

# Derivation of flood frequency curves through a bivariate rainfall distribution based on copula functions: application to an urban catchment in northern Italy's climate

M. Balistrocchi and B. Bacchi

## ABSTRACT

The utilization of continuous approaches, namely analytical-probabilistic methods, has often been advocated for hydraulic device sizing, in order to overcome some deficiencies of the design event method. In the analytical distribution derivation, however, strong simplifying hypotheses are usually adopted. Rainfall depth and duration independency is the most unrealistic, even if it usually leads to satisfactory agreements between derived and benchmarking distributions. The reason can lie in drawbacks related to conventional assessment techniques of multivariate rainfall distributions. Copula functions recently provided a significant improvement in statistical inference capabilities and greatly simplified the distribution assessment. Nonetheless, the generalization of the return period concept, well defined in the univariate case, to multivariate cases has not found a blanket solution yet. Effective estimate methods can, however, be developed for the design and performance assessment of specific hydraulic devices. With regard to urban catchment applications, a criterion to derive flood frequency curves from a rainfall volume and duration distribution is herein proposed. Further, a method to estimate the return period of bivariate rainfall events based on a device-targeted approach is developed. Hydrologic simulations are conducted to support model reliability through a test case, featuring a northern Italian rainfall regime.

**Key words** | bivariate return period, copula functions, derived distributions, flood frequency, Italian rainfall, urban hydrology

**M. Balistrocchi** (corresponding author)

**B. Bacchi**

Department of Civil, Environmental, Architectural Engineering and of Mathematics,  
University of Brescia,  
via Branze 43,  
Brescia 25123,  
Italy  
E-mail: [matteo.balistrocchi@unibs.it](mailto:matteo.balistrocchi@unibs.it)

## INTRODUCTION

Although the design event method is still a very popular approach for sizing hydraulic devices and for evaluating drainage system performances, its limitations have long been highlighted and debated (Adams & Howard 1986; Yen 1990). The main reason lies in assuming that the frequency of occurrence of the total depth of the design rainfall hyetograph equals that of the derived flood hydrograph. It is well known that the flood hydrograph depends on a number of factors, which are not accounted for by the rainfall statistics on which design event methods are based. Therefore, crucial characteristics, such as the

catchment initial condition, the wet weather duration and the rainfall time pattern, are arbitrarily set. As a result, estimates derived by design event methods are always biased (Bacchi *et al.* 1993).

In order to overcome such deficiencies, continuous approaches should be preferred. In particular, the analytical-probabilistic methods, originally developed by Eagleson (1972), recently aroused researcher interest (Guo *et al.* 2012). In fact, these methods feature all the advantages of continuous approaches, even if they yield closed-form analytical expressions analogous to those of the design event methods.

According to the derived distribution theory, complete cumulative distribution functions (CDFs) of the runoff variables of interest can be defined from joint distribution functions (JDFs) of the rainfall variables (Adams & Papa 2000). Nevertheless, to have the possibility of analytically integrating the derived distributions, simplifying hypotheses must be adopted in the rainfall probabilistic model: usually exponential CDFs are utilized for the marginals and the mutual variable dependency is disregarded (Guo & Adams 1998; Balistrocchi *et al.* 2013).

When dealing with rainfall depth and wet weather duration JDFs, more complex CDFs are often needed to properly suit empirical marginal distributions, while the independence hypothesis is largely rejectable. Surprisingly, simplified JDFs produce results in better agreement with benchmarks, and more conservative, than those achieved by the more realistic ones (Adams & Papa 2000). Even if not totally explained, the reasons can lie in a compensation of errors in the simplified models and in drawbacks related to the JDF assessment by means of traditional techniques in the complete models.

Fortunately, copula functions (Nelsen 2006), recently introduced in hydrologic research (De Michele & Salvadori 2003; Favre *et al.* 2004), offer the opportunity to broaden the multivariate inference capability. Indeed, through copula approach, the assessment of the dependence structure relating to random variables can be carried out separately from those of marginal distributions (Genest & Favre 2007; Salvadori *et al.* 2007). As a consequence, the dependence structure analysis is no longer affected by marginal behaviours. Further, copula functions and marginal distributions belonging to different probability families, even complex, can be utilized to develop the JDF. The derivation procedure therefore becomes straightforward, eliminating additional sources of uncertainty, such as preliminary sample data transformations, and permitting an effective verification of model reliability by means of blanket test statistics (Genest *et al.* 2009).

Owing to these significant advances, the longstanding debate regarding the generalization of the return period concept, from the univariate case to the multivariate one, has been notably reopened. Traditionally, two methods based on simple logical expressions 'AND' 'OR', already utilized in conventional multivariate statistics but susceptible to copula interpretation, were utilized. More recently,

Salvadori & De Michele (2010) proposed a new estimation method relying on Kendall function.

However, as can be seen in Gräler *et al.* (2013), where a review of approaches to develop multivariate design events is provided, all these methods lead to statistically different outcomes; moreover, a generally applicable solution cannot be suggested, so that the most suitable method should be selected in consideration of the specific hydraulic application at hand. Hydrologic event severity is actually related to hydraulic device performances: on the one hand, all events yielding identical performances can be associated with the same return period; on the other hand, the same event can be associated with different return periods depending on the type of device.

According to this device-targeted approach, the return period of a derived variable, exploited to express device performances, can be estimated by conventional univariate techniques and thus be associated with the input multivariate event. For instance, the return period of bivariate flood events, defined by using peak discharge and runoff volumes, was estimated with reference to routing reservoir performances by Requena *et al.* (2013) and Volpi & Fiori (2014), by using univariate statistics of maximum water level of the stored volume and maximum routed discharge, respectively.

Bearing in mind these encouraging results, herein the derivation of a flood frequency curve (FFC) from a JDF of storm volume and wet weather duration is developed for urban applications in medium-size catchments (area less than 100:200 ha) located in northern Italy. To do so, a hydrologic loss model and a runoff routing model must be established. As demonstrated by Candela *et al.* (2014) dealing with extended natural watersheds, a copula-based rainfall stochastic generator can be successfully coupled with a fully conceptual distributed model. Nevertheless, a lumped modelling based on hydrologic parameters operatively adopted in practical engineering was preferred in this study. Owing to rainfall-runoff transformation dynamics in urban catchments, this type of modelling was considered to be sufficiently reliable and more appealing for practitioners.

The objective of this paper is two-fold: first, quantifying the modelling improvement achievable by a complete JDF with respect to a simplified analytical-probabilistic model, in a case where the speculated compensation of errors does not occur; second, developing a device-targeted criterion to

estimate the return period of bivariate rainfall events from univariate statistics of the runoff peak discharge.

Hereafter, copula analysis technique and JDF construction are briefly recalled in the first section, while special attention is given to the return period estimate in the second one, focusing on the existing method drawbacks and on the development of the proposed method. The continuous simulation model employed to derive the benchmarking FFC, available data and the test watershed are described in the following sections. JDF fitting results and the FFC derivation are then discussed. Conclusions are drawn in the last section.

## BIVARIATE ANALYSIS OF RAINFALL DEPTH AND DURATION

To cope with JDF assessment by using copula functions, random variables uniformly distributed on  $\mathbb{I} = [0,1]$  must be derived from original random variables through probability integral transform. In this bivariate case, uniform random variables  $u$  and  $v$  are defined with respect to their natural counterparts, the storm volume  $h$  and the wet weather duration  $d$ , by means of the respective CDFs  $P_H$  and  $P_D$  as shown in Equation (1):

$$\begin{cases} u = P_H(h) \\ v = P_D(d) \end{cases} \quad (1)$$

A 2-copula is a bivariate JDF  $C(u,v): \mathbb{I}^2 \rightarrow \mathbb{I}$ , such that constrains (2) and (3) are satisfied. The first one states that copula's marginals are uniformly distributed on  $\mathbb{I}$ , while the second one establishes a bivariate increasing trend.

$$C(u, 0) = 0, \quad C(u, 1) = u, \quad C(0, v) = 0, \quad C(0, 1) = v \quad (2)$$

with  $u, v \in \mathbb{I}$

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0 \quad (3)$$

for every  $u_1, u_2, v_1, v_2 \in \mathbb{I}$   
such that  $u_1 \leq u_2, v_1 \leq v_2$

According to the fundamental Sklar theorem (Sklar 1959), the bivariate JDF  $P_{HD}(h,d): \mathbb{R}^+ \rightarrow \mathbb{I}$  can be expressed by means of an underlying 2-copula  $C$ , as shown

in Equation (4). The Sklar theorem ensures that this function exists and, if the marginals are continuous, is unique.

$$P_{HD}(h, d) = C[P_H(h), P_D(d)] \quad (4)$$

It is evident that, unlike  $P_{HD}$ , the copula function is not affected by marginal distributions and exclusively expresses the dependence structure. Thus, the assessment of  $P_{HD}$  can be carried out in two distinct phases, involving that of the copula  $C$  and those of univariate distributions  $P_H$  and  $P_D$ .

To perform an event based analysis, however, the continuous series must preliminarily be separated in independent occurrences by a discretization procedure, to obtain a bivariate sample of rainfall depths  $\hat{h}_i$  and wet weather durations  $\hat{d}_i$ . Plotting positions of natural occurrences, usually called pseudo-observations, define sample versions  $\hat{u}_i$  and  $\hat{v}_i$  of uniform random variables and allow construction of the empirical copula (Ruymgaart 1973). Equation (5) reports the expression of the bivariate empirical copula  $C_n$ , in which  $\mathbf{1}(\cdot)$  is the indicator function,  $R(\cdot)$  is the rank operator and  $n$  is the sample size.

$$\begin{aligned} C_n(u, v) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{u}_i \leq u, \hat{v}_i \leq v) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}\left(\frac{R(\hat{h}_i)}{n+1} \leq u, \frac{R(\hat{d}_i)}{n+1} \leq v\right) \end{aligned} \quad (5)$$

$C_n$  being a consistent estimator of the underlying copula  $C$  (Deheuvels 1979), it plays a pre-eminent role when the most suitable copula family is selected and fitted to pseudo-observations. Conversely, the assessment of marginal distributions  $P_H$  and  $P_D$  can be carried out by using well-established univariate inference techniques with regard to natural observations  $\hat{h}_i$  and  $\hat{d}_i$ .

In the following sub-sections, the three phases outlined above – (i) discretization procedure, (ii) underlying copula assessment and (iii) marginal distribution assessments – are discussed in more detail, justifying the selection of theoretical functions with reference to existing literature.

### Continuous rainfall series discretization

The separation of a continuous precipitation record into independent events can be carried out by applying two

discretization thresholds: an interevent time definition (*IETD*) and a rainfall depth threshold (Balistrocchi *et al.* 2009). The first one represents the minimum dry weather period needed for two following hyetographs to be considered independent. Hence, if two rain bursts are detached by a dry weather period shorter than *IETD*, they are aggregated into a unique event, whose duration and depth are computed from the beginning of the first one to the end of the latter one.

The second parameter corresponds to the minimum rainfall depth that must be exceeded in order to have a rainfall relevant to analysis purposes. When this condition is not satisfied, the event is suppressed and the corresponding wet weather duration is assumed to be rainless and joined to the adjacent dry weather periods.

Discretization parameters strongly affect the derived sample statistics, namely the mean annual number of rainfall events and marginal parameters. Then, their values must be chosen very carefully. When dealing with FFC derivation, a suitable criterion is focusing on runoff discharge characteristics: the *IETD* can be estimated as the minimum time that avoids the hydrographs generated by two subsequent rainfalls to overlap, while the depth threshold can be identified with the initial abstraction (*IA*) of the catchment hydrological losses.

### Dependence structure modelling

As demonstrated by Balistrocchi & Bacchi (2011), the Gumbel–Hougaard 2-copula (Nelsen 2006; Salvadori *et al.* 2007) provides a suitable model to represent the dependence structure of storm volume and wet weather duration in Italian climates. This family belongs to Archimedean copulas and derives from the generator function  $\psi$  (6), in which  $\theta$  represents the dependence parameter:

$$\psi(t) = (-\ln t)^\theta \quad \text{with } t \in \mathbb{I} \quad (6)$$

Hence, bivariate members of this family can be obtained as shown in Equation (7), by using  $\psi$  and its pseudo-inverse function  $\psi^{-1}$ :

$$C(u, v) = \psi[\psi(u) + \psi(v)]^{-1} \quad (7)$$

Substituting the function (6) in Equation (7), the bivariate function (8) is written for the Gumbel–Hougaard 2-copula. This is a symmetric, mono-parametric and extreme value copula, which is able to suit only concordant associations:

$$C(u, v) = \exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\} \quad (8)$$

In this copula  $\theta$  must be greater than or equal to one and is algebraically related to the Kendall coefficient  $\tau_K$  through the relationship (9), so that the stronger the concordance, the larger  $\theta$  is. However, the Gumbel–Hougaard copula is comprehensive of the independence copula, which is obtained when  $\tau_K$  is zero and  $\theta$  is equal to one:

$$\tau_K = \frac{\theta - 1}{\theta} \quad \text{with } \theta \geq 1. \quad (9)$$

The satisfactory adaptation of the Gumbel–Hougaard copula to pseudo-observations is mainly due to its behaviour both in the upper tail and in the lower tail: in the Gumbel–Hougaard copula, the upper tail dependence coefficient (10) exists and increases with  $\theta$ , while the lower tail dependence coefficient is identically null:

$$\lambda_u = 2 - 2^{1/\theta} \quad (10)$$

Consequently, an attitude to generate strongly concordant rainfall events arises in the upper tail; on the contrary, in the lower tail the association is weaker. This matches with the empirical evidence in the sub-alpine climate (Balistrocchi *et al.* 2009; Balistrocchi & Bacchi 2011). Short wet weather durations are often associated with heavy rainfall depths, especially in spring and summer, when intense convective storms can easily occur. Conversely, extended frontal rainfall events, which prevail in other seasons, determine a more proportional relationship between wet weather durations and rainfall depths. When all these events are joined in a single sample, the global effect is to decrease the concordance of short duration events, while the strong concordance of long duration ones persists.

The Gumbel–Hougaard 2-copula (8) fitting to pseudo-observations can be performed by using the moment-like

method or the maximum likelihood criterion. In the first method, the dependence parameter  $\theta$  is simply expressed as a function of the sample version of the Kendall coefficient, by inverting Equation (9); in the second one, the dependence parameter is estimated as the one that maximizes the pseudo-log-likelihood estimator (11), in which  $c$  is the copula density:

$$L(\theta) = \sum_{i=1}^n \log [c(\hat{u}_i, \hat{v}_i)] \quad (11)$$

The null hypothesis that the underlying copula is the fitted Gumbel-Hougaard 2-copula must be verified by test statistics. In this regard, an effective blanket test has been developed by Genest *et al.* (2009) to assess the goodness-of-fit of the selected theoretical copula with respect to the empirical one. In this test, the residuals between the theoretical copula (8) and the empirical copula (5) are summarized in a Cramér-Von Mises statistic  $S_n$  (12) and compared to those of samples generated by using Monte Carlo-like simulations, under the null hypothesis:

$$S_n = \sum_{i=1}^n [C_n(\hat{u}_i, \hat{v}_i) - C(\hat{u}_i, \hat{v}_i)]^2 \quad (12)$$

An empirical estimate  $p$  of the test significance, according to which the null hypothesis cannot be rejected, is hence given by Equation (13), where  $N$  is the number of simulation runs, much larger than the sample size  $n$ , and  $S_{n,k}$  are corresponding statistics:

$$p = \frac{1}{N} \sum_{k=1}^N \mathbf{1}(S_{n,k} > S_n) \quad (13)$$

Finally, as already highlighted by Poulin *et al.* (2007), the proper representation of the tail behaviour is crucial to achieving a reliable return period estimate. Thus, in addition to the assessment of the overall goodness-of-fit, the consistency between the upper tail dependence coefficient of the fitted copula (10) and that of the empirical copula  $\hat{\lambda}_u$  must be ensured. Herein, the non-parametric estimator (14) was employed as a term of comparison, since Frahm *et al.*

(2005) demonstrated that it performs well if the upper tail dependence exists.

$$\hat{\lambda}_u = 2 - 2 \exp \left\{ \frac{1}{n} \sum_{i=1}^n \log \left[ \sqrt{\log \frac{1}{u_i} \log \frac{1}{v_i}} / \log \frac{1}{\max(u_i, v_i)^2} \right] \right\} \quad (14)$$

### Marginal distribution modelling

An appropriate model for both marginals was identified in the Weibull CDF, that can be viewed as a generalization of the exponential model (Balistocchi & Bacchi 2011):  $P_H$  and  $P_D$  are therefore represented by means of CDFs (15) and (16), respectively:

$$P_H(h) = \begin{cases} 1 - \exp \left[ - \left( \frac{h - IA}{\zeta} \right)^\beta \right] & \text{for } v \geq IA \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

$$P_D(d) = \begin{cases} 1 - \exp \left[ - \left( \frac{d}{\lambda} \right)^\gamma \right] & \text{for } d \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Exponents  $\beta$  and  $\gamma$  are dimensionless parameters that rule the distribution shape, while the denominators  $\zeta$  (mm) and  $\lambda$  (h) are scale parameters. The lower limit  $IA$  in CDF (15) is set in accordance with the volume threshold employed in the rainfall discretization procedure. In fact, the shape parameter makes the distribution very versatile and leads to fittings as satisfactory as those achievable by using more complex models. The exponent less than one delineates a probability density function monotonically decreasing from the lower limit, where a vertical asymptote is present. The exponent greater than one shows that the probability density function has a finite mode and exhibits a right tail.

In order to quantify the impact on the whole model reliability of assuming exponential marginals, as in the most common analytical-probabilistic models, JDF with exponential marginals were fitted as well. Equations (15) and (16) consequently simplify, being the shape parameters

$\beta$  and  $\gamma$  unitary, and feature a finite mode in the lower limit, equal to the reciprocal of the scale parameter.

## RETURN PERIOD ESTIMATE

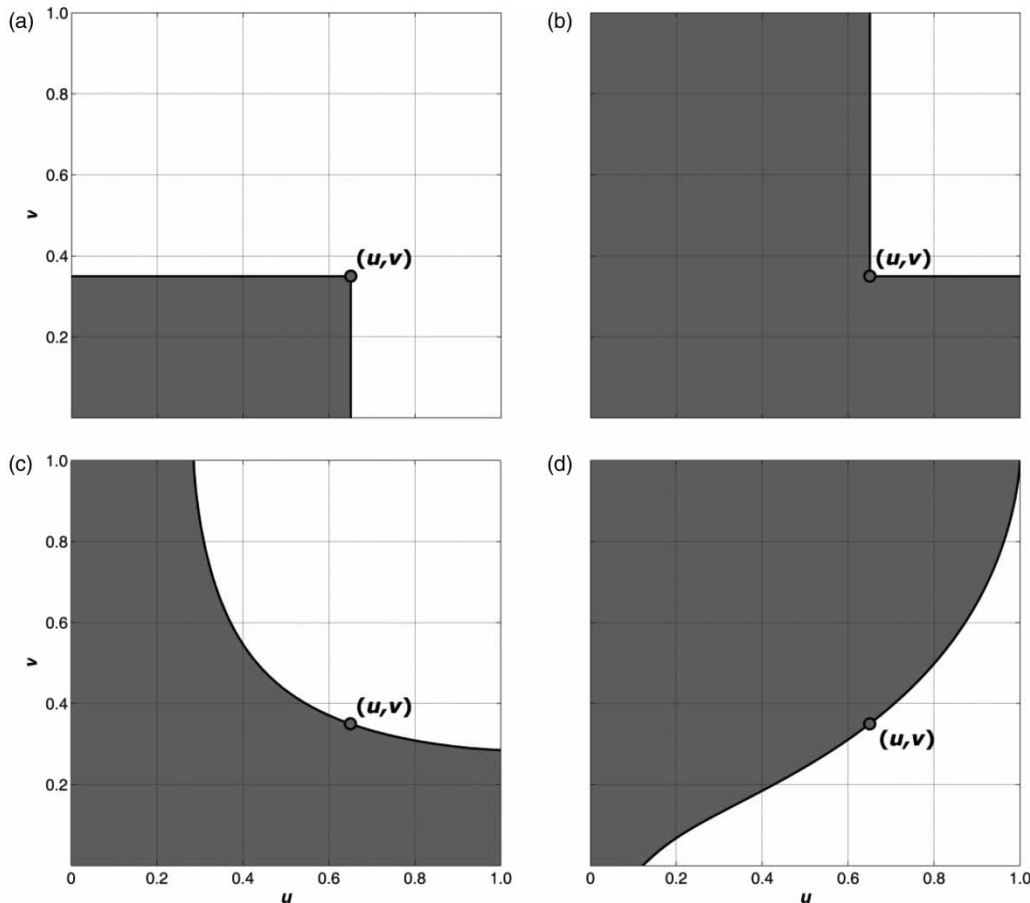
In general, the estimate of the return period  $T_r$  associated with a given value  $x$  of the random variable of interest  $X$  requires the population to be split in two dichotomic regions: the sub-critical one and the super-critical one. The first one collects the less severe events ( $X \leq x$ ), while the second one collects the more severe ones ( $X > x$ ). The return period  $T_r$ , defined as the average period elapsing between two subsequent occurrences of the expected event, is operatively estimated in (17) by using the non-exceedance probability  $P_X$  and the annual average number

of occurrences  $\omega$ :

$$T_r(x) = \frac{1}{\omega[1 - P_X(x)]} \quad (17)$$

Unfortunately, reproducing this simple procedure in a multivariate case is problematic, since a total order relation does not exist and the population splitting is not univocal. The oldest methods attempted to mimic the sub-critical subset definition of univariate analysis, by exploiting intuitive logical expressions. The concepts of 'OR' return period  $T_r^{OR}$  and 'AND' return period  $T_r^{AND}$  were thus suggested (Salvadori *et al.* 2007; Serinaldi 2015 and references therein).

In the  $T_r^{OR}$  formulation, a super-critical event occurs when at least one of the random variables defining the event of interest is exceeded; a graphical illustration of the following unitary square partition is given in Figure 1(a).



**Figure 1** | Sub-critical regions (shaded areas) of a generic bivariate event  $(u, v) = (0.65, 0.35)$  for the analysed test case by  $T_r$  estimate methods: (a)  $T_r^{OR}$ , (b)  $T_r^{AND}$ , (c)  $T_r^{KEN}$  and (d)  $T_r^{OP}$ .

Owing to copula function properties,  $C(u,v)$  can formally substitute the non-exceedance probability in Equation (17) yielding the  $T_r^{OR}$  estimate (18):

$$\begin{aligned} T_r^{OR}(u, v) &= \frac{1}{\omega(1 - \text{Prob}\{U \leq u \wedge V \leq v\})} \\ &= \frac{1}{\omega[1 - C(u, v)]} \end{aligned} \quad (18)$$

Otherwise, in the  $T_r^{AND}$  formulation, a super-critical event occurs when all the random variables defining the event of interest are exceeded (see Figure 1(b)). To estimate  $T_r^{AND}$ , the exceedance probability can be more conveniently exploited, as shown in Equation (19), where the survival copula  $\bar{C}$ , direct extension of the univariate survival probability function, is used. The survival copula  $\bar{C}$  is related to copula  $C$ , so that the last estimator in Equation (19) is obtained:

$$\begin{aligned} T_r^{AND}(u, v) &= \frac{1}{\omega \text{Prob}\{U > u \wedge V > v\}} \\ &= \frac{1}{\omega \bar{C}(1 - u, 1 - v)} \\ &= \frac{1}{\omega[1 - u - v + C(u, v)]} \end{aligned} \quad (19)$$

These formulations usually yield very different return period estimates and the real value is arbitrarily supposed to be included in this range. In addition, neither approach induces a dichotomic splitting of the population, as in the univariate case. Isolines of constant return period are actually given by copula contours, in the  $T_r^{OR}$  formulation, or survival copula contours in the  $T_r^{AND}$  one. Distinct events belonging to such lines therefore have the same return period, even if their sub-critical regions are different and partially overlap.

In order to overcome this crucial conceptual drawback, Salvadori & De Michele (2010) proposed to utilize the Kendall measure  $K_C$  (20). This function estimates the probability of occurrence of an event belonging to the region included between the lower-left corner of the unit square and the copula contour of level  $t = C(u,v)$ ; such a region is illustrated in Figure 1(c):

$$K_C(t) = \text{Prob}\{(U, V) \in \mathbb{I}^2 : C(U, V) \leq t\} \quad \text{with } t \in \mathbb{I} \quad (20)$$

The function  $K_C(t)$  is a univariate probability distribution exclusively depending on the copula and associated with a dichotomic splitting of  $\mathbb{I}$ , since all events belonging to the contour line of level  $t$  have the same sub-critical region. Unlike previous formulations, the sub-critical region depends on the copula function and its parameters. Formally, Equation (20) can be substituted in the univariate return period expression (17), yielding the estimate of the Kendall return period  $T_r^{KEN}$  (21):

$$T_r^{KEN}(u, v) = \frac{1}{\omega[1 - K_C(C(u, v))]} \quad (21)$$

For Archimedean copulas, function (20) is explicit and can be expressed in terms of generator function as shown in Equation (22):

$$K_C(t) = t - \frac{\psi(t)}{\psi'(t)} \quad (22)$$

Substituting the generator function (6) in (22), the simple Equation (23) of the Kendall function for Gumbel-Hougaard 2-copula is obtained:

$$K_C(t) = t - \frac{t}{\theta} \ln(t) \quad (23)$$

Owing to the increasing portion of  $\mathbb{I}$  that  $T_r^{OR}$ ,  $T_r^{KEN}$  and  $T_r^{AND}$  involve, respectively (compare Figure 1(a)-1(c)), larger non-exceedance probabilities are estimated, so that inequality (24) must hold:

$$T_r^{OR} < T_r^{KEN} \leq T_r^{AND} \quad (24)$$

Despite the fundamental progress in understanding and overcoming the inherent limitations of the previous ones,  $T_r^{KEN}$  estimate method demonstrates that it is affected by conceptual drawbacks and yields unacceptable estimates when applied to FFC derivation from the constructed JDF.

Indeed, in hydrologic practice, the problem of multivariate return period estimate arises when performances of any hydraulic device of interest are strongly sensitive to various characteristics of the input hydrologic event. When

such performances can be referred to a single derived variable, the multivariate return period estimate can be brought back to that of a univariate return period. In fact, according to the derived distribution theory, the non-exceedance probability of the derived variable must equal that of the input hydrological event.

The transformation function relating to the input variables and the derived variable, called by some authors structure function (Volpi & Fiori 2014; Salvadori *et al.* 2015; Pappadà *et al.* 2016), must be a subjective function such that a sub-set of the input variable population is related to a single occurrence of the derived variable. This sub-set separates the sub-critical region from the super-critical one, by discriminating input events leading to lower or greater values of the derived variable, respectively. It is apparent that these regions are dichotomous and that infinite input events are associated with a unique return period.

Following this criterion, for instance, in Requena *et al.* (2013), copula simulation techniques were exploited to generate from a bivariate distribution of peak discharge and runoff volume a number of flood events, forcing a real-world routing reservoir. Hydraulic simulations were utilized as transformation function to obtain maximum water levels occurring in the storage volume during the routing process. Thus, flood events were associated with the return period of corresponding maximum water levels, estimated by means of their empirical frequencies.

Conversely, Volpi & Fiori (2014) analysed the same problem from a theoretical point of view, by assuming a constant inflow discharge and a linear behaviour for the reservoir. Under such simplifying hypotheses, an analytically closed-form transformation function relating inflow peak discharge and runoff volume to routed peak discharge can be derived. Such a function was exploited to delimitate sub-critical regions in the population of inflow flood variables for constant values of routed peak discharge. The bivariate return period estimate was therefore carried out with reference to non-exceedance probabilities, obtained by integrating copula density functions suggested in the literature on such regions.

Differently, peak discharge must be accounted for when dealing with open channel design or safety verification. This variable depends on the rainfall volume and the wet weather duration and can be derived from their JDF. Further, all

rainfall events leading to an identical peak discharge can be associated with the same severity, or return period.

To derive peak discharges from an input rainfall event, a simplified lumped hydrological model, similar to those commonly adopted in practical applications of urban hydrology, is herein set up. In order to approximate the natural depletion of the hydrological losses during the wet weather period, the rainfall excess  $h_r$  (25) is evaluated by means of a runoff coefficient  $\Phi$  applied to rainfall portion exceeding the initial abstraction  $IA$ , chosen in accordance with the volume threshold of the discretization procedure. Consequently, the number of rainfall event  $\omega$  equals the number of flood events:

$$h_r = \Phi (h - IA) \quad (25)$$

Further, similarly to Wycoff & Singh (1976), the flood hydrograph shape is assumed to be triangular, with a base given by the sum of the rainfall excess duration  $d_r$  and the catchment time of concentration  $t_c$ . If a constant rainfall intensity is assumed,  $d_r$  can be estimated by expression (26):

$$d_r = d \frac{h - IA}{h} \quad (26)$$

The runoff volume (25) must equal the area of the flood hydrograph, so that the peak discharge  $q_p$  can be expressed in terms of the rainfall random variables  $h$  and  $d$  as shown below. Equation (27) represents an analytical transformation function relating to the bivariate input rainfall event and the peak discharge:

$$q_p(h, d) = \frac{2 \Phi h (h - IA)}{d (h - IA) + h t_c} \quad (27)$$

Through the derived distribution theory, the non-exceedance probability of the peak discharge  $P_{QP}$  can be expressed in the terms of the following equation. In the proposed estimate method, non-exceedance probability (28) is thus assigned to all infinite rainfall events leading to the same discharge peak  $q_p$  (27):

$$P_{QP}(q_p) = \text{Prob}\{Q_p \leq q_p\} = \text{Prob}\left\{\frac{2 \Phi h (h - IA)}{d (h - IA) + h t_c} \leq q_p\right\} \quad (28)$$



Such a formulation allows to define in  $\mathbb{I}^2$  the sub-critical region  $\Lambda^{QP}$  (29), if uniform random variables  $u$  and  $v$  are made explicit in relationship (27), by means of inverse CDFs (1).  $\Lambda^{QP}$  depends on marginals and on two hydrologic catchment parameters  $IETD$  and  $IA$  but, differently from the formulation relying on the Kendall function, it is independent of the underlying copula. An example of the unitary square splitting related to Equation (29) is given in Figure 1(d):

$$\Lambda^{QP}(u, v) = \{(U, V) \in \mathbb{I}^2: q_p(P_H^{-1}(U), P_D^{-1}(V)) \leq q_p(P_H^{-1}(u), P_D^{-1}(v))\} \quad (29)$$

The return period associated with a bivariate rainfall event  $T_r^{QP}$  can finally be estimated through the univariate return period associated with  $q_p$ , as shown in expression (30), where the non-exceedance probability of the peak discharge  $P_{QP}$  is estimated by integrating the copula density  $c$  over the super-critical region  $\Lambda^{QP}$  (29):

$$T_r^{QP}(u, v) = \frac{1}{\omega [1 - P_{QP}(u, v)]} = \frac{1}{\omega \left[ 1 - \int_{\Lambda^{QP}(u, v)} c(x, y) dx dy \right]} \quad (30)$$

The integral in (30) must be computed numerically, unless the JDF is constructed by coupling exponential functions through the independence copula and Equation (27) is simplified by assuming the wet weather duration to be equal to the rainfall excess duration, as in Balistrocchi *et al.* (2013). Numerical computing is, however, facilitated by operating in the unitary square, which allows only proper integrals to be used.

## CONTINUOUS SIMULATION MODEL

Hydrological simulations were conducted to obtain the continuous discharge series generated by the test catchment, as a result of the observed rainfalls. A conceptual lumped model was developed, aiming at limiting discrepancies with respect to the one implemented in the derivation of the  $T_r^{QP}$  expression. Hence, the rainfall–discharge

transformation process was again represented by the loss model (25), while the rainfall excess routing was executed by using a triangular instantaneous unitary hydrograph, with duration equal to  $t_c$ .

Finally, following the individual event statistics criterion, the simulation output was separated in independent floods, detecting their peak discharges and counting their average annual number  $\omega_j$ . Thus, the experimental return period  $\hat{T}_r^{QP}$  of peak discharges was estimated by relationship (31), where  $F_{QP}$  is the  $q_p$  Weibull plotting position. The  $IETDs$  adopted to account for the hydrologic losses and to identify the flood events were equal to that utilized in the previous rainfall discretization procedure. A very small discharge threshold was further used to delete from the flood event samples too small to be considered simulation nuisances:

$$\hat{T}_r^{QP} = \frac{1}{\omega_j(1 - F_{QP})} \quad (31)$$

## TEST WATERSHED

The JDF  $P_{HD}$  is herein constructed with reference to a 45-year long time series of observed rainfall depths, recorded every 30 minutes from 1949 to 1993 by the ITAS Pastori raingauge, located in Brescia, northern Italy, next to the transition between the Padan Plain and the southern foothills of the Alps. The rainfall regime is classified as sub-alpine, characterized by two maxima, in spring and autumn, and two minima, in winter and summer. During summer, short duration and high intensity storms separated by long dry weather periods occur, while longer but less intense precipitations are more common in the other seasons. The annual average rainfall depth is quite large with respect to other Italian climates and amounts to about 1.000 mm.

Bearing in mind small–medium size urban catchment applications, analyses referred to a synthetic test watershed characterized by this parameter set: catchment area  $A$  100 ha, time of concentration  $t_c$  20 minutes and hydrological losses expressed by an initial abstraction  $IA$  5 mm and a runoff coefficient  $\Phi$  0.45, for the exceeding portion. The

discretization of the continuous rainfall time series into independent rainfall events was hence conducted by using an *IETD* corresponding to the minimum suggested value of 3 h (Adams & Papa 2000) and a volume threshold equal to 5 mm.

## RESULTS AND DISCUSSION

The application of the discretization procedure to Brescia's rainfall time series yields a mean annual number  $\omega$  of 48.2 independent rainfall events. This result can be considered reasonable in this rainfall regime, owing to the very small *IETD* adopted for the test catchment.

The maximum likelihood criterion was employed to fit the theoretical copula (8) to pseudo-observations, estimating a value of 1.41 for the dependence parameter  $\theta$ . Although this value corresponds to a concordance coefficient  $\tau_K$  equal to 0.29, a significant association is however shown. Indeed, when a goodness-of-fit test is performed according to statistics (12) by assuming that the dependence structure is given by the independence copula, that is, disregarding the association degree, the null hypothesis can be rejected for a  $p$  value (13) less than 0.1%.

Conversely, the null hypothesis that the underlying copula is expressed by copula (8) cannot be rejected for a  $p$  value greater than 80.0%. Moreover, the upper tail dependence coefficient  $\lambda_U$  (10) corresponding to  $\theta$  estimate is 0.366, whereas the non-parametric estimate (14) yields a value of 0.346.

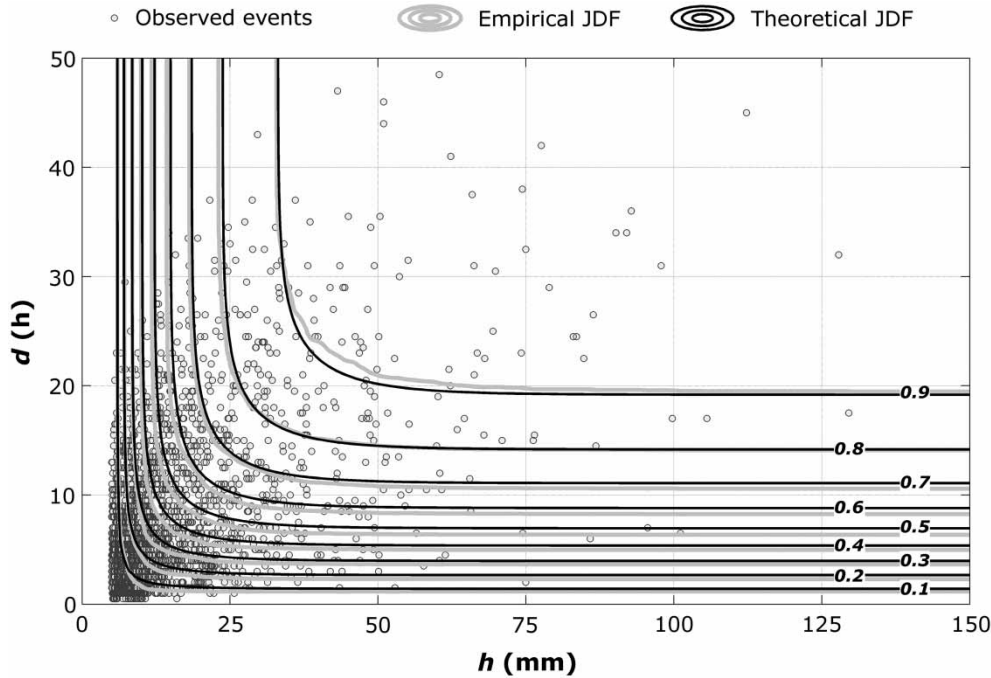
Marginals (15) and (16) were finally fitted to  $h$  and  $d$  samples by the maximum likelihood criterion, as well, leading to the following parameter set: 0.88, 1.18, 10.9 mm and 9.5 h for  $\beta$ ,  $\gamma$ ,  $\zeta$  and  $\lambda$ , respectively. A  $\beta$  value less than one evidences a significant presence of rainfall events having very small depth, while a  $\gamma$  value greater than one shows that wet weather durations tend to be less asymmetric. The goodness-of-fit was verified by means of the confidence boundary test, revealing that these distributions cannot be rejected for conventional levels of significance of 5–10%. When marginals are simplified in exponential distributions, shape parameters are set equal to one, while scale parameter estimates are: 12.0 mm for  $\zeta$  and 8.9 h for  $\lambda$ . However, such distributions can be rejected for a significance less than 1.0%.

On the whole, the JDF constructed by coupling, according to Equation (4), Weibull marginals (15) and (16) by means of the Gumbel–Hougaard copula (8) appears to be a suitable model to represent the bivariate variability of the rainfall volume and the wet weather duration in this rainfall regime. For the sake of completeness, a visual goodness-of-fit of the fitted JDF is proposed in Figure 2, where contour lines of the theoretical JDF are compared to those of the empirical JDF; the satisfactory agreement stated by performed test statistics is clearly evident.

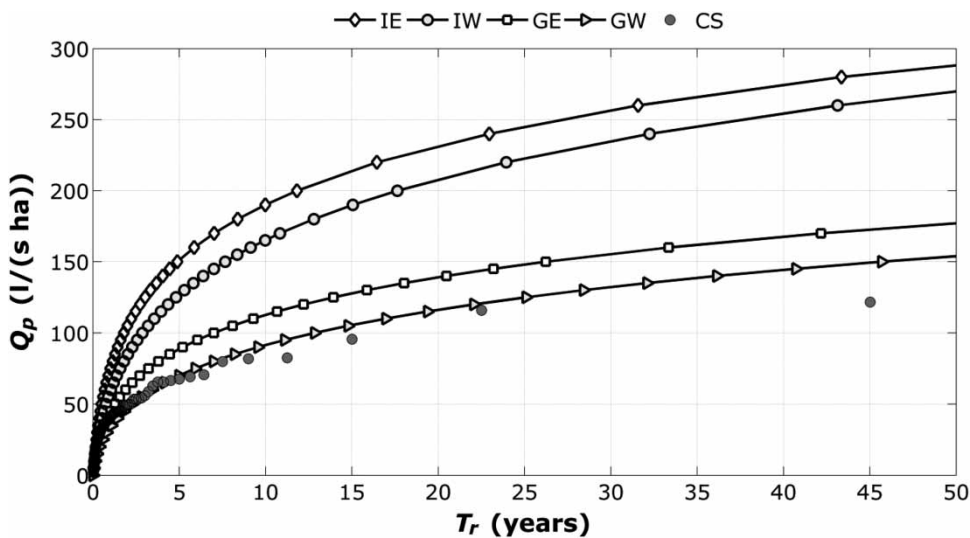
FFC curves derived for the test catchment through continuous simulations CS are drawn in Figure 3, along with those derived from the probabilistic approach by assuming diverse change in combinations of copula and marginals: independence copula and exponential marginals IE, independence copula and Weibull marginals IW, Gumbel–Hougaard copula and exponential marginals GE and, finally, Gumbel–Hougaard copula and Weibull marginals GW. In this chart, return period estimates are carried out only according to the proposed method  $T_r^{QP}$  (30).

As can be seen, a good agreement between CS FFC and GW FFC is evidenced. The largest residual, about 23%, is given by the peak estimate corresponding to the 45-year return period, that is the series length, and it can be considered acceptable. Instead, in the range of 5:20 years  $T_r$ , residuals are much smaller and the probabilistic model slightly overestimates CS FFC. This can be justified by considering that the rainfall series time step is comparable to  $t_c$ , so that simulated peak discharges are expected to be more attenuated than the real ones. However, GW FFC appears to be a little more conservative than benchmarking outcomes and its values match those suggested in the analysed rainfall regime for urban sewer design (a term of reference adopted by practitioners is about 100 l/(s ha) for  $T_r$  10 years). This result definitively supports both the reliability of the JDF based on such functions and the device-targeted method to estimating the return period of bivariate rainfall events (30) herein developed.

On the contrary, in the other couplings, unacceptably large estimates are obtained. The worse result is provided by IE FFC, since almost double peak discharges than the benchmarking ones are derived. Hence, the most common assumptions on which analytical-probabilistic models commonly rely, in this situation, do not yield the



**Figure 2** | Contour lines of the theoretical JDF and of the empirical JDF, obtained for rainfall events derived from continuous rainfall series by using  $IETD = 3$  h and  $IA = 5$  mm.

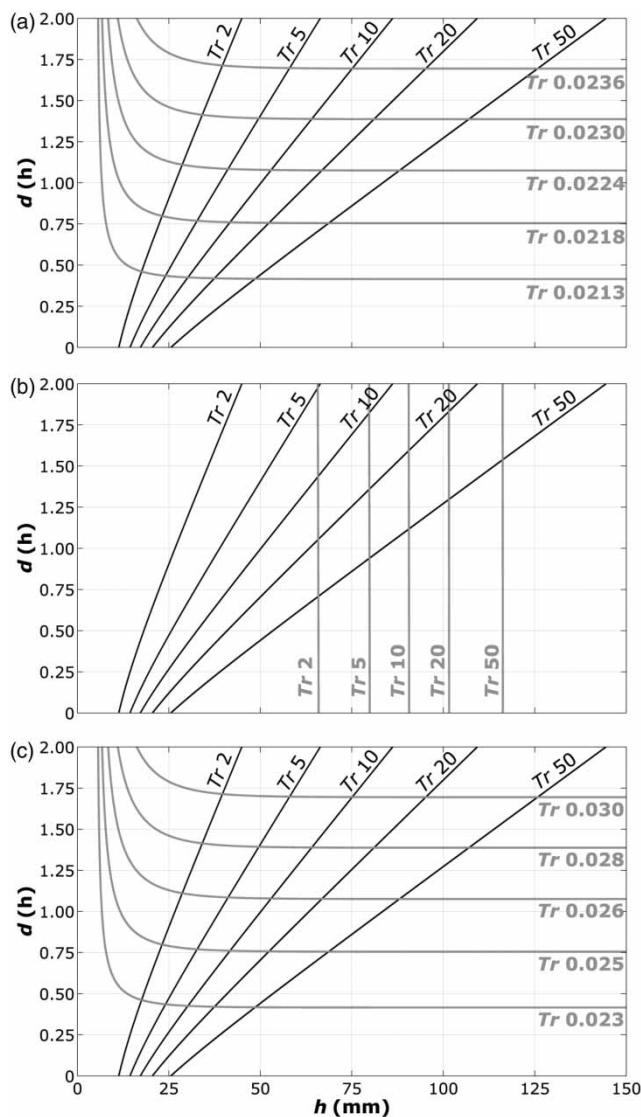


**Figure 3** | FFCs derived according to various methods: continuous simulations CS, probabilistic models derived by independence copula and exponential marginals IE, independence copula and Weibull marginals IW, Gumbel-Hougaard copula and exponential marginals GE and Gumbel-Hougaard copula and Weibull marginals GW.

compensation of errors occurring elsewhere. The most significant reliability improvement is due to a correct representation of the dependence structure rather than a better fitting of marginals. In fact, GE FFC is much closer to the CS FFC than IW FFC. This occurrence may be

explained by considering that, even if rejected by test statistics, exponential distributions approximately suit the empirical marginal distributions, while the observed dependence structure is substantially different from the independence copula.

Isolines of bivariate rainfall events characterized by constant  $T_r^{QP}$ , which at this stage may be considered realistic, are compared to those derived for constant  $T_r^{OR}$ ,  $T_r^{AND}$  and  $T_r^{KEN}$ , respectively (Figure 4(a)–4(c)). The catchment time of concentration  $t_c$  being equal to 20 minutes, sub-hourly wet weather durations are basically of interest for open channel applications. A broader range up to 2 h has, however, been considered, to provide a more comprehensive illustration.



**Figure 4** | Detail of isolines for constant  $T_r$  (years) obtained by  $T_r^{QP}$  method (black), compared to those (grey) obtained by using (a)  $T_r^{OR}$ , (b)  $T_r^{AND}$  and (c)  $T_r^{KEN}$  methods.

First, it is well known that, for a given  $T_r$ , the rainfall volume is expected to increase with respect to the wet weather duration. This behaviour is respected by  $T_r^{QP}$  isolines, which also provide a rainfall volume greater than zero for null durations. These trends are a consequence of the super-critical region delimited by the  $T_r^{QP}$  method, which is mainly located in the lower-right corner of the unitary square. As can be seen in the example in Figure 1(d), in this approach, super-critical events are those featured by large depth but short duration, which exceed the hydrologic losses; the conceptual soundness of the proposed model is therefore further supported.

The extremely large peak discharge obtained from JDFs based on the independence copula can be explained, as well. In the independent copula, the copula density is uniformly distributed on  $\mathbb{I}$ , while, in the Gumbel–Hougaard copula, it concentrates on the main diagonal in the lower-left and in the upper-right corners. Thus, in the first case, the non-exceedance probability is underestimated, consequently leading to an excessively low return period.

When return periods are estimated by using  $T_r^{OR}$  and  $T_r^{KEN}$  methods very similar outcomes are obtained. Isolines exhibit a decreasing trend of the rainfall volume with respect to the wet weather duration. In addition, since return period isolines correspond to JDF contour lines, a significant range of short durations next to the axis origin is excluded: the greater the return period, the larger this range is. On the contrary, the wet weather duration range must span from zero to infinite. Furthermore, in both methods, estimated return periods are evidently meaningless. More reasonable values are instead assessed by means of the  $T_r^{AND}$  method, although the rainfall volumes appear to be almost independent of the wet weather duration. This occurs because, in the region taken into consideration, survival copula contour lines are almost straight and parallel to wet weather duration axis.

Such huge discrepancies between  $T_r^{QP}$  and the others can be explained by the substantially different splitting of the bivariate population, by which they estimate the non-exceedance probability. The extreme sensitivity of the return period with respect to this estimate further accentuates such discrepancies. In complete agreement with Serinaldi (2015),  $T_r^{OR}$ ,  $T_r^{AND}$  and  $T_r^{KEN}$  methods can be adopted only if they match the failure mechanism of the analysed device.

For example, fewer discrepancies were found by Requena *et al.* (2013), when the estimate method based on routing reservoir performances is compared to  $T_r^{OR}$  and  $T_r^{KEN}$ . In this application, peak discharge must actually decrease with the runoff volume when the maximum water level, or the routed peak discharge, is constant. Therefore, the failure mechanisms on which  $T_r^{OR}$  and  $T_r^{KEN}$  rely are more similar to the device one. Conversely, as in the problem dealt with in this study, they do not provide suitable return period estimates.

## CONCLUSIONS

In this paper, a FFC for a small–medium size urban catchment in a northern Italian rainfall regime has been derived, from a JDF of rainfall depths and wet weather durations. To do so, a Gumbel–Hougaard copula and Weibull marginals were combined and a bivariate return period estimate method was developed, by using a lumped conceptual scheme for the rainfall–runoff transformation and routing processes. The proposed model soundness is supported by a satisfactory agreement with benchmarking continuous simulations. Two main consequences can be drawn from this.

First, although a method of general applicability to estimating multivariate return periods cannot reasonably be established, it is however possible to delineate a common strategy, addressing the real-world application dynamics. In this circumstance, if device performances can be referred to a single constituent variable, the derived distribution theory allows the multivariate case to be traced back to the univariate one. In fact, with regard to the derived hydrologic variable, the population of the input random variable can effectively be split in the dichotomous regions, collecting sub-critical and super-critical events. Furthermore, it is important to point out that, in a multivariate framework, defining a single change in event for a given return period appears to be meaningless. When an estimate method is chosen, a single event corresponds to a unique return period but, on the contrary, infinite events share the same return period.

Second, the utilization of the copula approach in the rainfall JDF construction demonstrated to yield a substantial improvement in the overall model reliability. Differently from what is reported in the literature concerning the analytical-probabilistic modelling, JDFs properly accounting

for the observed dependence structure of random variables, featuring a non-negligible association, performed better than those assuming their independence. Nonetheless, in the latter case, more conservative results are obtained. In any case, the model choice should be carried out carefully, balancing the definitive advantage of an analytical formulation and the real need for accuracy improvement following a more precise probabilistic model.

Future developments regarding the method herein proposed to estimating the multivariate return period are still desirable. On the one hand, the FFC derivation could be generalized towards natural watershed applications, whose scale and hydrological processes require more complex modelling. On the other hand, it could be attractive to more deeply understand, with regard to real-world test cases, whether the method can be suitably exploited to face other multivariate problems, or not. Indeed, a number of hydraulic design and verification problems, in which the hydrologic input would be more efficaciously outlined by a multivariate continuous approach, exists, namely landslide triggering risk assessment, routing reservoir design, spillway safety verification and flooding area delimitations.

## REFERENCES

- Adams, B. J. & Howard, C. D. D. 1986 [Design storm pathology](#). *Can. Water Resour. J.* **11** (3), 49–55.
- Adams, B. J. & Papa, F. 2000 *Urban Stormwater Management Planning with Analytical Probabilistic Models*. John Wiley & Sons, New York.
- Bacchi, B., Brath, A. & Maione, U. 1993 Sul dimensionamento delle reti di drenaggio con la metodologia dell'evento critico. *Idrotecnica* **1**, 33–43.
- Balistrocchi, M. & Bacchi, B. 2011 [Modelling the statistical dependence of rainfall event variables through copula functions](#). *Hydrol. Earth Syst. Sci.* **15** (6), 1959–1977.
- Balistrocchi, M., Grossi, G. & Bacchi, B. 2009 [An analytical probabilistic model of the quality efficiency of a sewer tank](#). *Water Resour. Res.* **45** (12), W12420.
- Balistrocchi, M., Grossi, G. & Bacchi, B. 2013 [Deriving a practical analytical-probabilistic method to size flood routing reservoirs](#). *Adv. Water Resour.* **62**, 37–46.
- Candela, A., Brigandi, G. & Aronica, G. T. 2014 [Estimation of synthetic flood design hydrographs using a distributed rainfall-runoff model coupled with a copula-based single storm rainfall generator](#). *Nat. Hazard. Earth Sys.* **14** (7), 1819–1833.

- Deheuvels, P. 1979 Empirical dependence function and properties: nonparametric test of independence. *Bulletin de la Classe des Sciences Academie Royale de Belgique* **65** (6), 274–292.
- De Michele, C. & Salvadori, G. 2003 A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.* **108** (2), D2.
- Eagleson, S. P. 1972 Dynamics of flood frequency. *Water Resour. Res.* **8** (4), 878–898.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N. & Bobée, B. 2004 Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* **40** (1), W01101.
- Frahm, G., Junker, M. & Schmidt, R. 2005 Estimating the tail-dependence coefficient: properties and pitfalls. *Insur. Math. Econ.* **37** (1), 80–100.
- Genest, C. & Favre, A.-C. 2007 Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.* **12** (4), 347–368.
- Genest, C., Rémillard, B. & Beaudoin, D. 2009 Goodness-of-fit tests for copulas: a review and a power study. *Insur. Math. Econ.* **44** (2), 199–213.
- Gräler, B., Van Den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B. & Verhoest, N. E. C. 2013 Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol. Earth Syst. Sci.* **17** (4), 1281–1296.
- Guo, Y. & Adams, B. J. 1998 Hydrologic analysis of urban catchments with event-based probabilistic models. 2. Peak discharge rate. *Water Resour. Res.* **34** (12), 3433–3443.
- Guo, Y., Liu, S. & Baetz, B. W. 2012 Probabilistic rainfall-runoff transformation considering both infiltration and saturation excess runoff generation processes. *Water Resour. Res.* **48** (6), W06513.
- Nelsen, R. B. 2006 *An Introduction to Copulas*. Springer, New York.
- Pappadà, R., Perrone, E., Durante, F. & Salvadori, G. 2016 Spin-off extreme value and archimedean copulas for estimating the bivariate structural risk. *Stoch. Environ. Res. Risk Assess.* **30** (1), 327–342.
- Poulin, A., Huard, D., Favre, A. C. & Pugin, S. 2007 Importance of tail dependence in bivariate frequency analysis. *J. Hydrol. Eng.* **12** (4), 394–403.
- Requena, A. I., Mediero, L. & Garrote, L. 2013 A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation. *Hydrol. Earth Syst. Sci.* **17** (8), 3023–3038.
- Ruymgaart, F. H. 1973 *Asymptotic Theory for Rank Tests for Independence*. MC Tract 43, Mathematisch Instituut, Amsterdam.
- Salvadori, G. & De Michele, C. 2010 Multivariate multiparameter extreme value models and return period: a copula approach. *Water Resour. Res.* **46** (10), W10501.
- Salvadori, G., De Michele, C., Kottegoda, N. T. & Rosso, R. 2007 *Extremes in Nature: An Approach Using Copulas*. Springer, Dordrecht.
- Salvadori, G., Durante, F., Tomasicchio, G. R. & D'Alessandro, F. 2015 Practical guidelines for the multivariate assessment of the structural risk in coastal and off-shore engineering. *Coast. Eng.* **95**, 77–83.
- Serinaldi, F. 2015 Dismissing return periods! *Stoch. Environ. Res. Risk Assess.* **29** (4), 1179–1189.
- Sklar, A. 1959 Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.
- 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.
- Wycoff, R. L. & Singh, U. P. 1976 Preliminary hydrologic design of small flood detention reservoirs. *Water Resour. Bull.* **12** (2), 337–349.
- Yen, B. C. 1990 Return period, risk and probability in urban storm drainage. *From the Experience of the 20th Century to the Science of the 21st Century. Proceedings of the 5th Conference on Urban Storm Drainage*, Osaka, Japan, pp. 59–72.

First received 15 April 2016; accepted in revised form 5 November 2016. Available online 9 February 2017