

Dismissing return periods!

Francesco Serinaldi

© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract The concept of return period in stationary univariate frequency analysis is prone to misconceptions and misuses that are well known but still widespread. In this study we highlight how nonstationary and multivariate extensions of such a concept are affected by additional misconceptions, thus easily resulting in further ill-posed procedures and misleading conclusions. We also show that the concepts of probability of exceedance and risk of failure over a given design life period provide more coherent, general and well devised tools for risk assessment and communication.

Keywords Return period · Nonstationary frequency analysis · Multivariate frequency analysis · Copulas · Risk of failure · Design values · Design life

1 Introduction

Return period \mathcal{T} is probably one of the most used and misused concepts in hydrological and geophysical risk analysis. \mathcal{T} is commonly written as:

$$\mathcal{T} = \frac{\mu}{p} = \frac{\mu}{\mathbb{P}[X > x]} = \frac{\mu}{1 - F(x)} \quad (1)$$

where X is a random variable describing the process under study (e.g. flow or rainfall peaks above a given threshold), $\mu > 0$ denotes the average inter-arrival time between two

realizations of the process, $p = \mathbb{P}[X > x]$ is the probability to observe realizations exceeding a specific value x , and $F(x) = 1 - p = \mathbb{P}[X \leq x]$ indicates the distribution function of X . Even though the attitude of considering observations of physical processes as realizations of random variables is questionable (e.g., Klemeš 1986, 2000, 2002), pure statistical frequency analyses and computation of \mathcal{T} values and corresponding return levels x are widespread in engineering, environmental sciences, and many other disciplines.

In this context, \mathcal{T} is usually preferred to values of the underlying probability of exceedance p as it seems to be (apparently) more friendly than the concept of probability. However, experience tells us that this feeling is generally not well founded, and often leads to misleading statements such as “The 50-year return period flood peak of $100 \text{ m}^3\text{s}^{-1}$ occurs once every 50 years” or “A flood peak of $100 \text{ m}^3\text{s}^{-1}$ has been recorded recently in this area. Therefore the value of $100 \text{ m}^3\text{s}^{-1}$ 50-year return period flood peak is now wrong”.

The causes of these incorrect conclusions (still widespread in technical reports and scientific literature) are discussed in textbooks and guidelines referring to frequency analyses of a single variable under the hypothesis that the observations are independent and identically distributed (*iid*) (see e.g., McCuen 1998; Fleming et al 2002; Gupta 2011, among others). However, the fast growth of nonstationary and multivariate frequency analyses occurred in the last decade led to extend the concept of return period to these frameworks. The aim of this study is to show how the causes of misconceptions mentioned above propagated in nonstationary and multivariate frequency analyses yielding further ill-posed procedures and misleading statements. In addition, we also show that the concepts of probability of exceedance and risk of failure over a given

F. Serinaldi (✉)
School of Civil Engineering and Geosciences, Newcastle
University, Newcastle Upon Tyne NE1 7RU, UK
e-mail: francesco.serinaldi@newcastle.ac.uk

F. Serinaldi
Willis Research Network, 51 Lime St., London EC3M 7DQ, UK

design life period provide more coherent, general and suitable tools to measure and communicate the risk corresponding to hydrological and geophysical hazards.

2 Reviewing some basic concepts: what risk do we really need to measure?

Before discussing nonstationary and multivariate cases, it is worth reviewing some basic concepts concerning the stationary univariate setting. Let us assume that a geophysical phenomenon is described by a random variable X and we observe realizations of the phenomenon at fixed time intervals (e.g. daily or annual time scales). Under the hypotheses that the phenomenon is stationary (i.e. the distribution function F is independent of time or other covariates) and each realization is independent of the previous ones (i.e. the realizations represent outcomes of a series of independent experiments under the same (controlled) conditions), \mathcal{T} can be defined in different ways: (1) as the expected value of the number of realizations (observed at fixed time steps) that one has to wait before observing an event whose magnitude exceeds a fixed value x ; (2) as the expected value of the number of trials between two successive occurrences of events exceeding x . The first definition is known as “average occurrence interval” (Douglas et al. 2002) and implies that a finite time τ has elapsed since a past exceedance, and the interest is in the residual or remaining waiting time for the next occurrence (Fernández and Salas 1999a), whereas the latter is known as “average recurrence interval” (Douglas et al. 2002) and conveys information about the mean elapsed time between occurrences of critical events. In the first case, τ can be known (from historical records) or unknown, whereas in the second case, we have $\tau = 0$, meaning that an exceedance has just occurred. The difference between these definitions can be (and actually is) commonly overlooked just because they both lead to Eq. 1 under *iid* hypotheses.

Referring to Chow et al. (1988, p. 382) or Cooley (2013) for the formal derivation of Eq. 1, we only recall that it corresponds to the Wald equation (Wald 1944) allowing for the calculation of the expected value of the sum of a random number n of *iid* random variables $(T_n)_{n \in \mathbb{N}}$ with common mean $E[T]$ under the hypothesis that the non-negative integer-value random variable N is also independent of the sequence $(T_n)_{n \in \mathbb{N}}$, that is (e.g., Shiau 2003; Salvadori 2004):

$$\mathcal{T} = E\left[\sum_{n=1}^N T_n\right] = E[T]E[N] = \frac{\mu}{p} \quad (2)$$

where $\mu := E[T]$, whereas $E[N] = 1/p$ results from the expected value of the geometric distribution describing the

number of events occurring before observing a realization larger than x .

As \mathcal{T} can be always expressed in years, the return period is deemed a friendly measure of the degree of rarity of an event, which however leads to statements such as “This event is expected to occur on average once each \mathcal{T} years”. This statement is formally correct but also possibly misleading because, as is well known, the underlying probability p actually says that there is a probability p to observe the so-called \mathcal{T} -year event “each year”, or better, each time interval of duration μ .

In fact, what really matters in systems’ design and planning is not \mathcal{T} and the corresponding x values obtained by inverting Eq. 1, but the risk of failure, which is the probability p_M to observe a critical event at least once in M years of design life. Under *iid* conditions, p_M is defined as (Chow et al. 1988, p. 383):

$$p_M := 1 - \prod_{j=1}^M (1 - p_j) = 1 - (1 - p)^M = 1 - (F(x_d))^M. \quad (3)$$

However, in the common practice, Eq. 3 is not used directly to define the design value x_d as it should, but only to verify the value of p_M corresponding to the \mathcal{T} -year value x , resulting in the well-known expression

$$p_M = 1 - (F(x))^M = 1 - \left(1 - \frac{1}{\mathcal{T}}\right)^M. \quad (4)$$

In other words, Eq. 3 allows us to compute a design value x_d with an appropriate probability p_M describing the actual risk of observing at least a failure in the entire design life period

$$x_d = F^{-1}((1 - p_M)^{1/M}), \quad (5)$$

whereas Eq. 4 provides the risk of failure corresponding to a value x which only accounts for a fraction $1/M$ of the time of exposure to a hazard. For example, when $M = \mathcal{T}$, the probability p_M is about 63 %, which is an almost unacceptable level of risk in the most applications. A criticism to such remarks could be that one can always choose a value of \mathcal{T} yielding a design value equal to x_d . However, this only introduces an intermediate step that makes the derivation of x_d unnecessarily more complicated. Indeed, the variables of true interest are p_M and x_d , whereas starting from the value of p corresponding to x_d is superfluous (as $p = 1 - (1 - p_M)^{1/M}$), and even more superfluous is the reciprocal transformation of p introduced by Eq. 1.

The shortcomings of \mathcal{T} and the advantages of reasoning in terms of p and p_M definitely emerge when we move from univariate *iid* conditions to univariate non-independent and/or non-identically distributed (*ni/nid*) data, and to *iid* multivariate framework.

3 Univariate nonstationary analyses: highlighting the limits of \mathcal{T}

When we move from stationary to nonstationary conditions (i.e. independent non-identically distributed *i/nid* data), the concept of return period becomes further ambiguous (Cooley 2013). However, it can still be defined in two ways for operational purposes. The first definition is the extension to nonstationary conditions of the concept of expected occurrence interval (expected waiting time until an exceedance occurs; Olsen et al 1998; Salas and Obeysekera 2014). In more detail, under nonstationarity, $p_j = \mathbb{P}[X_j > x] = 1 - F_j(x)$ is no longer constant and equal to p but changes for each trial (time step) j along the time series. Therefore, the return period in Eq. 1 becomes (Cooley 2013; Salas and Obeysekera 2014)

$$\mathcal{T} = 1 + \sum_{k=1}^{\infty} \prod_{j=1}^k (1 - p_j) = 1 + \sum_{k=1}^{\infty} \prod_{j=1}^k F_j(x). \quad (6)$$

Parey et al. (2007, 2010) extended to *i/nid* conditions an alternative definition of return period such that x is the value for which the expected number of exceedances in \mathcal{T} years (trials) is equal to one. Therefore, x is the solution of the equation (Cooley 2013)

$$1 = \sum_{j=1}^{\mathcal{T}} p_j = \sum_{j=1}^{\mathcal{T}} (1 - F_j(x)). \quad (7)$$

Unlike the *iid* case, both Eqs. 6 and 7 need to be solved numerically to obtain the required x value corresponding to the assigned \mathcal{T} (see e.g., Cooley 2013; Salas and Obeysekera 2014, for numerical details).

Whatever definition is used, the return periods and/or corresponding return levels x given by Eqs. 6 and 7 simply summarize (exactly or approximately) the average annual probability of exceedance similar to the *iid* case. Indeed, dividing both terms in Eq. 7 by \mathcal{T} and taking the reciprocal we obtain:

$$\frac{1}{\mathcal{T}} = \frac{1}{\mathcal{T}} \sum_{j=1}^{\mathcal{T}} p_j = \bar{p} \Rightarrow \mathcal{T} = \frac{1}{\bar{p}} = \frac{1}{1 - \overline{F_j(x)}}. \quad (8)$$

Therefore, similar to \mathcal{T} under *iid* conditions, Eq. 8 reveals that nonstationary \mathcal{T} under *i/nid* hypotheses does not provide additional information compared with the average value of the probabilities of exceedance p_j over the period \mathcal{T} . In other words, one can choose a prescribed average annual probability of exceedance \bar{p} to be met in the \mathcal{T} period and compute $x \equiv x_{\bar{p}}$ solving $\bar{p} = 1 - \overline{F_j(x)}$ without introducing the redundant concept of return period.

At this stage it is worth recalling that the two definitions of \mathcal{T} as average occurrence and recurrence intervals can yield different relationships between \mathcal{T} and p when the data are not independent but identically distributed (i.e. under *ni/id* conditions; Loaiciga and Mariño 1991; Fernández and Salas 1999a, b; Douglas et al 2002). This happens for instance for trials following a simple first-order Markov chain (see e.g., Fernández and Salas 1999a). Moreover, also for *ni/id* data, the final expressions of \mathcal{T} are functions of the unconditional and conditional probabilities of failure and safe events at each trial (time step).

The above discussion highlights that (1) the definition of \mathcal{T} is not unique and depends on some hypotheses about data that seldom if ever hold true for real world records, making the concept ambiguous (e.g., Fernández and Salas 1999a; Cooley 2013); (2) \mathcal{T} only summarizes the average probability of exceedance (or failure) at single time steps (trials), thus generally underestimating the actual risk of failure; and (3) in the most common applications (i.e. applying Eq. 1 under *iid* conditions), \mathcal{T} does not add information compared with p .

On the contrary, p_M has a well devised and unique general definition that suitably specializes for each situation (*iid*, *i/nid*, etc.). This is better emphasized by resorting to copula notation (Nelsen 2006). Since p_M is defined as the complement to unity of the joint probability of observing no failures in the design life period (e.g., Şen 1999; Şen et al. 2003), it can be written as:

$$\begin{aligned} p_M &= 1 - \mathbb{P}[X_1 \leq x_d \cap X_2 \leq x_d \cap \dots \cap X_M \leq x_d] \\ &= 1 - H_M(X_1 \leq x_d, X_2 \leq x_d, \dots, X_M \leq x_d) \\ &= 1 - C_M(F_1(x_d), F_2(x_d), \dots, F_M(x_d)), \end{aligned} \quad (9)$$

where H_M denotes the joint distribution of a set of random variables X_j , for $j = 1, \dots, M$, describing the process at each time step, F_j indicates the marginal univariate distribution of X_j , and C_M is the copula describing the margin-free serial dependence structure. Under *iid* conditions, $F_j = F$, $\forall j \in \{1, \dots, M\}$, and $C_M = \Pi$, so that Eq. 9 reduces to Eq. 3. For *i/nid* data (i.e. independent but nonstationary conditions), $C_M = \Pi$, whereas F_j changes at each time step (trial). For *ni/id* data (i.e. serially correlated stationary random variables), $C_M \neq \Pi$ and $F_j = F$, $\forall j \in \{1, \dots, M\}$. Therefore, Eq. 9 can account for whatever condition, and copula notation allows us to explicitly distinguish the role of temporal dependence/independence (summarized by C_M) and stationarity/nonstationarity (i.e. the assumption of identical/non-identical marginal distributions F_j).

To summarize, p_M in Eq. 9 (1) describes the actual risk of failure in the design life period; (2) has a unique definition independent of the nature of data, comprising every combination of (in)dependence and (non)stationarity

assumptions as special cases; (3) does not imply elaborated analytical derivations and/or reasoning, and extrapolations beyond the design life (unlike the summations in Eqs. 6 and 7); and (4) has an easy and straightforward interpretation. In this respect, the so-called “design life levels” proposed by Rootzén and Katz (2013) for univariate nonstationary data (i.e. *i/nid* conditions) provide x_d values yielded by one of the special cases of Eq. 9 mentioned above [see also Sivapalan and Samuel (2009), for the rationale of risk assessment under nonstationary conditions].

4 Multivariate analyses: the sleep of p reason produces \mathcal{T} monsters

4.1 Preliminary remarks

In this section, we extend the discussion to a multivariate framework involving multiple *iid* random variables. In the literature dealing with the application of multivariate frequency analysis to hydrological variables, some effort has been made to understand how Eq. 1 can be adapted to be applied in a multivariate context. Indeed, moving from the univariate to multivariate framework, the following apparent problem seems to arise: in the univariate case, a critical value x defines a unique critical region, i.e. the set of values so that $X > x$, and the denominator in Eq. 1 is uniquely defined as $1 - F(x)$, whereas in a multivariate context it seems that we have a multiple choice. Referring for instance to a bivariate case involving two random variables X and Y , they can combine in different ways yielding for instance the events $(X > x \cap Y > y)$, $(X > x \cup Y > y)$, $(X > x | Y > y)$, among many others. Such combinations of events are described by different joint and conditional distributions summarizing the corresponding joint and conditional probabilities (e.g., $\mathbb{P}[X > x \cap Y > y]$, $\mathbb{P}[X > x \cup Y > y]$, etc.). Moreover, unlike the univariate case, several (actually infinite) pairs of values (x, y) share the same joint probability t because an infinite set of pairs (x, y) fulfills for instance the equation $t = H(x, y)$, where H denotes the joint distribution of X and Y .

In light of this variety of possible cases, several studies attempted to examine the relationships between the \mathcal{T} values yielded by Eq. 1 replacing different conditional and joint probabilities of exceedance into the denominator in order to define the most appropriate choice, also making comparisons in terms of \mathcal{T} values and corresponding return levels. However, as is shown in the following, these analyses are essentially not well founded and related to the misleading nature of \mathcal{T} . The evolution of such a literature is an interesting example of how misconceptions tend to

spread more easily than good procedures and recommendations. Therefore the chronological path offers an outline for the discussion and an interesting interpretative lens.

4.2 Setting the stage

The examination of multivariate return periods and return levels requires the preliminary introduction of some concepts. For the sake of simplicity and without loss of generality, let us focus on the bivariate case so that X and Y are two random variables representing two hydrological/geo-physical variables (e.g. the rainfall intensity at two measurement stations). Let F_X and F_Y be the marginal distributions of X and Y , C their copula, and $H(x, y) = C(F_X(x), F_Y(y)) = C(u, v)$, where H is the bivariate joint distribution function of X and Y , and $U = F(X)$ and $V = G(Y)$ are standard uniform random variables. The following discussion is based on $C(u, v)$ as this allows us to work in the unit square $[0, 1]^2$ (where C is defined), making the results independent of the marginal distributions, and the graphical visualization more effective and easier. Moreover, we deal with *iid* data, i.e. temporally independent and identically distributed two-dimensional observations (x, y) . We also introduce the expressions of some joint and conditional probabilities corresponding with some bivariate return periods commonly studied in the literature. Using copula notation, we define

$$p_{\text{AND}} := \mathbb{P}[U > u \cap V > v] = 1 - u - v + C(u, v), \quad (10)$$

$$p_{\text{OR}} := \mathbb{P}[U > u \cup V > v] = 1 - C(u, v), \quad (11)$$

$$p_{\text{COND1}} := \mathbb{P}[U > u | V > v] = (1 - u)(1 - u - v + C(u, v)), \quad (12)$$

$$p_{\text{COND2}} := \mathbb{P}[U > u | V \leq v] = 1 - \frac{C(u, v)}{u}, \quad (13)$$

$$p_{\text{COND3}} := \mathbb{P}[U > u | V = v] = 1 - \frac{\partial C(u, v)}{\partial u}, \quad (14)$$

$$p_K := \mathbb{P}[\mathbb{P}[U \leq u \cap V \leq v] > t] = \mathbb{P}[C(U, V) > t] = 1 - K_C(t), \quad (15)$$

$$p_S := \mathbb{P}[g(U, V) > z] = 1 - F_Z(z). \quad (16)$$

Equations 10–14 describe the probabilities of exceedance used by Yue and Rasmussen (2002), Shiau (2003) and Salvadori (2004) in the first works providing a systematic discussion of multivariate return periods. Equation 15 describes the probability of exceedance corresponding to the so-called “secondary” or “Kendall” return period introduced by Salvadori (2004) and further investigated by Durante and Salvadori (2010), Salvadori and De Michele (2010), and Salvadori et al (2011). The function K_C in Eq. 15 is the so-called Kendall function and represents the distribution function of the copula [see e.g., Salvadori and De Michele (2010), for further details]. Finally, Eq. 16 refers to the so-called “structure-based” return period introduced by Volpi and Fiori (2014), where g is a functional relationship linking the forcing (environmental) variables with the design (structural) variable $Z = g(U, V)$. The meaning of all these probabilities will be examined in more detail in the following sections.

4.3 Some classical definitions of multivariate return periods: is something better than something else?

To our knowledge, Yue and Rasmussen (2002) provided the first systematic discussion about multivariate return periods. The aim of that work was praiseworthy recognizing that the conditional distributions, conditional return periods, and joint return periods, were misused in spite of their importance for understanding and interpreting a multivariate event. Indeed, “incorrect interpretations of these concepts will lead to misinterpretation of frequency analysis results. Thus, for both practitioners and researchers to apply these concepts appropriately in the future, the authors feel that it is necessary to assemble these concepts together and to give a clear illustration of them”. Thus, Yue and Rasmussen (2002) collected and discussed some concepts related to conditional and joint distributions and return periods, and derived some relationships between univariate and bivariate return periods. Unfortunately, the road to hell is paved with good intentions, and the same work also introduced some ambiguous final recommendations whose negative consequences still persist. Based on a bivariate model describing the relationship between flood peak and volume, Yue and Rasmussen (2002) concluded that “under a given return period, the flood peak/volume value given by the single frequency analysis is greater than those by the joint distribution. This implies that if one neglects the close correlation between flood peak and volume, and carries out single-variable frequency analysis on flood peak or volume only, the severity of a flood event may be overestimated. If a hydrologic engineering design is based on the results from the single-variable frequency analysis, then this over-evaluation will lead to an increased

cost. Hence, single-variable frequency analysis cannot provide a sufficient probabilistic assessment of a correlated multivariate event”.

Leaving out the actual nature of the correlation between flood peak and volume and the correctness of using joint distributions to describe such a relationship [see Serinaldi and Kilsby (2013), for a discussion], this sentence can be misleading, suggesting some comparisons that are actually illogical from both theoretical and practical point of view. To better understand this problem, it is worth starting from some very basic concepts. In applied sciences, probabilistic models are built and set up to describe specific situations concerning the behavior of a system. For example, hydraulic structures are designed to fulfill specific requirements, and are characterized by some key features (e.g., the length of a spillway) and operational rules. In these cases, if some variables of interest are known with uncertainty, a probabilistic model can be used to describe them and their interaction, according to physical constraints and device operating principles. In this respect, borrowing the example of flood events, if a device is designed to protect against flood peaks and is insensitive to flood volume, or the flood volume and/or duration are not of interest because the device does not manage these quantities in no way, therefore the variable of interest is only one and multivariate probabilistic models are not required. Thus, stating that the univariate frequency analysis of flood peak or volume yields an overestimation of the severity of a flood event is essentially meaningless without specifying (1) which variables are critical and are required to characterize a flood event, and (2) how these variables interact in light of the design/protection purposes.

This misconception, which is the basis of several ill-posed comparisons described in the literature, is partly related to the use of \mathcal{T} instead of the underlying probabilities. Indeed, based on Eq. 1, fixing \mathcal{T} does not mean to select a value expressed in a friendly measurement unit and summarized by the sentence “...under a given return period, e.g. 100 years...”; it means to select the probability of exceedance (or failure) corresponding to a specific and unique type of event. To clarify this point, let us consider the probabilities $p_X = \mathbb{P}[X > x]$, $p_Y = \mathbb{P}[Y > y]$, p_{OR} , p_{AND} and p_{COND1} , and the corresponding return periods given by Eq. 1 (i.e. the reciprocal of such probabilities up to the multiplying factor μ). Referring to Yue and Rasmussen (2002) and Vandenberghe et al. (2011) for analytical derivations, it can be shown that

$$p_{OR} \geq \max\{p_X, p_Y\} \geq \min\{p_X, p_Y\} \geq p_{AND} \geq p_{COND1}, \quad (17)$$

and therefore

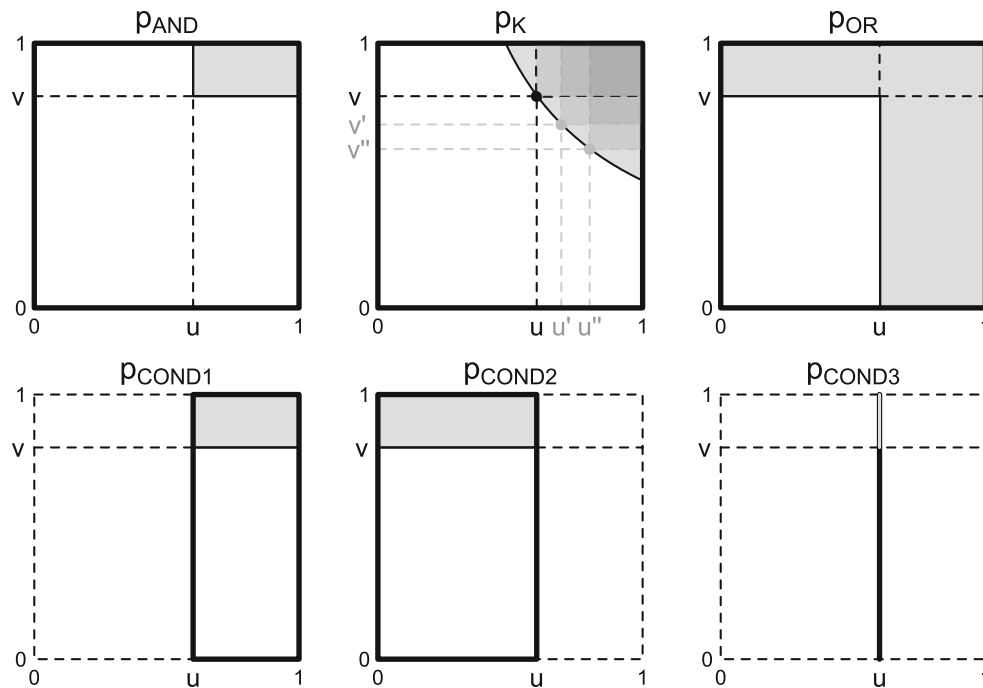


Fig. 1 Synopsis of the domains and critical regions corresponding to different types of probabilities (p_{AND} , p_K , etc.) described in the text. *Bold black lines* define the domains where the probability is computed, whereas *grey areas* denote the critical regions fulfilling the condition related to each type of probability. For example, p_{AND} is

defined over the unit square $[0, 1]^2$ and describes the chance to observe events in the *top right corner* fulfilling the condition ($U > u \cap V > v$)

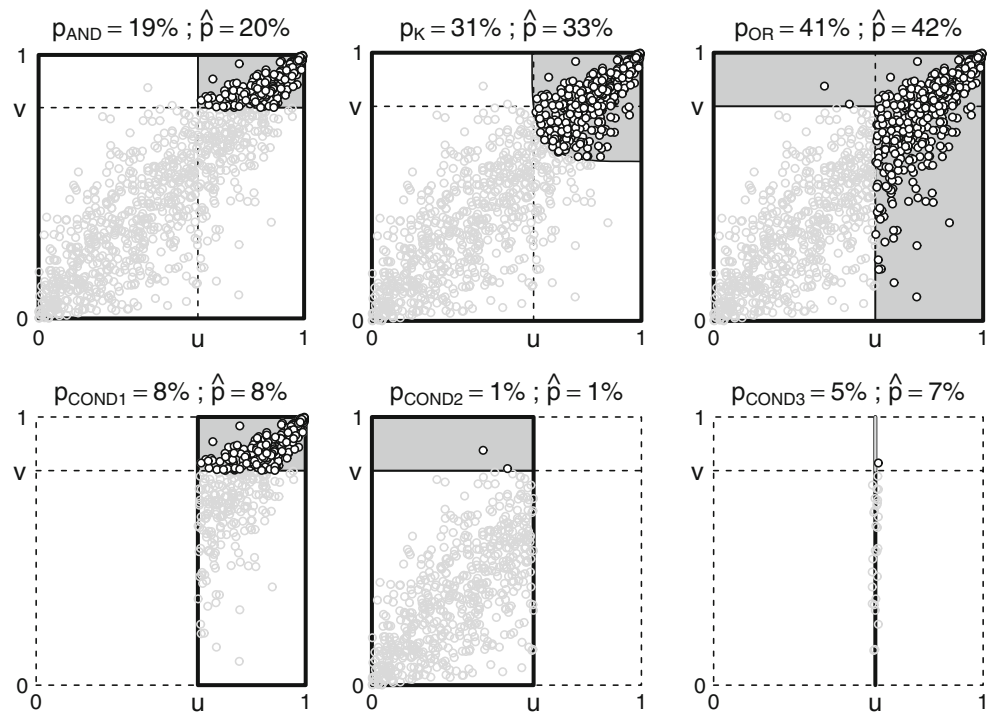
$$\mathcal{T}_{\text{OR}} \leq \min\{\mathcal{T}_X, \mathcal{T}_Y\} \leq \max\{\mathcal{T}_X, \mathcal{T}_Y\} \leq \mathcal{T}_{\text{AND}} \leq \mathcal{T}_{\text{COND1}}. \quad (18)$$

The conclusions of Yue and Rasmussen (2002) reflect these theoretical inequalities, but overlook that the probabilities in Eq. 17 describe events that cannot be ordered and compared, making statements about risk overestimation or underestimation essentially meaningless. A visualization of such concepts can help understand this issue. Figure 1 highlights the domains where the probabilities p_{AND} , p_K , p_{OR} , p_{COND1} , p_{COND2} , and p_{COND3} are defined (black bold lines) along with the subsets where the critical events corresponding to such probabilities fall (grey areas). For example, p_{AND} is a bivariate joint probability defined on the entire unit square (as U and V can assume every value in $[0, 1]^2$) and measures the chance to observe an event in the upper right quadrant defined by ($U > u \cap V > v$), whereas p_{COND1} is defined over the subset $(u, 1] \times [0, 1]$ and measures the probability to observe an event in the upper part of such a subset. Therefore, even though the analytical relationships in Eqs. 17 and 18 hold true because of pure mathematical constraints, it is evident that the different probabilities refer to different sets of events defined over different domains. In this respect, the inequality $p_{\text{OR}} \geq p_{\text{AND}}$ (or $\mathcal{T}_{\text{OR}} \leq \mathcal{T}_{\text{AND}}$) naturally derives

from the fact that both probabilities are defined on $[0, 1]^2$ but describe the chance to observe events over different subsets, being the OR subset always larger than the AND subset for each fixed pair (u, v) .

Salvadori and De Michele (2004) clearly described these aspects and highlighted that the univariate analyses are fine if only one variable is significant in the design process, whereas multivariate approaches are obviously required when several variables are involved. However, this did not prevent subsequent comparisons reported in several works. Referring to a case study discussed by De Michele et al. (2005), Salvadori and De Michele (2004) showed that \mathcal{T}_{OR} is about 20 % smaller than $\mathcal{T}_X (= \mathcal{T}_Y) = \pi$, which in turn is about 30 % larger than \mathcal{T}_{AND} , thus concluding that values different from $x_\pi (= y_\pi)$ (corresponding to univariate return periods) must be used to obtain joint events with a return period \mathcal{T}_{OR} or \mathcal{T}_{AND} equal to π . Even though this line of reasoning seems to be correct, the following question arises. If the critical configuration is described by e.g. the OR sets of bivariate events, why should one use the sets of univariate events as a reference? In other words, given that $\mathcal{T}_X (= \mathcal{T}_Y)$, \mathcal{T}_{OR} and \mathcal{T}_{AND} are different (excluding some limiting cases) as they refer to different mechanisms of failure and different sets of events, why should the values of $x_\pi (= y_\pi)$ corresponding to a univariate return period π

Fig. 2 Similar to Fig. 1, but showing 1,000 pairs (u, v) simulated from a Gumbel copula. Each panel highlights the sets of pairs falling within the domains where each probability (p_{AND} , p_K , etc.) is defined (grey circles and filled circles), and the subsets of pairs fulfilling the condition related to each type of probability (filled circles)



match the values corresponding to \mathcal{T}_{OR} or \mathcal{T}_{AND} ? Based on the inequalities in Eq. 18 and Fig. 1, it is evident that the pairs $(x_{\pi}^{\text{AND}}, y_{\pi}^{\text{AND}})$ yielding $\mathcal{T}_{\text{AND}} = \pi$ are generally different from those giving univariate $\mathcal{T}_X (= \mathcal{T}_Y) = \pi$. Of course, the comparison makes sense if one wants to quantify the error of using a probabilistic model that does not provide a suitable description of the actual mechanisms of failure. However, without specifying such mechanisms, there is no way to make comparisons and draw conclusions about possible underestimation or overestimation. Moreover, advocating the multivariate nature of some geophysical phenomena (such as floods, droughts or storms) is also insufficient to assert that a multivariate approach is better than the univariate. Indeed, every phenomenon can be described in principle by multiple variables; however, as mentioned above, sometimes only one variable is of interest for design purposes (Salvadori and De Michele 2004).

We provide a visual description to further highlight that every probability in Eqs. 10–15 (and the corresponding return periods) is perfectly coherent with the scenarios of events that it describes. Figure 2 shows 1,000 pairs (u, v) simulated from a Gumbel copula with parameter corresponding to a value of Kendall correlation equal to 0.7 and $u = 0.6$ and $v = 0.8$. Each panel displays both the set of pairs falling within the domains over which a specific probability is defined, and the subsets of pairs falling within the critical regions according to the different definitions. Theoretical and empirical values of the

probabilities of observing critical events are also reported. The agreement between theoretical and empirical probabilities, and the visualization of the sets and subsets of interest in each case should definitely clarify that (1) there is no definition better than others, (2) each definition is coherent with the scenario that it describes, and (3) making comparisons between probabilities defined over different sets and subsets of data is allowable only to show the error related to an incorrect choice of the probabilistic model. It should also be noted that reasoning in terms of probabilities allows us to bear in mind the underlying scenarios, whereas reasoning in terms of \mathcal{T} -year return period easily leads to miss the meaning of the underlying probability and to make comparisons of values that seem to be similar in terms of measurement unit, but actually describe incomparable mechanisms of failure.

4.4 Kendall and structure-based return periods:

something better than classical definitions or other facets of the same die?

Referring to OR and AND cases and flood peak and volume, Shiau (2003) anticipated a summary of the above discussion stating that “The use of \mathcal{T}_{OR} or \mathcal{T}_{AND} as the design criterion depends on what situations will destroy the structure. Under the condition that either flood peak or flood volume exceeding a certain magnitude will cause damage, then \mathcal{T}_{OR} can be used to evaluate the average recurrence interval. On the other hand, when the flood

volume and flood peak must exceed a certain magnitude that will cause damage, then \mathcal{T}_{AND} is used". As this recommendation holds true not only for OR and AND cases but also for any other case, it raises some considerations about p_K and p_S and the corresponding return periods \mathcal{T}_K and \mathcal{T}_S .

\mathcal{T}_K was proposed by Salvadori (2004) and therefore extensively applied (e.g., Salvadori and De Michele 2010, 2013; Durante and Salvadori 2010; Salvadori et al 2011; Vandenbergh et al. 2011, among others). The idea behind \mathcal{T}_K is to overcome an apparent shortcoming of \mathcal{T}_{OR} (and \mathcal{T}_{AND}) based on the following arguments. Different pairs of (U, V) , e.g. (u, v) , (u', v') and (u'', v'') , lying on the same level curve of a bivariate joint distribution share the same joint probability, i.e. $\mathbb{P}[X \leq x \cap Y \leq y] = \mathbb{P}[X \leq x' \cap Y \leq y'] = \mathbb{P}[X \leq x'' \cap Y \leq y'']$, but define different and partially overlapping p_{AND} critical regions (see e.g. panel p_K in Fig. 1). Thus, we have infinite OR (AND) critical regions characterized by the same joint probability, making a choice among them impossible (e.g., Salvadori and De Michele 2010). Since this lack of correspondence between each \mathcal{T}_{OR} (\mathcal{T}_{AND}) value and a unique critical region is incorrect from a measure theoretic point of view, Salvadori (2004) introduced \mathcal{T}_K , which relies on the Kendall distribution (or measure) K_C and measures the chance to observe an event in one of the two unique subregions defined by a level curve characterized by a unique value of joint probability. This solves the lack of dichotomy mentioned above.

However, is \mathcal{T}_K a really better tool for dealing with multivariate return periods? In other words, is \mathcal{T}_K better than \mathcal{T}_{AND} (or \mathcal{T}_{OR})? Also in this case, removing the concealing effect of Eq. 1 and reasoning in terms of probabilities, a positive answer to the above question implies for example that $\mathbb{P}[\mathbb{P}[U \leq u \cap V \leq v] \leq t] = K_C$ is better than $\mathbb{P}[U \leq u \cap V \leq v] = C$. Of course, both the probabilities legitimately exist along with every other joint and conditional probability describing the infinite possible combinations of bivariate events. They are simply different because describe different situations, cannot be interchanged, and their use only depends on which one better describes the design requirements and mechanisms of failure. In terms of critical regions, AND and OR (which rely on C) describe the probabilities associated with critical regions defined by a *unique pair of values* (u, v) , whereas \mathcal{T}_K (which relies on K_C) measures the probability associated with critical regions defined by an *infinite set of points lying on a t -level curve*. In the first case, the design criterion intrinsically focuses on (u, v) , whereas in the second case, the focus is on t . In other words, in the first case the implicit requirement is that the final *unique* design pair (u, v) guarantees a prescribed joint probability of exceedance, provided that a failure occurs when both *specific values* u and v are exceeded. In the second case, we

implicitly deal with a system which is sensitive to and can fail for a *set* of bivariate events characterized by the same joint probability of exceedance. Thus, \mathcal{T}_K , \mathcal{T}_{AND} and \mathcal{T}_{OR} simply describe different mechanisms of failure associated with different systems and must be used accordingly.

In this context, the structure-based return period introduced by Volpi and Fiori (2014) allows us to further expand the above discussion. The authors highlighted that "being strictly dependent on the particular structure under examination, the return period of structure failure usually does not match that of the hydrological loads. This entails that the multivariate approach may not fully rely on the assumption of hydrological design events, i.e., a multivariate event or an ensemble of events which all share the same (multivariate) return period". These remarks led Volpi and Fiori (2014) to introduce a so-called structure-based return period \mathcal{T}_S . Also in this case, reasoning in terms of probabilities provides a clearer picture than working with return periods. The idea was to move from the (multivariate) distribution of the hydraulic loads X and Y (e.g. peak and volume of the input hydrograph in a reservoir) to that of the actual design variable Z (e.g. the spillway design discharge) by propagating the probability density function of the hydrological loads through the function $Z = g(X, Y)$, which describes the physical dynamics of the system (e.g. the reservoir routing through the spillway). This approach is known as transformation of two random variables (e.g., Papoulis and Pillai 2002, p. 139), its univariate version ($Z = g(X)$) has been used in several applications (e.g. Kunstmann and Kastens 2006; Ashkar and Aucoin 2011; Serinaldi 2013), and in the present case it yields Eq. 16.

The comparison of Eqs. 15 and 16 highlights that p_K and p_S have the same form, meaning that C is just a particular case of g . Both C and g are used to define sets of events that fulfill some specific requirements (a prescribed value of the joint probability or a physical law) and identify two sub- and super-critical regions uniquely defined by a critical region (i.e. a curve on $[0, 1]^2$). In other words, if the generic function g describes a physical transformation of (X, Y) , the resulting design variable Z has a physical meaning, whereas if g specializes as C , the resulting design variable Z is implicitly the value of the joint probability. Is p_S (\mathcal{T}_S) better than p_K (\mathcal{T}_K) or vice versa? It is not actually. They simply focus on two different design variables, among the infinite options that can be selected using different forms of g . The choice depends on the final aim as for p_{OR} , p_{AND} , etc. Therefore the comparison between \mathcal{T}_{OR} , \mathcal{T}_K and \mathcal{T}_S critical regions is unfortunately once again not very informative. Indeed, p_{OR} , p_K and p_S correctly describe their own underlying probabilistic structures, which are different and cannot be compared. Moreover, in that

specific example, only p_S is correct as it is the only probability describing the physical mechanism under study, and stating that p_{OR} and p_K underestimate or overestimate the probability of failure is not meaningful as it is known a priori that they do not describe the critical scenarios corresponding to the mechanism of failure at hand. These comparisons may only be useful to show the error corresponding with the use of probabilities (return periods) that are known a priori to be inappropriate for the physical process of interest. Finally, the reduction of dimensionality given by the use of \mathcal{T}_K and \mathcal{T}_S , that is, the use of the univariate distributions K_C and F_Z instead of the bivariate distribution H , can be ineffective if the design (structural) variable is not unique (e.g., $\{(Z, W) : Z = g(X, Y) \text{ and } W = h(X, Y)\}$).

4.5 Multivariate risk of failure

Similar to univariate case, the definition of risk of failure p_M easily adapts to a multivariate *iid* framework. Denoting \mathcal{E}_j , for $j = 1, \dots, M$, a generic safe event, p_M can be written as

$$\begin{aligned} p_M &= 1 - \mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2 \cap \dots \cap \mathcal{E}_M] \\ &= 1 - H_M(\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_M). \end{aligned} \quad (19)$$

As for the return period, the choice of \mathcal{E}_j is not arbitrary but must describe the mechanism of failure. Thus, it can be $\{\mathcal{E}_j : X_j \leq x_d \cup Y_j \leq y_d\}$, $\{\mathcal{E}_j : X_j \leq x_d \cap Y_j \leq y_d\}$, $\{\mathcal{E}_j : g(X_j, Y_j) \leq z_d\}$ or whatever else related to the design process. Also in this case, p_M provides a more transparent and suitable measure of risk compared to \mathcal{T} as it measures the overall risk of failure along the entire design life period and requires to explicitly and carefully think about what type of event \mathcal{E}_j (and corresponding probability) must be used to obtain meaningful results fulfilling the true design requirements.

5 Conclusions

In this study, we attempted to show that the concept of return period is prone to additional misinterpretations when we move from the classical univariate frequency analysis of *iid* data to nonstationary and multivariate settings. Even though we used examples referring to hydrological processes and corresponding engineering problems, it should be noted that the discussion and methodological framework are fully general and concern the risk assessment of whatever process (environmental, geophysical, anthropogenic, etc.). Therefore, referring to a generic system (Dodge 1968) which can fail under critical conditions

according to a given mechanism of failure, our conclusions can be summarized as follows:

1. Independent of the particular framework (univariate/multivariate and stationary/nonstationary), the concept of return period \mathcal{T} does not add information compared with the underlying probabilities of exceedance p_j measuring the risk of failure each time or time interval j in which there is exposure to a specific hazard. Using financial terminology, \mathcal{T} can be seen as a derivative of the underlying p_j , and as we learned from the financial crisis of 2007–2008, derivatives can be toxic. Indeed, in spite of the simple relationships linking \mathcal{T} and p_j , return period tends to conceal the actual meaning of p_j and the underlying mechanisms of failure by an apparently friendly and understandable measurement unit.
2. Focusing on the univariate nonstationary case, we have shown that the effort to define \mathcal{T} resulted in two measures that simply summarize the average value of p_j over the \mathcal{T} period, thus better highlighting an aspect that is well known in the classical analysis of univariate *iid* data, but concealed by the compact form of Eq. 1.
3. While the concealing nature of \mathcal{T} can have a limited impact in a univariate (stationary or nonstationary) context, it easily leads to incoherent calculations and misleading conclusions in the multivariate *iid* case. Since multiple variables can combine in almost infinite ways, the multiple definitions of \mathcal{T} (\mathcal{T}_{OR} , \mathcal{T}_{AND} , etc.) introduced the belief that the choice is somewhat arbitrary and subjective, and can be object of debate. However, looking at the underlying probabilities, it is clear that such a belief is not well founded, and no meaningful debate does exist because each type of probability (p_{OR} , p_{AND} , etc.) describes in a unique way a specific mechanism of failure. Therefore the choice between the multiple definitions depends on how the system (e.g., a hydraulic device or whatever else) responds to a specific forcing. This mechanism has a unique probabilistic description that results in a specific type of p (univariate, multivariate, conditional, etc.), which in turn corresponds to a unique type of \mathcal{T} according to the mere reciprocal transformation $\mathcal{T} = \mu/p$.
4. Provided that multivariate return periods are not interchangeable because the underlying probabilities are not interchangeable, also comparisons between different definitions (so widespread in the literature) lose their meaning. Indeed, comparing different multivariate return period means to compare the probabilities describing different sets of events corresponding to different mechanisms of failure, only one of which

describes the response of the system at hand. Therefore, conclusions about supposed overestimation or underestimation are illogical and misleading because every univariate, multivariate and conditional \mathcal{T} and p correctly describes its own events' set (as shown in Fig. 2 and discussed in Sect. 4). Such comparisons may make sense only to assess the error of using a type of \mathcal{T} different from the correct one. However, also in this case the usefulness is limited as the different return periods usually correspond to very different combinations of critical events. In this context, the chain of inequalities linking some types of \mathcal{T} (see Eq. 18) results from pure mathematical constraints and provides numerical boundaries for the values of different return periods for fixed values of U and V ; however, the existence of these relationships should not be confused with the possibility of comparing probabilities that describe heterogeneous types of events defined over different domains (as is shown in Fig. 2).

5. Unlike \mathcal{T} , the risk of failure in the design life period $p_M(1)$ has a unique and general definition that can fit every situation (univariate/multivariate and stationary/nonstationary); (2) has an easy and coherent interpretation; and (3) provides a well devised measure of the actual risk to observe at least a critical event in the design life period moving from average "annual" risk summarized by p and \mathcal{T} to the actual joint probability of failure in the entire design life.

Acknowledgments This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant EP/K013513/1 "Flood MEMORY: Multi-Event Modelling Of Risk & recoverY", and Willis Research Network. The comments of four anonymous reviewers are gratefully acknowledged. The analyses were performed in R (R Development Core Team 2013) by using the contributed packages CDvine (Schepsmeier and Brechmann 2012) and copula (Yan 2007; Kojadinovic and Yan 2010).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Ashkar F, Aucoin F (2011) A broader look at bivariate distributions applicable in hydrology. *J Hydrol* 405(3–4):451–461
- Chow VT, Maidment DR, Mays LW (1988) *Applied hydrology*. McGraw-Hill, New York
- Cooley D (2013) Return periods and return levels under climate change. In: AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S (eds) *Extremes in a changing climate, water science and technology library*, vol 65. Springer, Netherlands, pp 97–114
- De Michele C, Salvadori G, Canossi M, Petaccia A, Rosso R (2005) Bivariate statistical approach to check adequacy of dam spillway. *J Hydrol Eng* 10(1):50–57
- Dooge JCI (1968) The hydrologic system as a closed system. *Bull Int Assoc Sci Hydrol* 13(1):58–68
- Douglas E, Vogel R, Kroll C (2002) Impact of streamflow persistence on hydrologic design. *J Hydrol Eng* 7(3):220–227
- Durante F, Salvadori G (2010) On the construction of multivariate extreme value models via copulas. *Environmetrics* 21(2):143–161
- Fernández B, Salas J (1999a) Return period and risk of hydrologic events. I: mathematical formulation. *J Hydrol Eng* 4(4):297–307
- Fernández B, Salas J (1999b) Return period and risk of hydrologic events. II: applications. *J Hydrol Eng* 4(4):308–316
- Fleming G, Frost L, Huntington S, Knight D, Law F, Rickard C (2002) *Flood risk management: learning to live with rivers*. Institution of Civil Engineers, London
- Gupta SK (2011) *Modern hydrology and sustainable water development*. Wiley, Chichester
- Klemeš V (1986) Dilettantism in hydrology: transition or destiny? *Water Resour Res* 22(9S):177S–188S
- Klemeš V (2000) Tall tales about tails of hydrological distributions. *I. J Hydrol Eng* 5(3):227–231
- Klemeš V (2002) Risk analysis: the unbearable cleverness of bluffing. In: Bogardi JJ, Kundzewicz ZW (eds) *Risk, reliability, uncertainty, and robustness of water resource systems*. International hydrology series. Cambridge University Press, Cambridge, pp 22–29
- Kojadinovic I, Yan J (2010) Modeling multivariate distributions with continuous margins using the copula R package. *J Stat Softw* 34(9):1–20
- Kunstmann H, Kastens M (2006) Direct propagation of probability density functions in hydrological equations. *J Hydrol* 325(1–4):82–95
- Loaiciga H, Mariño M (1991) Recurrence interval of geophysical events. *J Water Resour Plan Manag* 117(3):367–382
- McCuen RH (1998) *Hydrologic analysis and design*. Prentice Hall, NJ
- Nelsen RB (2006) *An introduction to copulas*, 2nd edn. Springer, New York
- Olsen JR, Lambert JH, Haines YY (1998) Risk of extreme events under nonstationary conditions. *Risk Anal* 18(4):497–510
- Papoulis A, Pillai SU (2002) *Probability, random variables, and stochastic processes*. McGraw-Hill, New York
- Parey S, Malek F, Laurent C, Dacunha-Castelle D (2007) Trends and climate evolution: statistical approach for very high temperatures in France. *Clim Change* 81(3–4):331–352
- Parey S, Hoang TTH, Dacunha-Castelle D (2010) Different ways to compute temperature return levels in the climate change context. *Environmetrics* 21(7–8):698–718
- R Development Core Team (2013) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. ISBN 3-900051-07-0
- Rootzén H, Katz RW (2013) Design life level: quantifying risk in a changing climate. *Water Resour Res* 49(9):5964–5972
- Salas J, Obeysekera J (2014) Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *J Hydrol Eng* 19(3):554–568
- Salvadori G (2004) Bivariate return periods via 2-copulas. *Stat Methodol* 1(1–2):129–144
- Salvadori G, De Michele C (2004) Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour Res* 40(12):WR003133
- Salvadori G, De Michele C (2010) Multivariate multiparameter extreme value models and return periods: a copula approach. *Water Resour Res* 46(10):WR009040
- Salvadori G, De Michele C (2013) Multivariate extreme value methods. In: AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S (eds) *Extremes in a changing climate, water science and technology library*, vol 65. Springer, Netherlands, pp 115–162

- Salvadori G, De Michele C, Durante F (2011) On the return period and design in a multivariate framework. *Hydrol Earth Syst Sci* 15:3293–3305
- Scheppsmeier U, Brechmann EC (2012) CDVine: statistical inference of C- and D-vine copulas. <http://CRAN.R-project.org/package=CDVine>. R package version 1.1-9
- Şen Z (1999) Simple risk calculations in dependent hydrological series. *Hydrol Sci J* 44(6):871–878
- Şen Z, Altunkaynak A, Özger M (2003) Autorun persistence of hydrologic design. *J Hydrol Eng* 8(6):329–338
- Serinaldi F (2013) On the relationship between the index of dispersion and Allan factor and their power for testing the Poisson assumption. *Stoch Environ Res Risk Assess* 27(7):1773–1782
- Serinaldi F, Kilsby CG (2013) The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs. *Water Resour Res* 49(6):3423–3442
- Shiau JT (2003) Return period of bivariate distributed extreme hydrological events. *Stoch Environ Res Risk Assess* 17(1–2): 42–57
- Sivapalan M, Samuel JM (2009) Transcending limitations of stationarity and the return period: process-based approach to flood estimation and risk assessment. *Hydrol Process* 23(11):1671–1675
- Vandenberghe S, Verhoest NEC, Onof C, De Baets B (2011) A comparative copula-based bivariate frequency analysis of observed and simulated storm events: a case study on Bartlett-Lewis modeled rainfall. *Water Resour Res* 47(7):W07529
- Volpi E, Fiori A (2014) Hydraulic structures subject to bivariate hydrological loads: Return period, design, and risk assessment. *Water Resour Res* 50(2):885–897
- Wald A (1944) On cumulative sums of random variables. *Ann Math Stat* 15(3):283–296
- Yan J (2007) Enjoy the joy of copulas: With a package copula. *J Stat Softw* 21(4):1–21
- Yue S, Rasmussen P (2002) Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrol Process* 16:2881–2898