to low intensity may cause a flood event, whereas the same storm does not cause any problem to the sewer system. Now, as a matter of principle, the design realization  $\delta_w$  for the given return period (i.e. the “typical” critical storm calculated according to the strategy illustrated here) may not cause any problem in both systems, and therefore these would be wrongly designed. Practically, the sewer systems should be designed using critical design storms of short durations and high intensities, whereas a structure in the main river of the watershed should be designed using storms of long durations and low intensities. However, the problem is more apparent than real. In fact, there are neither theoretical nor practical limitations to restrict the search for the maxima in Eq. (13) over a suitable sub-region of  $\mathcal{L}_{t^*}^F$ : remember that all the realizations on the critical layer share the same prescribed KRP. Thus,

when a sewer system is of concern, only storms having short durations and high intensities could be considered, whereas a critical design storm for a structure in the main river could be spotted by restricting the attention to storms of long durations and low intensities. Roughly speaking, in the approach outlined here, the calculation of the critical design realization can be made dependent on both the environment in which a structure should be designed, as well as on the stochastic dynamics of the phenomenon under investigation.

Overall, the procedure to identify the design realization could be described as follows. Let  $X$  be a random vector with distribution  $F = C(F_1, \dots, F_d)$ .

1. Fix a RP  $T$ .
2. Calculate the corresponding probability level  $p = 1 - \mu/T$ .
3. Compute the Kendall’s quantile  $q_p$  as in Eq. (12), either analytically or by using Algorithm 2.
4. Fix a suitable weight function  $w$ .
5. Calculate the point of maximum  $\delta_w$  of  $w$  on the critical layer  $\mathcal{L}_{q_p}^F$ .

The resulting  $\delta_w$  represents a “typical” realization in  $R^d$  with a given RP. Roughly speaking, it denotes the design realization obtained by considering the very stochastic dynamics of the phenomenon. Note that, in general,  $\delta_w$  (or, better, the corresponding critical layer) should be considered together with other information (e.g. the physical features of the structure) in order to be correctly used in practice. For the sake of illustration, below we introduce two weight functions.

### 5.1 Component-wise excess design realization

A realization lying on the critical layer  $\mathcal{L}_t^F$  may be of major interest when all of its marginal components are exceeded with the largest probability. In simple words, we suggest to look for the point(s)  $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{L}_t^F$  such that it is maximum the probability that a dangerous realization  $\mathbf{y} = (y_1, \dots, y_d)$  satisfies all the following component-wise inequalities:

$$y_1 \geq x_1, \dots, y_d \geq x_d, \tag{14}$$

or  $\mathbf{y} > \mathbf{x}$  using a simplified notation. The next definition is immediate.

**DEFINITION 7.** The Component-Wise Excess weight function  $w_{CE}$  is defined as

$$w_{CE}(\mathbf{x}) = \mathbb{P}(X \in [\mathbf{x}, \infty)), \tag{15}$$

where  $X$  has distribution  $F = C(F_1, \dots, F_d)$ , and  $[\mathbf{x}, \infty)$  is the hyper-rectangle in  $R^d$  whose points satisfy all the inequalities stated in Eq. (14).

**Table 2.** Estimates of the critical design realizations, for a millenary return period, according to different strategies – see text. Also shown are the estimates of the univariate quantiles of order  $p=0.999$  of the variables of interest. The right-most column shows the Maximum Water Level of the dam associated with the flood event ( $Q, V, L$ ) reported on the corresponding row.

Strategy	$Q$ ( $\text{m}^3 \text{s}^{-1}$ )	$V$ ( $10^6 \text{ m}^3$ )	$L$ (m a.s.l.)	M. W. L. (m a.s.l.)
C.-E.	352.76	25.21	781.25	782.08
M.-L.	316.23	19.64	781.29	781.98
$F^{(-1)}(0.999)$	1208.9	172.58	781.44	784.80

Then, by restricting our attention to the critical layer  $\mathcal{L}_t^F$ , the following definition is immediate.

**DEFINITION 8.** The Component-wise Excess design realization  $\delta_{\text{CE}}$  of level  $t$  is defined as

$$\delta_{\text{CE}}(t) = \operatorname{argmax}_{\mathbf{x} \in \mathcal{L}_t^F} w_{\text{CE}}(\mathbf{x}), \tag{16}$$

where  $t \in (0, 1)$ .

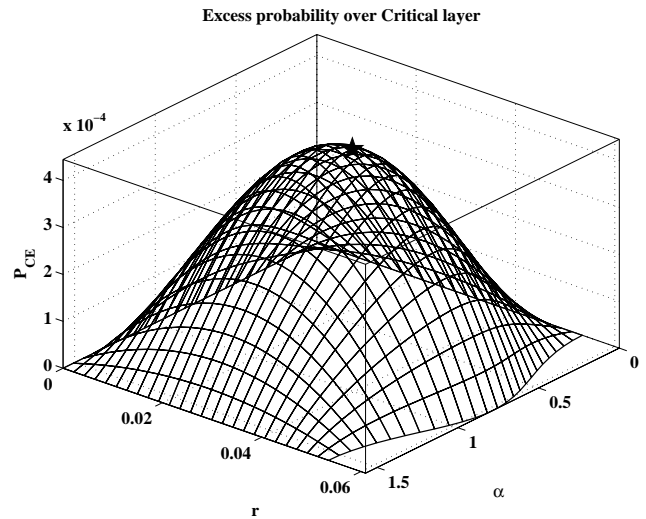
**REMARK 3.** Via the Probability Integral Transform and Sklar’s Theorem, it is easy to show that

$$w_{\text{CE}}(\mathbf{x}) = \mathbb{P}(U \in [\mathbf{u}(\mathbf{x}), \mathbf{1}]), \tag{17}$$

where  $U$  has the same copula  $C$  as of  $X$  and Uniform marginals,  $\mathbf{u}(\mathbf{x}) = (F_1(x_1), \dots, F_d(x_d))$ , and  $[\mathbf{u}, \mathbf{1}]$  is the hyper-rectangle in  $I^d$  with lower corner  $\mathbf{u}$  and upper corner  $\mathbf{1}$ . Thus, the probabilities of interest can be directly computed in the unit hyper-cube (see, e.g. Joe, 1997) by working directly on the critical layer  $\mathcal{L}_t^C$  (instead of  $\mathcal{L}_t^F$ ), a solution numerically more convenient. Note that, for large  $d$ -dimensional problems, the CPU time involved may become prohibitive, though clever solutions have been proposed for large  $d$ ’s (see, e.g. Cherubini and Romagnoli, 2009). In some cases,  $\delta_{\text{CE}}$  can be calculated analytically; otherwise, it can be empirically estimated (e.g. by calculating it over suitable points of  $\mathcal{L}_t^C$  or  $\mathcal{L}_t^F$ ).

In Fig. 6 we show the behavior of  $w_{\text{CE}}$  over  $\mathcal{L}_{t^*}^F$ , as well as the Component-wise Excess design realization  $\delta_{\text{CE}}(t^*)$  calculated for the case study investigated here (see also Fig. 5). This point has the largest probability to be component-wise exceeded by an extreme realization with KRP larger than 1000 years, and therefore it should be regarded as a sort of (statistical) “safety lower-bound”: viz., the structure under design should, at least, withstand realizations having (multivariate) size  $\delta_{\text{CE}}(t^*)$ , as reported in Table 2.

As anticipated in Sect. 2, using the design realization  $\delta_{\text{CE}}(t^*)$ , we operated the reservoir routing of the corresponding flood hydrograph. Then, we checked whether or not the reservoir level exceeds the crest level of the dam



**Fig. 6.** Polar re-mapped plot of the Component-Wise Excess weight function  $w_{\text{CE}}$  over the critical layer  $\mathcal{L}_{t^*}^F$ , corresponding to the (millenary KRP) critical level  $t^* \approx 0.946537$ . The star marker indicates where the maximum is attained – see text and Table 2.

(i.e. 784 m a.s.l.). The column “M. W. L.” in Table 2 reports the value 782.08 m a.s.l.: thus, no over-topping occurs, i.e. the dam seems to be safe against Component-Wise Excess millenary realizations.

### 5.2 Most-likely design realization

A further approach to the definition of a characteristic design event consists in taking into account the density of the multivariate distribution describing the overall statistics of the phenomenon investigated: in fact, assuming that the density  $f$  of  $F$  is well defined over  $\mathcal{L}_t^F$ , we may think of using it as a weight function.

Clearly, the restriction  $f_t$  of  $f$  over  $\mathcal{L}_t^F$  is not a proper density, since it does not integrate to one. However, it may provide useful information, since it induces a (weak) form of likelihood over  $\mathcal{L}_t^F$ : in fact, it can be used to weigh the realizations lying on  $\mathcal{L}_t^F$ , and spot those that are (relatively) “more likely” than others. Indeed,  $f_t$  inherits all the features of interest directly from the true global density  $f$ . The next definition is immediate.

**DEFINITION 9.** The Most-Likely weight function  $w_{\text{ML}}$  is defined as

$$w_{\text{ML}}(\mathbf{x}) = f(\mathbf{x}), \tag{18}$$

where  $f$  is the density of  $F = C(F_1, \dots, F_d)$ .

Then, by restricting our attention to the critical layer  $\mathcal{L}_t^F$ , the following definition is immediate.

**DEFINITION 10.** The Most-Likely design realization  $\delta_{ML}$  of level  $t$  is defined as

$$\delta_{ML}(t) = \operatorname{argmax}_{x \in \mathcal{L}_t^F} w_{ML}(x) = \operatorname{argmax}_{x \in \mathcal{L}_t^F} f(x), \quad (19)$$

where  $t \in (0, 1)$ .

**REMARK 4.** As a rough interpretation,  $\delta_{ML}$  plays the role as of a “characteristic critical realization”, i.e. the one that has to be expected if a critical event with given KRP happens. In some cases,  $\delta_{ML}$  can be calculated analytically; otherwise, it can be empirically estimated (e.g. by calculating  $f$  over suitable points of  $\mathcal{L}_t^F$ ).

In general, provided that weak regularity conditions are satisfied,  $f$  can be calculated by using the marginal densities  $f_i$ 's of  $X$ , and the density  $c = \frac{\partial^d}{\partial u_1 \dots \partial u_d} C(u_1, \dots, u_d)$  of the copula  $C$ :

$$\begin{aligned} f(x) &= \frac{\partial^d}{\partial x_1 \dots \partial x_d} C(F_1(x_1), \dots, F_d(x_d)) \\ &= c(F_1(x_1), \dots, F_d(x_d)) \cdot \prod_{i=1}^d f_i(x_i). \end{aligned} \quad (20)$$

Since our target is to compare the “weight” of different realizations, from a computational point of view it may be better to minimize  $-\ln(f)$  over  $\mathcal{L}_t^F$  (since the maxima are preserved).

As an illustration, in the present (absolutely continuous) case, the expression of the trivariate density  $f_{QVL}$  is given by

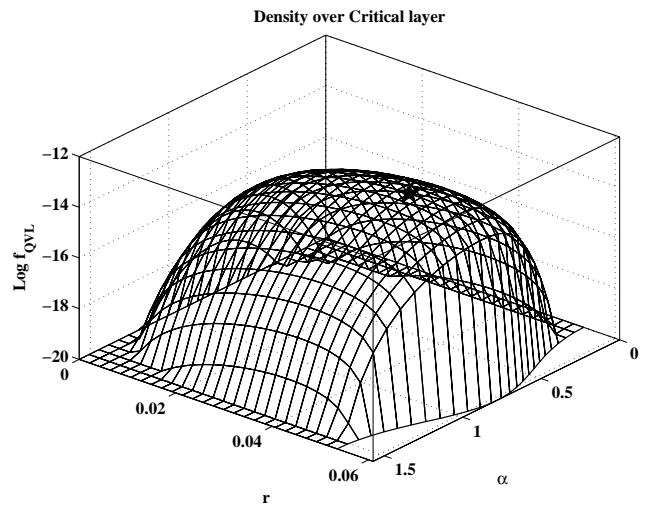
$$f_{QVL}(x, y, z) = c_{QV}(F_Q(x), F_V(y)) \cdot f_Q(x) \cdot f_V(y) \cdot f_L(z), \quad (21)$$

where  $(x, y, z) \in R^3$ , and  $c_{QV}$  is the density of the Gumbel copula modeling the pair  $(Q, V)$ . In Fig. 7 we show the behavior of (the logarithm of)  $w_{ML}$  (i.e.  $f_{QVL}$ ) over  $\mathcal{L}_{t^*}^F$ , as well as the Most-Likely design realization  $\delta_{ML}(t^*)$  calculated for the case study investigated here (see also Fig. 5). The actual values of the function  $w_{ML}$  are irrelevant: in fact, we are only interested in spotting where  $f_{QVL}$  is maximal. Therefore, the Most-Likely design realization could be regarded as the “typical” realization: viz., the structure under design should be expected to withstand events having (multivariate) size  $\delta_{ML}(t^*)$ , as reported in Table 2.

Again, as a test, using the design realization  $\delta_{ML}(t^*)$ , we operated the reservoir routing of the corresponding flood hydrograph, and checked whether or not the reservoir level exceeds the crest level of the dam. The column “M. W. L.” in Table 2 reports the value 781.98 m a.s.l.: thus, no over-topping occurs, i.e. the dam seems to be safe also against Most-Likely millenary realizations.

### 5.3 Additional remarks about design strategies

An interesting test concerning the misuse of univariate approaches in a multivariate framework can be carried out



**Fig. 7.** Polar re-mapped plot of the (log) Most-Likely weight function  $w_{ML}$  over the critical layer  $\mathcal{L}_{t^*}^F$ , corresponding to the (millenary KRP) critical level  $t^* \approx 0.946537$  (for the sake of presentation, the surface is clipped at  $-20$ ). The *star* marker indicates where the maximum is attained – see text and Table 2.

as follows. In fact, as a further possible strategy, suppose that a critical design realization  $\delta_{ID} = (x_{0.999}, y_{0.999}, z_{0.999})$  is defined in terms of the millenary univariate quantiles of the three variables of interest here (see the last row of Table 2). In turn, the layer  $\mathcal{L}_{t_{ID}^*}^{F_{QVL}}$  supporting  $\delta_{ID}$  has a critical level  $t_{ID}^* \approx 0.997754$  (see Fig. 4b), corresponding to a value of  $K_C(t_{ID}^*) \approx 0.999998$ , and a KRP of about  $5 \times 10^5$  years. It is then immediate to realize that, in order to provide a true millenary multivariate design realization, it may not be enough (or necessary) to rely upon millenary univariate quantiles. Also, operating the reservoir routing using  $\delta_{ID}$ , yields a reservoir level of about 784.80 m (see Table 2), which may cause an over-topping and a dam failure.

The example given above, as well as the illustrations presented in Sect. 5, may suggest the following empirical consideration (which, however, should be taken with care). Both the millenary multivariate design realizations  $\delta_{CE}$  and  $\delta_{ML}$  yielded a maximum water level of about 782 m a.s.l., whereas  $\delta_{ID}$  (with a KRP of the order of  $10^5$  years) generated a maximum level over-topping the dam crest by only about 80 cm. Thus, apparently, the dam is over-sized, i.e. it could withstand events with a RP much larger than 1000 years. Clearly, from the safety point of view, this is a good news. On the other hand, a smaller structure, correctly sized for withstanding true millenary multivariate events, would probably reduce the cost.

## 6 Conclusions

This paper is of methodological nature, and introduces original techniques for the calculation of design quantiles in a multivariate environment. In this work we made an effort to reduce the troublesome nature of multivariate analysis – which has always limited its practical application – by providing consistent frameworks (the KRP) and techniques (the weight functions on the critical layers) to address the identification of the critical design events when several dependent variables are involved. In particular, the “CE” and “ML” design values may provide basic realizations with given KRP, of interest in multivariate design problems.

It worth noting that the design phase should not be confused with risk assessment. In fact, the target of the former one is to provide characteristic realizations (e.g. the design realizations) useful for planning and managing a structure. In this case only the *hazard* component is taken into account, viz. the probabilistic behavior of the r.v.’s under consideration, but no specific information is exploited about the structure (e.g. the dam) under design. The risk assessment, instead, aims at pointing out possible dangerous situations by further introducing the *impact* ingredient, i.e. by considering the physical influence of the variables on the structure. In other words, the design phase only identifies the set of possible realizations (namely, those on the critical layer) that are associated with a given probabilistic level of confidence. These realizations should then be carefully used, together with additional information (e.g. the morphology of the basin), in order to provide specific parameters for quantifying the risk of a structural failure.

To the best of our knowledge, this is the first time that a similar study is presented. Clearly, further research is necessary, especially concerning the introduction of alternative design strategies, and their mutual comparison. In addition, a step towards a consistent framework for dealing with risk assessment in a multivariate environment is needed.

*Acknowledgements.* The Authors thank C. Sempì (Università del Salento, Lecce, Italy) for invaluable helpful discussions and suggestions, and M. Pini (Università di Pavia, Pavia, Italy) for software recommendations. We also thank I. Kojadinovic (Université de Pau, Pau, France) for useful comments, as well as the colleagues who provided valuable suggestions for improving the paper (B. Gräler, S. Grimaldi, S. Vandenberghe, and an anonymous referee). Also the comments and the discussions of the participants attending the S. T. A. H. Y. school “Copula Function: Theory and Practice” held in July 2011 are gratefully acknowledged. The research was partially supported by the Italian M. I. U. R. via the project “Metodi stocastici in finanza matematica”. [G. S.] The support of “Centro Mediterraneo Cambiamenti Climatici” (CMCC – Lecce, Italy) is acknowledged. [F. D.] The support of the “School of Economics and Management” (Free University of Bozen-Bolzano), via the project “Multivariate Dependence Models”, is acknowledged.

Edited by: E. Todini

## References

- Barbe, P., Genest, C., Ghoudi, K., and Rémillard, B.: On Kendall’s process, *J. Multivar. Anal.*, 58, 197–229, 1996.
- Belzunce, F., Castaño, A., Olvera-Cervantes, A., and Suárez-Llorens, A.: Quantile curves and dependence structure for bivariate distributions, *Comput. Stat. Data Anal.*, 51, 5112–5129, 2007.
- Berg, D.: Copula goodness-of-fit testing: an overview and power comparison, *Eur. J. Finance*, 15, 675–701, 2009.
- Bowman, A. and Azzalini, A.: *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, New York, 1997.
- Bras, R. L.: *Hydrology: an introduction to hydrologic science*, 1st Edn., Addison-Wesley, Boston, 1990.
- Briere, F.: *Drinking-water distribution, sewage, and rainfall collection*, Polytechnic International, Montreal, Canada, 1999.
- Chaouch, M. and Goga, C.: Design-Based estimation for geometric quantiles with application to outlier detection, *Comput. Stat. Data Anal.* 54, 2214–2229, 2010.
- Chebana, F. and Ouarda, T. B. M. J.: Index flood-based multivariate regional frequency analysis, *Water Resour. Res.*, 45, W10435, doi:10.1029/2008WR007490, 2009.
- Chebana, F. and Ouarda, T.: Multivariate extreme value identification using depth functions, *Environmetrics*, 22, 441–455, doi:10.1002/env.1089, 2011a.
- Chebana, F. and Ouarda, T. B. M. J.: Multivariate quantiles in hydrological frequency analysis, *Environmetrics*, 22, 63–78, doi:10.1002/env.1027, 2011b.
- Cherubini, U. and Romagnoli, S.: Computing the Volume of  $N$ -Dimension Copulas, *Appl. Math. Finance*, 16, 307–314, 2009.
- Chow, V. T., Maidment, D., and Mays, L. W.: *Applied Hydrology*, 1st Edn., McGraw-Hill, Singapore, 1988.
- Davison, A. C. and Hinkley, D. V.: *Bootstrap methods and their application*, vol. 1 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1997.
- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R.: Bivariate statistical approach to check adequacy of dam spillway, *J. Hydrol. Eng.-ASCE*, 10, 50–57, 2005.
- De Michele, C., Salvadori, G., Passoni, G., and Vezzoli, R.: A multivariate model of sea storms using copulas, *Coast. Eng.*, 54, 734–751, 2007.
- DiCiccio, T. J. and Efron, B.: Bootstrap Confidence Intervals, *Stat. Sci.*, 11, 189–228, 1996.
- Durante, F. and Salvadori, G.: On the construction of Multivariate Extreme Value models via copulas, *Environmetrics*, 21, 143–161, 2010.
- Embrechts, P., Klüppelberg, C., and Mikosch, T.: *Modelling Extremal Events for Insurance and Finance*, 4th Edn., Springer-Verlag, 2003.
- Feller, W.: *An introduction to probability and its applications*, vol. 1, 3rd Edn., J. Wiley & Sons, New York, 1971.
- Genest, C. and Favre, A.: Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.-ASCE*, 12, 347–368, 2007.
- Genest, C. and Rivest, L.-P.: Statistical inference procedures for bivariate Archimedean copulas, *J. Am. Stat. Assoc.*, 88, 1034–1043, 1993.
- Genest, C. and Rivest, L.-P.: On the multivariate probability integral transformation, *Stat. Probab. Lett.*, 53, 391–399, 2001.

- Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, *Insur. Math. Econom.*, 44, 199–213, 2009.
- Ghoudi, K., Khoudraji, A., and Rivest, L.: Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles, *Can. J. Stat.*, 26, 187–197, 1998.
- Grimaldi, S. and Serinaldi, F.: Asymmetric copula in multivariate flood frequency analysis, *Adv. Water Resour.*, 29, 1155–1167, 2006.
- Härdle, W. and Okhrin, O.: De copulis non est disputandum – Copulae: an Overview, *Adv. Stat. Anal.*, 94, 1–31, 2010.
- Hering, C., Hofert, M., Mai, J.-F., and Scherer, M.: Constructing hierarchical Archimedean copulas with Lévy subordinators, *J. Multivar. Anal.*, 101, 1428–1433, 2010.
- Jaworski, P., Durante, F., Haerdle, W., and Rychlik, T., eds.: *Copula Theory and its Applications*, vol. 198 of *Lecture Notes in Statistics – Proceedings*, Springer, Berlin, Heidelberg, 2010.
- Jaynes, E. T.: *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.
- Joe, H.: *Multivariate models and dependence concepts*, Chapman & Hall, London, 1997.
- Kojadinovic, I., Yan, J., and Holmes, M.: Fast large-sample goodness-of-fit tests for copulas, *Stat. Sinica*, 21, 841–871, 2011.
- Kottegoda, N. and Rosso, R.: *Probability, statistics, and reliability for civil and environmental engineers*, McGraw-Hill, 1997.
- McNeil, A. J. and Nešlehová, J.: Multivariate Archimedean copulas,  $d$ -monotone functions and  $L^1$ -norm symmetric distributions, *Ann. Stat.*, 37, 3059–3097, 2009.
- Midttomme, G., Honningsvag, B., Repp, K., Vaskinn, K., and Westernen, T.: Dams in a European Context, in: *Proceedings of the 5th ICOLD European Symposium, Geiranger (Norway), 25–27 June 2001*, edited by: Midttomme, G., Honningsvag, B., Repp, K., Vaskinn, K., and Westernen, T., Balkema, Lisse, The Netherlands, 2001.
- Nappo, G. and Spizzichino, F.: Kendall distributions and level sets in bivariate exchangeable survival models, *Inform. Sci.*, 179, 2878–2890, 2009.
- Nelsen, R.: *An introduction to copulas*, 2nd Edn., Springer-Verlag, New York, 2006.
- Nelsen, R. B., Quesada-Molina, J. J., Rodríguez-Lallena, J. A., and Úbeda-Flores, M.: Distribution functions of copulas: a class of bivariate probability integral transforms, *Stat. Probab. Lett.*, 54, 277–282, 2001.
- Nelsen, R. B., Quesada-Molina, J. J., Rodríguez-Lallena, J. A., and Úbeda-Flores, M.: Kendall distribution functions, *Stat. Probab. Lett.*, 65, 263–268, 2003.
- Salvadori, G.: Bivariate return periods via 2-copulas, *Stat. Methodol.*, 1, 129–144, 2004.
- Salvadori, G. and De Michele, C.: Frequency analysis via Copulas: theoretical aspects and applications to hydrological events, *Water Resour. Res.*, 40, W12511, doi:10.1029/2004WR003133, 2004.
- Salvadori, G. and De Michele, C.: Multivariate multiparameter extreme value models and return periods: A copula approach, *Water Resour. Res.*, 46, W10501, doi:10.1029/2009WR009040, 2010.
- Salvadori, G., De Michele, C., Kottegoda, N., and Rosso, R.: *Extremes in nature. An approach using copulas*, vol. 56 of *Water Science and Technology Library*, Springer, Dordrecht, 2007.
- Serfling, R.: Quantile functions for multivariate analysis: approaches and applications, *Stat. Neerlandica*, 56, 214–232, 2002.
- Serinaldi, F. and Grimaldi, S.: Fully nested 3-copula: procedure and application on hydrologic data, *J. Hydrol. Eng.-ASCE*, 12, 420–430, 2007.
- Serinaldi, F. and Grimaldi, S.: Synthetic design hydrographs based on distribution functions with finite support, *J. Hydrol. Eng.-ASCE*, 16, 434–445, doi:10.1061/(ASCE)HE.1943-5584.0000339, 2011.
- Singh, V., Jain, S., and Tyagi, A.: *Risk and Reliability Analysis*, ASCE Press, Reston, Virginia, 2007.
- Sklar, A.: Fonctions de répartition à  $n$  dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, 8, 229–231, 1959.
- Soil Conservation Service: *Hydrology, Sec. 4 of National Engineering Handbook*, Tech. rep., U. S. Department of Agriculture, Washington, D.C., 1972.
- Vandenberghe, S., Verhoest, N. E. C., Buyse, E., and De Baets, B.: A stochastic design rainfall generator based on copulas and mass curves, *Hydrol. Earth Syst. Sci. Discuss.*, 7, 3613–3648, doi:10.5194/hessd-7-3613-2010, 2010.
- Zoppou, C.: Reverse routing of flood hydrographs using level pool routing, *J. Hydrol. Eng.-ASCE*, 4, 184–188, 1999.