



**HAL**  
open science

# Extreme rainfall analysis by a stochastic model: impact of the copula choice on the sub-daily rainfall generation

P. Cantet, P. Arnaud

## ► To cite this version:

P. Cantet, P. Arnaud. Extreme rainfall analysis by a stochastic model: impact of the copula choice on the sub-daily rainfall generation. *Stochastic Environmental Research and Risk Assessment*, 2014, 28 (6), p. 1479 - p. 1492. 10.1007/s00477-014-0852-0 . hal-01016960

**HAL Id: hal-01016960**

**<https://hal.science/hal-01016960>**

Submitted on 1 Jul 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

<b>Stochastic Environmental Research and Risk Assessment manuscript No.</b> (will be inserted by the editor)
---

---

## **Extreme rainfall analysis by a stochastic model : impact of the copula choice on the sub-daily rainfall generation**

**Philippe Cantet · Patrick Arnaud**

Received: date / Accepted: date

**Abstract** The hourly rainfall stochastic model SHYPRE is based on the simulation of descriptive variables. It generates long series of hourly rainfall and enables an at-site empirical estimation of distribution quantiles over France. The present study focuses on the improvement of the rainfall generator by modelling storm characteristics dependence by the copula approach. An evaluation framework is proposed to evaluate the goodness-of-fit of a given method over a territory with a particular care for the extreme part of the distribution. It is used to illustrate the impact of the copula choice on the estimation of rainfall quantiles. Contrary to Clayton copula, both the Gumbel's and Frank's permit to improve significantly the performance of the model in the sub-daily rainfall generation. According to our criteria, the final version of SHYPRE proposes a better estimation of rainfall quantiles than the classical extreme value distributions.

### **1 Introduction**

The utilization of stochastic models in a hydrological framework was introduced by (Eagleson 1972). He derived the peak flow rate frequency from average intensity and duration of storm, by considering the two random variables independent and exponentially distributed. This paper stimulated many subsequent works aimed at various purposes in which same hypotheses are assumed (Eagleson 1978a,b,c; Córdova and Rodríguez-Iturbe 1985; Díaz-Granados et al. 1984; Guo and Adams 1999; Li and Adams 2000).

---

P. Cantet, P. Arnaud

IRSTEA, 3275 route de Cézanne CS 40061, 13182 Aix-en-Provence, Cedex 5, FRANCE

Tel.: +33 (0)4.42.66.99.47

Fax: +33 (0)4.42.66.99.05

E-mail: [philippe.cantet@irstea.fr](mailto:philippe.cantet@irstea.fr)

Even if these papers led to remarkable results, observed data statistics undermined the assumption of independence between the depth (or intensity) and the duration of a rainfall. However, Adams (2000) compared analytical models by assuming both dependent and independent rainfall characteristics and showed that models have better performances and more conservative results by neglecting the association among the random variables. These results might be explained by the selection of an inappropriate dependence model.

The joint probability function allows modelling dependence between hydrological variables (Goel et al. 2000; Kurothe et al. 1997). The main limitation of this approach is that the individual behavior of the variables (marginal distributions) must be characterized by the same parametric family of univariate distributions. Exponential marginal distribution is generally used to model the intensity or duration of rainfall (Singh and Singh 1991; Bacchi et al. 1994). However, the exponential function does not always fit the sample distributions exactly and distinct marginal probability functions may be needed for the variables (Salvadori and De Michele 2006; Haberlandt et al. 2008).

An opportunity to overcome these modelling drawbacks has been achieved using copula functions introduced by (Hoeffding 1940; Sklar 1959). Copulas are functions that join or “couple” multivariate distribution functions to their one-dimensional marginal distribution functions (Nelsen 2006). Starting with the papers of De Michele and Salvadori (2003) and Favre et al. (2004), copula models have become more and more widespread in hydrological models (Salvadori and De Michele 2004; De Michele et al. 2005; Zhang and Singh 2007; Salvadori et al. 2007; ?) to improve their performance (Vandenberghe et al. 2011). The flexibility of copulas can be applied on different topics. Salvadori and De Michele (2006); Gargouri-Ellouze and Chebchoub (2008); Vandenberghe et al. (2010) used them to associate storm characteristics in a rainfall model while copulas make it possible to simulate space-time rainfall for several stations in Haberlandt et al. (2008); Bárdossy et al. (2009); Ghosh (2010); Salvadori et al. (2011). Most of the models described in these papers are tested on one station alone or on several stations subject to the same precipitation regime. Balistrocchi and Bacchi (2011) proposed similar marginal distributions and the same dependence structure to reproduce three Italian rainfall time series. The Frank, Clayton and Gumbel copulas often appear in the literature (Salvadori and De Michele 2006; Gargouri-Ellouze and Chebchoub 2008; Balistrocchi and Bacchi 2011) and are used to generate correlated rainfall characteristics. This specific selection was performed

---

because they allow to model no dependence tail (Frank's), a lower dependence tail (Clayton's) or an upper dependence tail (Gumbel's).

The proposed model, already presented in (Cernesson et al. 1996; Arnaud and Lavabre 1999, 2002; Arnaud et al. 2007), is designed for estimating extreme rainfall quantiles of different durations (1 to 72 hours). Like in Wu et al. (2006), it is applicable for simulating rainstorm at different sites. The model structure (marginal distribution functions of rainstorm characteristics or relationships between them) is the same for any station, shifting from one climate to another is possible uniquely based on the model's parameters. Arnaud et al. (2007) highlighted that the model can reproduce extreme rainfall for all types of climate by adding a dependence structure between the depths of successive rainstorms. The current version of the model has been regionalized on French territory providing a knowledge of the rain risk on ungauged sites (Arnaud et al. 2006) and reproduced in a satisfactory way the standard and extreme statistics of long duration maximum rainfall ( $\geq 24 h$ ) (Muller et al. 2009; Neppel et al. 2007; Carreau et al. 2013). However, these regional studies focused only on the reproduction of the daily rainfall, neglecting the hourly time-step though fundamental to reproduce observed flood peaks.

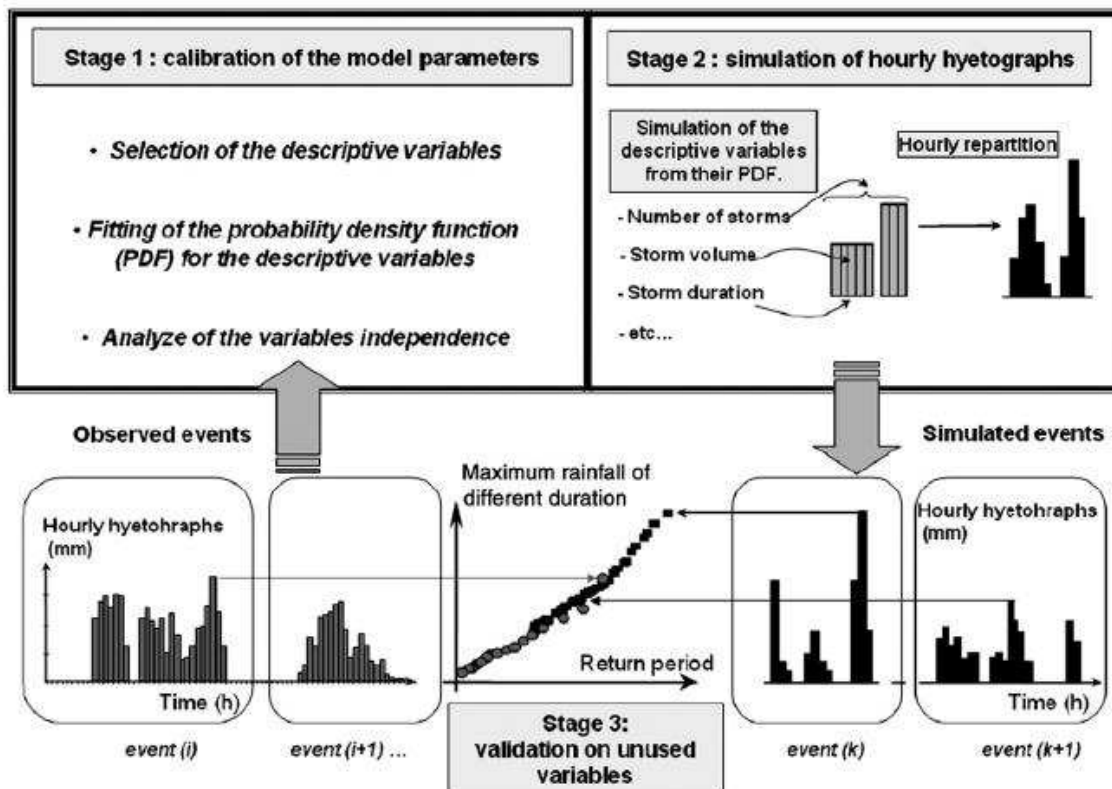
Two performance criteria are proposed to evaluate the goodness-of-fit with particular care for the extreme part of the distribution and permit to judge the reliability of a given method on a territory (Renard et al. 2013). This evaluation framework highlights that the sub-daily rainfalls generated by the model do not properly respect observations on several sites. To improve the generation of the sub-daily rainfall, the copula theory is applied to generate correlated rainfall characteristics, especially the depth and duration of a rainstorm. The paper focuses on the impact of the copula choice on the model performance to reproduce extreme rainfall for different time steps (1 to 72 hours). The criteria used to evaluate the model performance lead to the optimum copula choice out of three investigated families, which is not rejected by a goodness-of-fit test, for modelling the dependence between the depth and the duration of rainfall occurring in two seasons on several stations subject to different precipitation regimes.

## **2 The rainfall generator: SHYPRE**

This Section briefly presents the rainfall generator: principle and variables. For further details, a methodological guide has been published (Arnaud and Lavabre 2010) in French language. Arnaud et al. (2007) can be considered as the referential scientific paper about SHYPRE written in English.

## 2.1 The principle

SHYPRE is a sequential model of hydrograph simulation based on an hourly rainfall generation. It was developed at IRSTEA in Aix-en-Provence and can be coupled with a rainfall-runoff model (Cernesson 1993; Arnaud 1997). This generator is of the aggregation type and models only intense rainfall events. Descriptive variables are used to define the hourly rainfall signal into a rainfall event. Each variable was fitted by a probability law (Cernesson et al. 1996). Monte Carlo methods were used to reproduce the rainfall signal from the generation of these variables. Then time series, statistically equivalent to observations, can be reproduced for any desired time period. Quantiles are empirically estimated from these simulated times series. The robustness and the accuracy of these quantiles has been tested for the daily rainfall (Muller et al. 2009; Neppel et al. 2007). Figure 1 illustrates the generator's principle.



**Fig. 1** Principle of the hourly rainfall generator. Figure from (Arnaud et al. 2007).

## 2.2 Generator's descriptive variables

First, the descriptive analysis of rainfall was based on rainfall events selected on daily criteria, i.e. a succession of daily rainfall depths of more than 4 mm, including one daily rainfall depth of at least 20 mm<sup>1</sup>. The selection threshold of 20 mm leads to the determination of a first parameter, the average number of events per year ( $\overline{NE}$ ), that strongly varies according to the climate zone. We precisely chose to keep the same selection criterion for the rainy events to make a homogeneous analysis on a same territory.

Based on these events, selected at daily intervals, the hourly rainfall signal is characterized by seven other descriptive variables. These variables are the number of rainy periods within an event ( $N_{RP}$ ), the number of storms within a rainy period<sup>2</sup> ( $N_S$ ), and the dry duration that separates it from the next rainy period  $D_{RP}$ . The storm is the basic entity for the analysis of rainfall events, and is defined as a succession of hourly rainfall accumulations with a single local maximum. Each storm is characterized by its duration ( $D_S$ ) and its volume ( $V_S$ ). The quantitative analysis of storm volumes and durations showed the need to distinguish two types of storm called “major” and “ordinary” storms, and therefore to create a storm typology based on a daily criterion (Fine and Lavabre 2002). This storm typology enables us to extract the main information from rainfall modelling (Arnaud et al. 2007). Furthermore, two other variables have been introduced to characterize the hourly rainfall itself: the ratio between the hourly peak of the storm and its volume ( $1/D_S \leq RX_S \leq 1$ ) and the relative position of the maximum ( $1 \leq RPX_S \leq D_S$ ). These allow for a satisfactory representation of the different hourly rainfall patterns. Figure 2 illustrated an example of a rainy event where the different descriptive variables are presented.

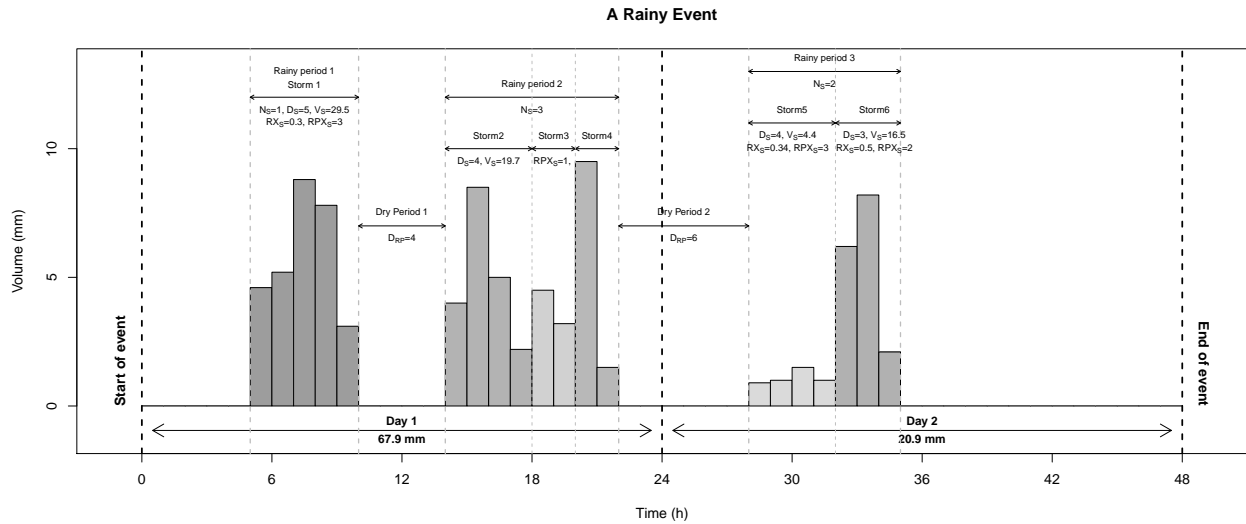
## 2.3 Model calibration

A first study carried out by Cernesson et al. (1996) determined the most adapted probability laws to the various descriptive variables. Because of the aim of the model regionalization (realized on the whole French territory), the same theoretical law has been defined to generate a given descriptive variable, whatever the studied station. For example, an exponential law has been chosen for the storm volume, and Poisson's law for the duration storm, whatever the studied station. Only parameters of these probability laws distinguish

---

<sup>1</sup> The threshold 20 mm is an assumption in order to take into consideration at least 5 events per year and to focus on extreme events.

<sup>2</sup> There is no dry period inside a rainy period.



**Fig. 2** Illustration of a rainy event (88.8 mm in 2 days) with 3 rainy periods ( $N_{RP} = 3$ ) and 6 storms (including 2 “major” storms: Storm1 and Storm2).

the climate (Arnaud et al. 2007). Calibrating the generator consists in estimating different parameters of the chosen probability laws with observed rainfall in a given rain gauge station. 20 parameters are required to fully calibrate the rainfall generator for two different seasons namely the “Winter” season from December to May and the “Summer” season from June to November. These were chosen for a maximum differentiation of the precipitation regimes. Note that some of the 20 parameters either vary slightly or have very little impact on the results.

#### 2.4 Simulation and rainfall quantiles estimation

After the model calibration, in order to simulate a rainy event, all descriptive variables are generated in a specific order (See Appendix A). Many rainy events are created to build time series as long as wanted in which the average number of observed events per year for each season are respected. Quantiles are empirically estimated from these simulated times series which are an interpretation of the phenomenon<sup>3</sup>. To reduce the sampling effect on the simulated events, rainfall is generated on periods which were a thousand times longer than the strongest wanted return period. For example, a 100 yr-quantile is determined by generated hyetographs on a 100 000 yr simulation period. It can be robustly estimated since it corresponds to the  $1000^{th}$  order statistics of this simulated series.

<sup>3</sup> Of course, no model can represent completely a real world phenomenon, it is a simplified reflection of reality.

At the beginning, the descriptive variables of the model were considered statistically independent. Many studies highlighted that some variables are dependent according to observations and that the dependence modelization is needed in order to reproduce the rainfall signal. Indeed, Arnaud et al. (2007) shows that the model can reproduce extreme rainfall for all types of climate by adding a dependence structure between the depths of successive rainstorms. In this paper, we focus on the dependence between two variables: the depth and the duration of a rainstorm.

## 2.5 An operational model

Prima facie, SHYPRE appears to be a complex model due to the number of variables or the different typologies used to define them. Nevertheless, an effort has been made to simplify it enabling an application on many hydrological problems. For example, Cantet et al. (2011) detected climate change impact on extreme rainfall throughout the model parameters; the SHYPRE outputs are also used to determine the dimension of a dam in (Carvajal et al. 2009) or to estimate the occurrence frequency of rainfall observed with a radar (Fouchier 2007) or in a flash flood warning (Javelle et al. 2010). A regionalized version of the model allows the estimation of rainfall quantiles for different time durations on a square of  $1\text{ km}^2$  everywhere in the French territory (Arnaud et al. 2006).

## 3 Modelling Dependence

The aim of this part is to introduce the mathematical tools to model a dependence between random variables, especially copulas. For further information, see Nelsen (2006) and Genest and Favre (2007).

### 3.1 Measuring dependence: Kendall's tau

Suppose that a random sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  is given from some pair  $(X, Y)$  of continuous variables. The empirical version of Kendall's tau,  $0 \leq \tau \leq 1$ , is given by:

$$\tau_n = \frac{P_n - Q_n}{n(n-1)/2} = \frac{4}{n(n-1)}P_n - 1 \quad (1)$$

where  $P_n$  and  $Q_n$  are the number of concordant and discordant pairs, respectively. Here, two pairs  $(X_i, Y_i), (X_j, Y_j)$  are said to be concordant when  $(X_i - X_j)(Y_i - Y_j) > 0$ , and discordant when  $(X_i - X_j)(Y_i - Y_j) < 0$ .



If  $X$  and  $Y$  are mutually independent, we have  $\tau_n \approx 0$ . An independence test can be based on  $\tau_n$ , since under  $H_0$ : independence between two variables, this statistic is close to normal with zero mean and variance  $2(2n + 5)/(9n(n + 1))$  (Genest and Favre 2007).

With discrete variables, this statistical test is biased by the ties. An unbiasing test consists in replacing  $n$  by  $n -$  number of ties in the variance calculus under  $H_0$ . However, this case will be discussed further.

### 3.2 Modelling dependence: copula approach

For simplicity purposes, we restrict attention to the bivariate case in this paper.

Any joint cumulative distribution function  $F_{XY}$  of a pair of continuous random variables  $(X, Y)$  can be written in the form

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)), \quad \forall x, y \in \mathbb{R} \quad (2)$$

where  $F_X$  and  $F_Y$  are the marginal functions and  $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$  is a copula. Sklar (1959) showed that  $C$ ,  $F_X$ , and  $F_Y$  are uniquely determined when  $F_{XY}$  is known, a valid model for  $(X, Y)$  arises from Eq. (2) whenever the three “ingredients” are chosen from given parametric families of distributions. The main advantage of the copula approach is that the choice of the dependence model between  $X$  and  $Y$  does not depend on the marginal distributions.

In this paper, only three archimedean copulas are considered: the Frank copula (Frank 1979), the Clayton copula (Clayton 1978), and the Gumbel copula (Gumbel 1961) for which their  $C(u, v)$  are defined in Table 1. The three copulas considered in this paper have a symmetric dependence structure -ie-  $C(u, v) = C(v, u)$ ,  $\forall u, v \in [0, 1]$ . Tests are proposed for the hypothesis that the underlying copula of a continuous random pair is symmetric (Genest et al. 2012).

### 3.3 Tail dependence

The tail dependence is an important concept in the multivariate extreme value theory. Contrary to the kendall's tau for which its estimation takes into account all realizations of the pair  $(X, Y)$ , the tail dependence focus on extreme values. The upper (resp. lower) tail dependence noted  $\lambda_U$  (resp.  $\lambda_L$ ) measures the

dependence between the variables in the upper-right (resp. bottom-left) quadrant and is defined by :

$$\begin{aligned}\lambda_U &:= \lim_{u \nearrow 1} \mathbb{P}(Y > F_Y^{-1}(u) | X > F_X^{-1}(u)) = \lim_{u \nearrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \\ \lambda_L &:= \lim_{u \searrow 0} \mathbb{P}(Y \leq F_Y^{-1}(u) | X \leq F_X^{-1}(u)) = \lim_{u \searrow 0} \frac{C(u, u)}{u}\end{aligned}\tag{3}$$

These parameters are nonparametric and depend only on the copula  $C$  of  $X$  and  $Y$ . The interpretation of the coefficients of tail dependency is that it measures the probability of two random variables both taking extreme values. If  $\lambda_U$  in  $(0, 1]$ , we say  $C$  has an upper tail dependence; if  $\lambda_U = 0$ , we say  $C$  has no upper tail dependence -ie- the asymptotic behavior of  $C$  is identical to the independence copula's; and similarly for  $\lambda_L$ . It is not tricky to evaluate both these coefficients for the Gumbel, Frank and Clayton copulas (see Table 1).

For illustration purposes only. We want to generate a couple  $(u, v)$  which respect a kendall's tau of 0.1, then the realization of the couple  $(u > 0.9999, v > 0.9999)$  is about 1000 times more likely with a Gumbel copula (upper dependence tail) than with a Frank copula (no dependence tail), whereas  $\mathbb{P}(u > 0.9, v > 0.9)$  are in the same order of magnitude for the both copulas.

### 3.4 Estimation and choice of models

Like usual statistic laws, different methods are used to estimate copula parameters (Genest et al. 1995; Joe 1997; Tsukahara 2005; Genest et al. 2008a). In this study, the copula parameter is estimated by the inversion of Kendall's tau. Analytic relations between these two quantities are illustrated in Table 1 for the 3 considered copulas.

Copulas	$C_\theta(u, v)$	Parameter $\theta$	Kendall's tau	$\lambda_L$	$\lambda_U$
Independence	$uv$	No	—	0	0
Frank	$\frac{1}{\theta} \ln \left( 1 + \frac{(e^{\theta u} - 1)(e^{\theta v} - 1)}{(e^\theta - 1)} \right)$	$\mathbb{R}^*$	$\tau(\theta) = 1 - \frac{4}{\theta} + \frac{4}{\theta^2} \int_0^\theta \frac{t}{e^t - 1} dt$	0	0
Gumbel	$\exp \{ - [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} \}$	$\theta \geq 1$	$\tau(\theta) = 1 - \frac{1}{\theta}$	0	$2 - 2^{1/\theta}$
Clayton	$(u^{-1/\theta} + v^{-1/\theta} - 1)^{-\theta}$	$\theta \geq -1$	$\tau(\theta) = \frac{\theta}{\theta + 2}$	$2^{-1/\theta}$	0

**Table 1** Definition of the four archimedean copulas with their parameter  $(\theta)$  space, and an expression for the population value of Kendall's tau  $(\tau)$ , and an expression for the tail dependence parameters  $(\lambda_L, \lambda_U)$  defined by Eq.3.

In typical modelling exercises, the user can choose between many different dependence structures. Consequently, a method is necessary to select, among different copulas, the best adapted dependence structure for the studied data. Genest et al. (2008b) compared a lot of measures to choose the best copula. A bootstrap procedure proposed in (Genest and Rémillard 2008) can be used to test the goodness-of-fit of copulas. This test has been implemented : "gofcopula" function in the "copula" package (Yan 2007; Ivan Kojadinovic and Jun Yan 2010) from the language R (R Core Team 2013). The tail dependence can also underpin the selection of copulas (Poulin et al. 2007).

### 3.5 Generating a pair from a copula

Simple simulation algorithms are available for most copula models, e.g. Devroye (1986, Ch. 2), or Whelan (2004) for the Archimedean copulas. In the bivariate case, a good strategy for generating a pair  $(U, V)$  from a copula  $C$  consists in using the conditional distributions:

1. Generate  $u$  from a uniform distribution on the interval  $[0, 1]$ ,
2. Given  $U = u$ , generate from the conditional distribution:

$$Q_u(v) = \mathbb{P}(V \leq v | U = u) = \frac{\partial}{\partial u} C(u, v)$$

by setting  $V = Q_u^{-1}(U^\circ)$ , where  $U^\circ \sim \mathcal{U}[0, 1]$

The explicit formulas for  $Q_u^{-1}$  are illustrated in the Table 2 for the Frank and Clayton copulas. For the Gumbel copula, no explicit formula exists, the value  $v = Q_u^{-1}(u^\circ)$  can be determined by a numerical approach<sup>4</sup>.

To avoid using an optimization algorithm, Embrechts et al. (2003) or Mc Neil (2008) propose to generate directly the pair  $(U, V)$ . In our case, the latter approach is not suitable since the storm duration must be generated for a given volume storm (See Appendix A).

<sup>4</sup> In our case, three iterations of the bisection method give the starting point of the Newton-Raphson algorithm. A  $Q_u^{-1}(u^\circ)$ -estimation as accurate as desired in this application is possible in a relatively short time.

Copulas	$C(u, v)$	$Q_u^{-1}(u^*)$
Frank	$\frac{1}{\theta} \ln \left( 1 + \frac{(e^{\theta u} - 1)(e^{\theta v} - 1)}{(e^\theta - 1)} \right)$	$-\frac{1}{\theta} \ln \left( 1 + \frac{u^*(e^{-\theta} - 1)}{\tilde{u} + (1 + u^*)e^{-\theta \cdot u}} \right)$
Clayton	$(u^{-1/\theta} + v^{-1/\theta} - 1)^{-\theta}$	$\left( 1 - u^{-\theta} + (u^* \cdot u^{1+\theta})^{-\frac{\theta}{1+\theta}} \right)^{-1/\theta}$

**Table 2** The Frank and Clayton copulas and the inverse function of their conditional distributions.

### 3.6 The discrete variable case

In the context of dependence, the methods described above depend on the continuity assumptions for the marginal distributions. In the case of discrete variables, many desirable properties of dependence measures no longer hold. The main technical argument consists in a continuous extension of integer-valued random variables. Here, we used the method proposed by (Denuit and Lambert 2005).

Assume that  $X$  is a discrete variable and  $X \geq 0$ . We associate  $X$  with a continuous random  $X^*$  such as

$$X^* = X + (U - 1), \quad \text{where } U \sim \mathcal{U}[0, 1]. \quad (4)$$

## 4 Evaluation Framework

Two performance criteria are designed to evaluate the goodness-of-fit with particular care for the extreme part of the distribution. These criteria permit to judge the reliability of a given method on a territory: the model is able to assign correct exceedance probabilities for several stations. See (Renard et al. 2013) for details on these criteria.

Let  $L_j$  be the number of years of observation for the station  $j$ .

Let  $NE_j$  be the number of observed rainy events (see Sec. 2.2 for definition) occurring on the station  $j$ .

Let  $\overline{NE_j} \stackrel{def}{=} \frac{NE_j}{L_j}$  be the average number of rainy event per year.

Let  $X_j(D) = \{X_j^i(D)\}_{i=1..NE_j}$  be the maximum rainfall in  $D$  hours for each observed rainy event.

Note that  $L_j$  can differ according to stations.

### 4.1 Quantile violations

The first criterion is based on the number of quantile violations : it verifies if the number of observation above a quantile estimated by a given method is consistent with the quantile level.

Let  $q_j^M(D, T)$  be the return level of the maximum rainfall in  $D$  hours associated with the return period  $T$  years (called *Tyr*-quantile) estimated by the model  $M$  at site  $j$ .

Let  $N_j^M(D, T)$  be the number of quantiles violations corresponding to the number of  $(X_j^i(D))_{i=1..NE_j}$  which exceed  $q_j^M(D, T)$ .

$$N_j^M(D, T) = \sum_{i=1}^{NE_j} \mathbf{1}_{X_j^i(D) > q_j^M(D, T)}$$

For a given time step  $D$  and a given model  $M$ , the exceedance of a quantile  $q_j^M(D, T)$  is a Bernoulli trial. Under the reliability assumption, its success probability is  $\frac{1}{NE_j \times T}$ . Considering the independence between rainy events, the variable  $N_j^M(D, T)$  therefore corresponds to the number of successes in  $NE_j$  independent Bernoulli experiments and thus:

$$\mathcal{QV}_T : N_j^M(T) \sim \mathcal{B}(NE_j, \frac{1}{NE_j \times T}) \quad (5)$$

where  $\mathcal{B}$  indicates the Binomial distribution with the number of trials  $NE_j$  and success probability  $p = \frac{1}{NE_j \times T}$ . Note that, in an annual maxima modelling, we have  $\overline{NE_j} = 1$ , like in (Renard et al. 2013).

#### 4.2 Frequency of the maximum observation

The second criterion is based on the maximum observed rainfall  $X_j^{\max}(D) = \max_{i=1..NE_j} X_j^i(D)$  for a given station (Garavaglia et al. 2010).

Let  $\hat{F}_j^{(D, M)}$  be the distribution function of the maximum rainfall in  $D$  hours at site  $j$  estimated by a given model  $M$ .

Under the reliability assumption, we have :

$$\mathcal{FM} : \hat{F}_j^{(D, M)}(X_j^{\max}(D)) \sim \mathcal{K}(NE_j, 1) \quad (6)$$

where  $\mathcal{K}(NE_j, 1)$  is the Kumaraswamy distribution (Kumaraswamy 1980) whose cdf can be written  $F_K(t) = t^{NE_j}$  with  $0 \leq t \leq 1$ . Indeed,

$$\begin{aligned} \mathbb{P}\left(\hat{F}_j^{(D, M)}(X_j^{\max}(D)) \leq t\right) &= \mathbb{P}\left(X_j^{\max}(D) \leq (\hat{F}_j^{(D, M)})^{-1}(t)\right) \\ &= \mathbb{P}\left(X_j^i(D) \leq (\hat{F}_j^{(D, M)})^{-1}(t); \forall i \in 1, \dots, NE_j\right) \\ \text{(Considering the independence between rainy events)} &= \prod_{i=1}^{NE_j} \mathbb{P}\left(X_j^i(D) \leq (\hat{F}_j^{(D, M)})^{-1}(t)\right) \\ \text{(under the reliability assumption)} &= \left[\hat{F}_j^{(D, M)}\left((\hat{F}_j^{(D, M)})^{-1}(t)\right)\right]^{NE_j} \\ &= t^{NE_j} \end{aligned}$$

Note that, in our case  $\hat{F}_j^{(D, M)}(\cdot)$  is empirically determined with the generation of a large number of rainy events ( $NEsim$ ).  $\hat{F}_j^{(D, M)}(X_j^{\max}(D))$  is then estimated by the Hazen formula  $\frac{i-0.5}{NEsim}$  where  $i$  is the rank of  $X_j^{\max}(D)$  among the  $NEsim$  maximum rainfall in  $D$  hours generated by the rainfall model.

### 4.3 Graphical evaluation

The reliability of a model  $M$  is tested by computing the two above criteria on  $N$  stations. They are noted  $n^M(D, T) = \{n_j^M(D, T)\}_{j=1..N}$  for the number of  $T$ yr-quantile violations and  $ff^{(D, M)} = \{ff_j^{(D, M)}\}_{j=1..N}$  for the frequency of the maximum. For each criterion, the distribution of the  $N$  realizations is then compared to the theoretical distribution under the reliability assumptions (see Eq. 5 and Eq. 6). The validity of hypotheses  $\mathcal{QV}_T$  and  $\mathcal{FM}$  is evaluated with probability-probability plots. Let  $B_{n_j, T}$  and  $K_{NE_j}$  be the distribution functions of the binomial and Kumaraswamy distributions respectively. Under the reliability of the model, the cumulative frequencies of  $K_{NE_j}(ff_j^{(D, M)})$  should be uniformly distributed -ie- the pp-plot is close to the diagonal line. Since the cumulative binomial distribution is discrete, a smoothing procedure is applied by sampling uniform variates in the interval  $[B_{n_j, T}(n_j^M(D, T) - 1), B_{n_j, T}(n_j^M(D, T)_j)]$ . To judge the reliability of a model  $M$  with the criteria  $n^M(D, T)$ , a pp-plot is constructed for each return period  $T$  and for each rainfall duration  $D$ . To resume this information, a new criterion is computed which measures how close the pp-plot is from the diagonal line :

$$S^M(\mathcal{QV}_T) = 1 - 2 \frac{1}{N} \sum_{j=1}^N |B_{n_j, T}(n_j^T) - \hat{H}_j| \quad (7)$$

$$S^M(\mathcal{FM}) = 1 - 2 \frac{1}{N} \sum_{j=1}^N |K_{NE_j}(ff_j) - \hat{H}_j| \quad (8)$$

where  $\hat{H}_j$  is the appropriate empirical frequency. By construction,  $0 \leq C^M(\cdot) \leq 1$ . if  $C^{M1}(\cdot) > C^{M2}(\cdot)$  then the model  $M1$  is assumed a better model than  $M2$  to estimate extreme quantile for a given rainfall duration and a given return period(only for the quantile violation criterion).

## 5 Application into the rainfall generator: Depth/Duration dependence

The subject of this section is to apply the copula approach into the rainfall generator in order to simulate the relationship between the depth and duration of a rainstorm. This relationship, called further the *Depth/Duration dependence* is only applied on major storms.

First, the data used in the study are briefly presented. Then, the mathematical tools, presented in Sect. 3, are applied to model the *Depth/Duration dependence*. Performance of four models are compared with criteria presented in Sect. 4.

## 5.1 Presentation of data used

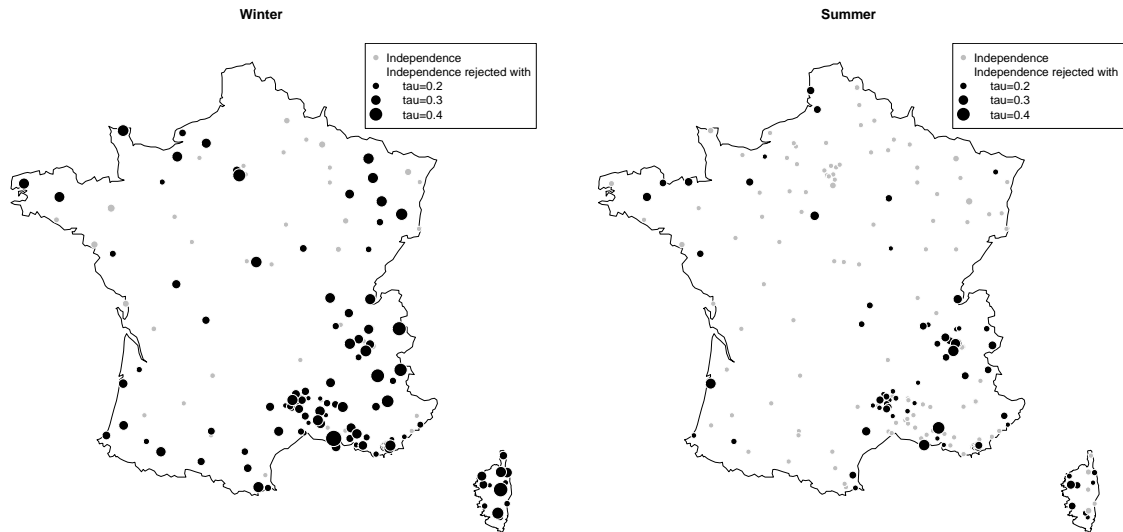
217 rain-gauge stations are used in metropolitan France (Fig. 3). Among the 217 stations studied, 173 are reference rainfall stations for the French weather office Météo-France (synoptic network). The others are stations with long observation records, and data that have been validated by management agencies – mainly Irstea, DDE, the local offices of the France Ministry of Equipment, and Diren, the regional environment authorities. If all stations are taken into account, the median observation period is 17.8 yr, with observation periods ranging from a few years for some of the alpine stations to 78 yr for the rainfall series in Marseille. The sampling of data used in this study indicates an extremely wide range of rainfall values (average annual precipitation varies between  $\sim 400\text{mm}$  and  $\sim 2000\text{mm}$ ), providing the opportunity to see how the hourly rainfall models perform in highly diverse contexts (continental, alpine and mediterranean climates).

## 5.2 The *Depth/Duration dependence* models

In the rainfall generator, the volume of a rainstorm, noted  $V$ , follows an exponential law while the duration of a rainstorm, noted  $D$ , follows a Poisson's law, a discrete law. Consequently, the method described in Sect. 3.6 is applied to transform  $D$  to  $D^*$  without losing information.

First Kendall's tau between  $V$  and  $D^*$  is estimated on each of the 217 rain gauge stations. The null hypothesis  $H_0$ : independence between  $V$  and  $D^*$  is not rejected for all stations. Actually the significance of the dependence seems to depend on the season and the geographical location (climate) (See Fig. 3). In "Winter", 140 among 217 rain gauge stations have a significant positive dependence, that is to say, a storm with a big volume is usually associated with a storm with a long duration (on these 140 stations). In "Summer", only 81 stations have a significant positive dependence, these stations are mostly located in the mountain landscape or near the ocean.

The goal is to maintain a single model structure: only model parameters can distinguish climate or rainfall pattern. Therefore only one copula should be used to model the *Depth/Duration dependence* for any station. Then, the Depth/Duration dependence modelling has been implemented into the rainfall generator by the independence, Clayton, Gumbel and Frank copulas as shown in the Sect. 3.5. For information, Joe (Joe 1993) (resp. Ali-Mikail-Haq (Ali et al. 1978)) copula which allows to model an upper (resp. no) dependence tail has been also considered but the Gumbel (resp. Frank) copula provides globally a better



**Fig. 3** Results of the independence test ( $\alpha = 0.05$ ) based on the Kendall's tau between the depth and duration of major storm tested on each 217 rain gauge stations with the False Discovery Rate approach.

fit on observed data pairs.

All considered copulas are symmetric. To assure the symmetry in the data, the test proposed by (Genest et al. 2012) has been performed<sup>5</sup> on each station where the dependence is significant in "Winter" and "Summer". The FDR procedure (See Appendix B) is applied on the p-values (140 for "Winter" and 81 for "Summer") and no null hypothesis - data come from exchangeable copulas - is rejected at a global significance level  $\alpha = 0.05$  for the both seasons.

### 5.3 Impacts of copula choice on the model performance

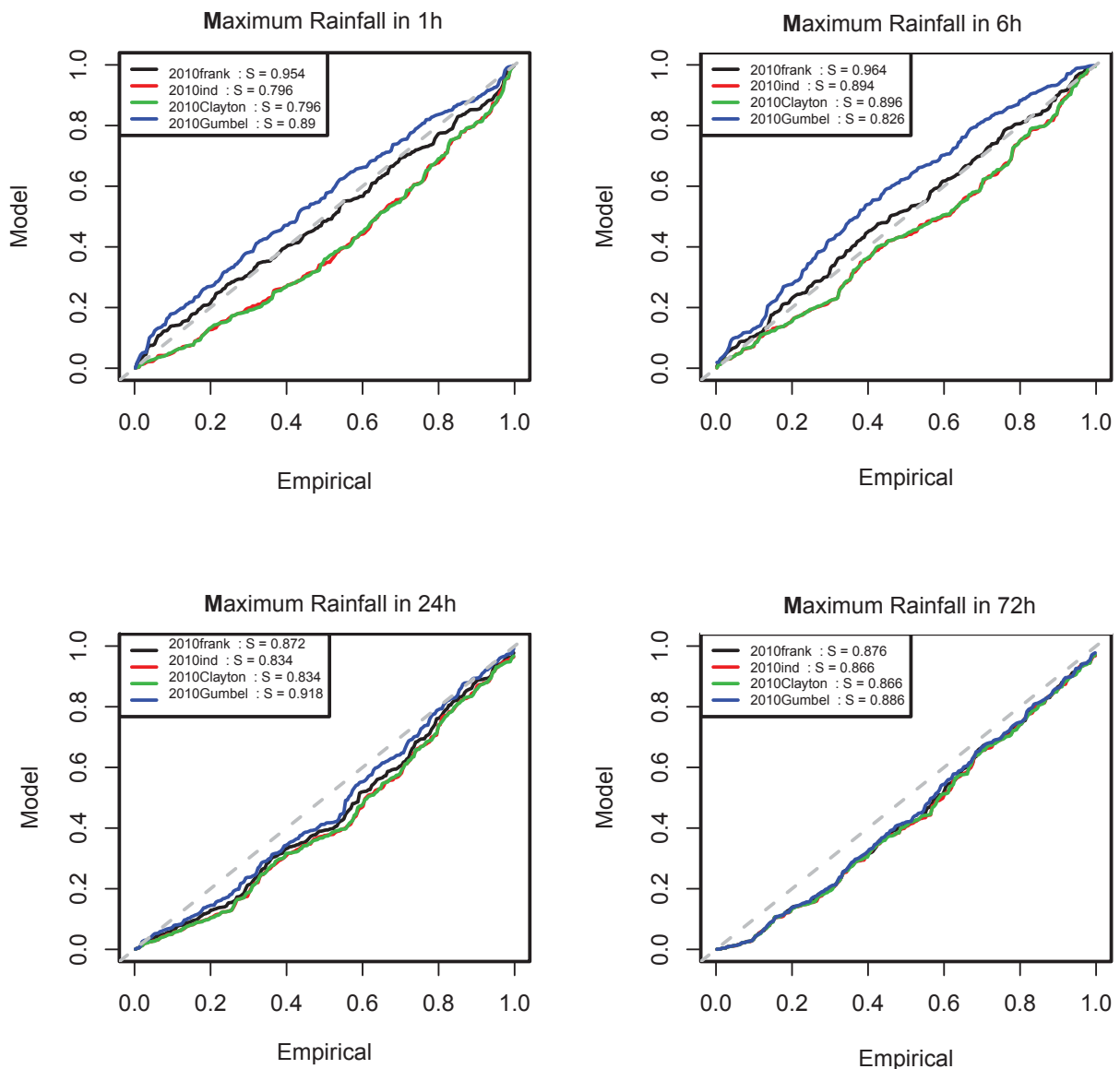
Simulations were performed on all 217 available stations and the performances of the four models were compared with the criteria presented in Sect. 4. Note that, for stations where the *Depth/Duration dependence* is not significant in both "Winter" and "Summer", rainfall quantiles are the same for the four models.

In Figure 4, we can assess how the dependence structure can affect the rainfall generator in terms of the distribution of the frequency of the maximum (criterion  $\mathcal{FM}$ ) for different rainfall durations ( $D = 1, 6, 24, 72$  hours). The performance of the Clayton copula model are very similar to the independence model's because of a lack of dependence in high frequencies. Rainfalls with long duration are less affected

<sup>5</sup> with the R function `exchTest` of the "copula" package (Kojadinovic and Yan 2012).

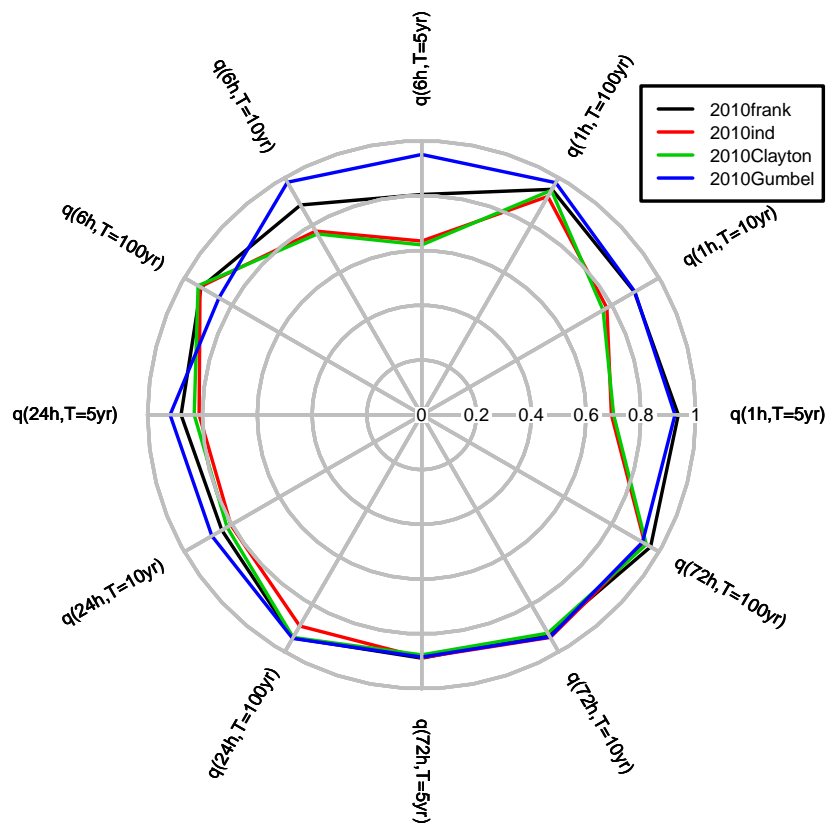


by the dependence structure and are generally over-estimated (the curves are below the diagonal for  $D \geq 24$ hours). The model with the Frank copula perform well the sub-daily rainfalls in terms of  $\mathcal{FM}$ . Indeed, the pp-plot curves are very close to the diagonal. The independence model overestimates the subdaily rainfall while for the model with the Gumbel copula, subdaily rainfalls are underestimate because of a too strong dependence in extreme values. According to the  $\mathcal{FM}$  criterion, modelling *Depth/Duration dependence* improves the performance of the rainfall generator and the Frank copula seems to be the best model to reproduce rainfall of different durations.



**Fig. 4** pp-plots of the  $\mathcal{FM}$  criterion calculated on 217 rain gauge stations. Each plot compares results of the four dependence structure for a given rainfall duration. The criterion  $S^M(\mathcal{FM})$  (Eq 8) is also illustrated in the legend.

The criterion  $QV_T$  is tested for  $T = 5, 10, 100$  years and different durations of rainfall ( $D = 1, 6, 24, 72$  hours). To limit the number of graphs, only the criterion  $S^M(QV_T)$  is illustrated by a spider plot in Figure 5. As the criterion  $\mathcal{FM}$ , the long duration rainfall is less affected by the dependence structure and the Clayton copula has no impact on the model performance. The lower quantiles ( $T \leq 10$  years) are more subject to the dependence structure than higher quantiles ( $T = 100$  years). Nevertheless, taking into account dependence with Frank or Gumbel copulas improve the model performance for any return period and any rainfall duration. According to the criterion  $QV_T$ , the Gumbel copula is generally the best dependence structure to reproduce rainfall.



**Fig. 5** Comparison of the  $S^M(QV_T)$  criterion calculated from the four dependence structures for  $T = 5, 10, 100$  years and for different durations of rainfall ( $D = 1, 6, 24, 72$  hours).

#### 5.4 Final version of the generator

The choice of the copula can be oriented by the model outputs (via criteria) but it must be validated by testing the suitability of copulas according to observation of considered variables. To assure the goodness-of-fit, the test presented by (Genest and Rémillard 2008) has been performed<sup>6</sup> on each station where the dependence is significant in "Winter" and "Summer". Like for testing the symmetry in the data (see Sec. 5.2), the FDR procedure is applied on the  $p$ -values (140 for "Winter" and 81 for "Summer"). For the Frank copula, no null hypothesis – Frank copula is well adapted – is rejected at a fixed global significance level  $\alpha = 0.05$  for the both seasons. The same test has been performed for the Gumbel copula and only 5 (resp. 2) null hypothesis are rejected for "Winter" (resp. "Summer"). For the Clayton copula, 102 (resp. 69) null hypothesis are rejected for "Winter" (resp. "Summer"). These results are in accordance to the performances of models (see Sec. 5.3): contrary to the Clayton copula, both the Gumbel and Frank copulas seem to be adapted to model the *Depth/Duration dependence* for "Winter" and "Summer" seasons. In the same way, the test of extreme-value dependence<sup>7</sup> defined in (Ben Ghorbal et al. 2009) has been performed on each station. The null hypothesis – the data come from an extreme-value copula<sup>8</sup> – is rejected on 40% (resp. 70%) stations for "Winter" (resp. "Summer") at a fixed global significance level  $\alpha = 0.05$ . Note that there is no specific geographic localization where the test is significant.

The rainfall generator distinguish rainfall occurring in "Winter" and "Summer", and so the copula can differ from season to season. Among the four possibilities, the dependence model which seems to have the best results used the Gumbel copula for "Winter" rainfall and the Frank copula for "Summer" rainfall. This configuration which is not in disagreement with the previous tests permits to have a compromise between the values of the two criteria ( $\mathcal{FM}$  and  $\mathcal{QV}_T$ ) and is considered as the actual final version of the generator.

#### 5.5 Comparison with classical extreme value theory

Extreme value theory provides the proper parametric framework to model the distribution of extremes. The approach we used is based on the Peak-Over-Threshold (POT) model. For a given duration  $D$  and a station  $j$ , the variables  $(X_j^i(D))_{i=1..NE_j}$  are distributed according to the same family and are independent

---

<sup>6</sup> with the R function `gofcopula` of the "copula" package.

<sup>7</sup> with the R function `evTestK` of the "copula" package.

<sup>8</sup> Among the three copulas, only Gumbel copula is an ev-copula

since events are considered as independent. We used the Generalized Pareto Distribution (GPD) family to model the threshold exceedances (Coles 2001). The GPD family has three parameters: a location parameter  $u$ , a scale parameter  $\sigma$  and a shape parameter  $\xi$ . The cumulative distribution function  $\mathbb{P}(X < x | X > u) = F_{(u,\sigma,\xi)}(x)$  is given by :

$$F_{(u,\sigma,\xi)}(x) = \begin{cases} 1 - (1 + \frac{\xi(x-u)}{\sigma})^{-1/\xi} & \text{with } \frac{x-u}{\sigma} < -\frac{1}{\xi} \text{ for } \xi < 0, \\ 1 - (1 + \frac{\xi(x-u)}{\sigma})^{-1/\xi} & \text{with } \frac{x-u}{\sigma} > -\frac{1}{\xi} \text{ for } \xi > 0, \\ \exp(-\frac{(x-u)}{\sigma}) & \text{for } \xi = 0. \end{cases}$$

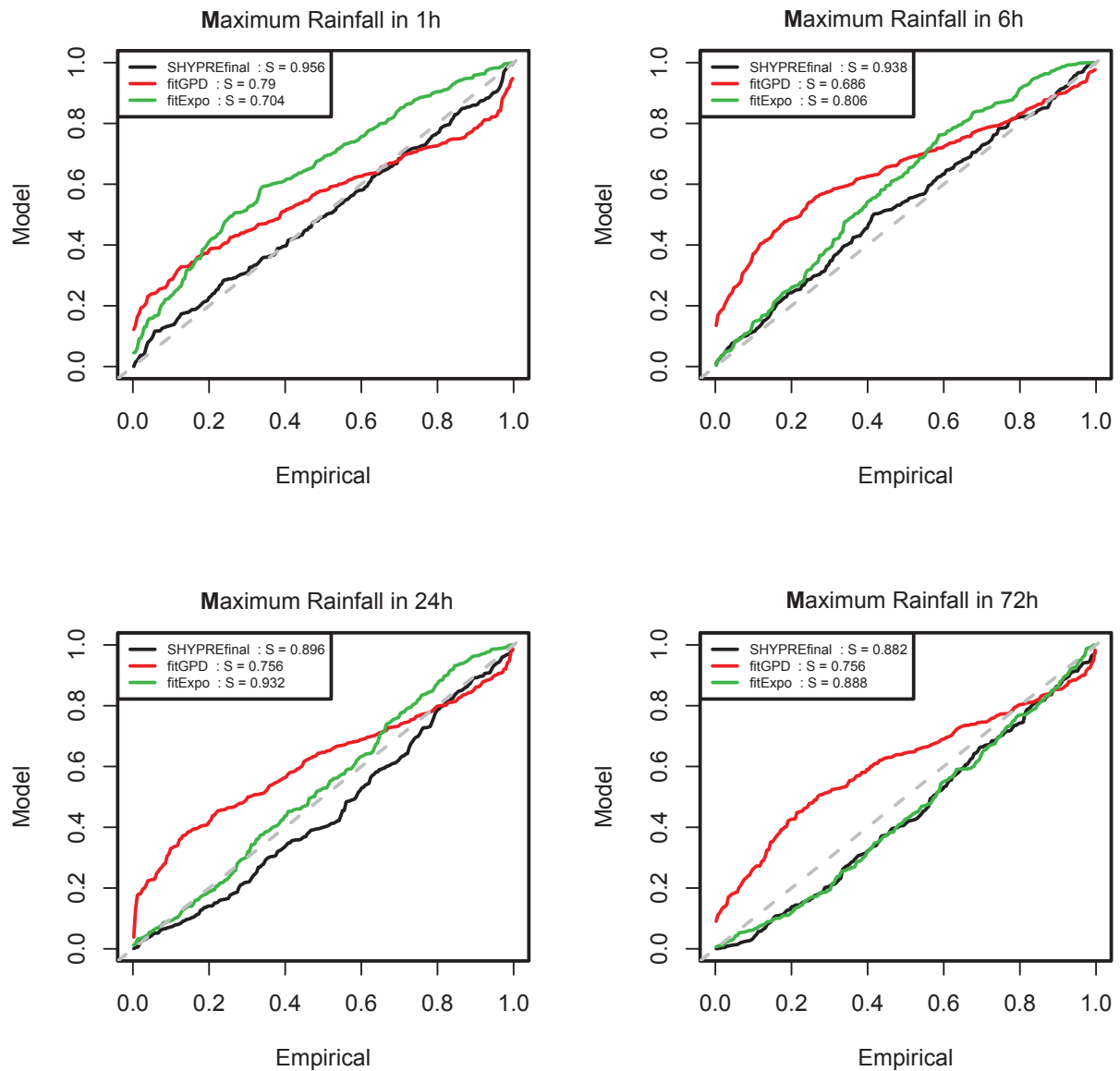
The location parameter  $u$  depends on the site and it fixes at the first quartile. The support of the shape parameter  $\xi$  is restricted to  $[-0.4, 0.4]$ . Parameters are estimated by the maximum likelihood method with the "fpot" R function from the "evd" package (Stephenson 2002). The performance of a such fitting law is tested with the criteria presented in Sec. 4 and is compared with the final version of the rainfall generator. The case where  $\xi$  is fixed to 0 (noted *fitExpo*) is also tested (see Fig.6 and 7).

Note that the GPD fitting applied here is performed separately for each rainfall duration whereas the generator estimates the quantiles of any rainfall durations simultaneously. Using information of different time steps simultaneously (for example with a scaling model) is generally more efficient and gives more accurate estimates (Bougadis and Adamowski 2006). Besides, a univariate approach is compared to a multivariate one leading to generally different results. Indeed the impact of the design choice of uni-, bi- and trivariate modelling in a rainstorm setting is for instance presented in Gräler et al. (2013).

The GPD fitting seems to be over parametrized because of the curves in "S" while an exponential fitting ( $\xi = 0$ ) under-estimates the frequency of the maximum for sub-daily rainfall (Fig. 6). For all durations and return periods, the rainfall generator is more reliable than a GPD fitting (Fig. 7). The exponential law is not adequate to estimate extreme quantiles, especially for the sub-daily rainfall.

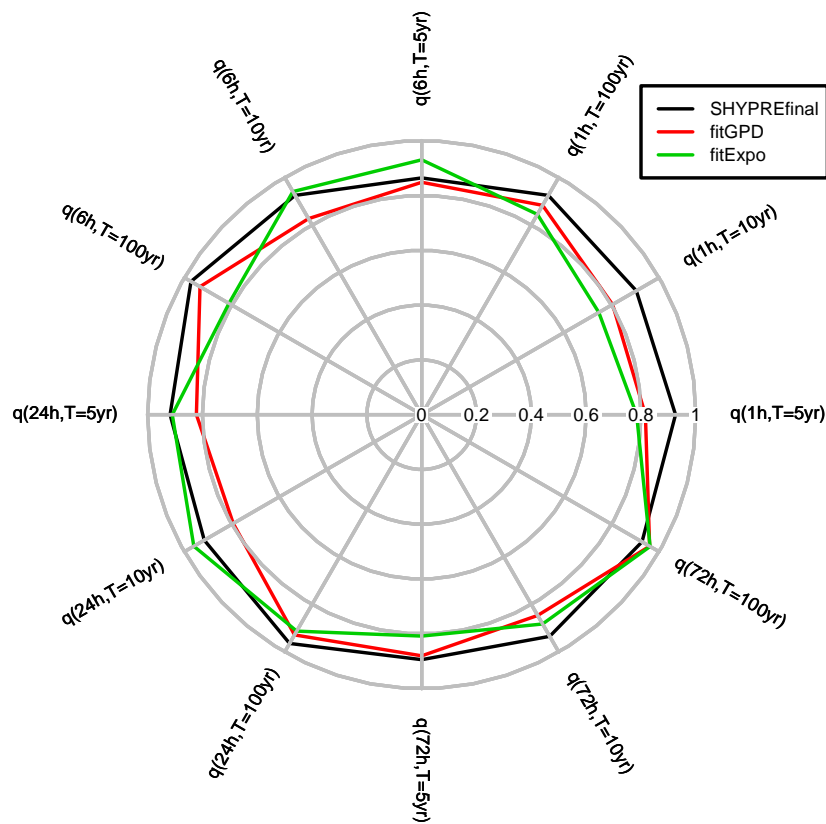
## 6 Discussion and conclusions

The copula approach was applied into a stochastic hourly rainfall generation model. With this tool, any dependence can easily be taken into account in an existing stochastic model, because a copula process permits the description of the dependencies between many random variables, independently of their marginal distributions. It has been applied to improve the generation of sub-daily rainfalls in generating a relation between the depth and the duration of a rainstorm.



**Fig. 6** pp-plots of the  $\mathcal{FM}$  criterion calculated on 217 rain gauge stations. Each plot compares results of the final model and two fitting laws for a given rainfall duration. The criterion  $S^M(\mathcal{FM})$  (Eq. 8) is also illustrated in the legend.

The rainfall generator has been developed to be applicable for all types of climate. The rainfall model structure is the same for any station, a shifting from one climate to another is possible based uniquely on the model's parameters estimated by at-site information. To continue in the same way, only one copula has been chosen to model the same structure of dependence on any sites. The copula parameter is estimated by the Kendall's Tau which is calculated from observations. Different copulas have been implemented in the generator and their performances have been tested.



**Fig. 7** Comparison of the  $S^M(QV_T)$  criterion calculated from the final model and two fitting laws for  $T = 5, 10, 100$  years and for different durations of rainfall ( $D = 1, 6, 24, 72$  hours).

The evaluation framework, applied here on several rainfall durations, focuses on the reproduction of extreme rainfall on 217 stations over France. Two criteria were used to test the fitness of the estimated quantiles through the number of quantile violations and the fitness of the probability attributed to the maximum of the observed rainfall. According to these criteria, the Gumbel and Frank copulas significantly enable the improvement of the model performance especially for the sub-daily rainfall while the Clayton copula shows similar results to the independence assumption. The goodness-of-fit of these two copulas has been tested and none of them can be rejected at a global significance level. Finally, the Gumbel (resp. Frank) copula has been chosen to model the dependence between the depth and the duration of rainfall occurring in "Winter" (resp. "Summer"). This version of the rainfall generator shows a better performance than the usual univariate extreme value distribution and proposes a good estimation of extreme quantile for rainfall of 1h to 72 hours.

Here, two seasons are distinguished into the generator calibration. A future version of the model will take into account different weather patterns based on meteorological circulation used in Garavaglia et al. (2010). This distinction could lead to use different copulas for each class of weather providing a better generation of all rainfall patterns.

### A Algorithm to generate hourly rainfalls inside a rainy event

The descriptive variables used below are described in Sec. 2.2 and in Fig.2 and their probability laws are defined in Arnaud et al. (2007). To facilitate notations, "major" and "ordinary" are not distinguished. The number of rainy events per year is generated from Poisson's law. For each rainy event :

1. The number of rainy periods  $N_{RP}$  is generated according to a geometric law;
2. The duration(in hours) between two rainy periods  $D_{RP}$  are generated  $N_{RP} - 1$  times according to a geometric law;
3. For the rainy period  $j = 1..N_{RP}$ , a number of storms  $N_S^j$  is generated from a geometric law;  
Let  $N_S = \sum_{j=1}^{N_{RP}} N_S^j$ , the total number of storms in the rainy event;
4. Let  $F_V = \{F_V^1, \dots, F_V^{N_S}\} \in [0, 1]^{N_S}$  be the non exceedance frequencies of the  $N_S$  storm depths. The  $N_S$  realizations  $f_V^i$  are generated sequentially to model the persistence phenomenon (dependence between the depths of successive storms) described in Arnaud et al. (2007) and Cantet (2009);
5. Let  $F_D = \{F_D^1, \dots, F_D^{N_S}\} \in [0, 1]^{N_S}$  be the non exceedance frequencies of the  $N_S$  storm durations. The  $N_S$  realizations  $f_D^i$  are generated separately according to the chosen copula in using the conditional distribution  $\mathbb{P}(F_D^i \leq f_D^i | F_V^i = f_V^i)$  as described in Sec. 3.5;
6. With the inverse probability integral transform, the non-exceedance frequencies are easily converted into depths (exponential law) and durations (Poisson's law).

As this stage, each storm of a rainy period is defined by a depth and a duration. The hourly rainfall are generated as follows;

For a  $D_S$ -hours storm associated to a  $V_S$  depth:

7.  $RX_S$  is generated from a truncated log-normal law ( $1/D_S \leq RX_S \leq 1$ ) and  $RPX_S$  from a truncated normal law ( $1 \leq RPX_S \leq D_S$ ). Then the hourly peak of the storm is characterized by its depth  $V_S \cdot RX_S$  and its position  $RPX_S$  inside the storm. Then the volume  $V_S \cdot (1 - RX_S)$  are uniformly distributed in the  $D_S - 1$  other hourly rainfalls. These  $D_S - 1$  hourly rainfalls are positioned to respect the storm definition (an unique local minimum).

A rejection sampling is performed to satisfy the definition of rainy events (at least a daily rainfall superior to 20mm).

## B Controlling the global significance level of a multiple tests approach using the False

### Discovery Rate (FDR): the Benjamini and Hochberg (BH) procedure

Benjamini and Hochberg (1995) proposed a procedure to control the global significance level  $\alpha_g$  of a multiple tests procedure. Assuming that  $K$  tests of a null hypothesis  $H_0$  are achieved, the BH procedure is the following:

1. Let  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(K)}$  be the sorted observed p-values related to the  $K$  tests;
2. Compute  $m = \max \{1 \leq j \leq K, p_{(j)} \leq \frac{j}{K} \alpha\}$ ;
3. If  $m$  exists, then reject among the  $K$  hypothesis the  $m$  ones corresponding to  $p_{(1)} \leq \dots \leq p_{(m)}$  p-values; else reject no hypothesis.

## References

- Adams, B. J. (2000). Urban stormwater management planning with analytical probabilistic models.
- Ali, M., Mikhail, N., and Haq, M. (1978). A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, 8(3):405–412.
- Arnaud, P. (1997). *Modèle de prédétermination de crues basé sur la simulation stochastique des pluies horaires*. PhD thesis, Université Montpellier II.
- Arnaud, P., Fine, J., and Lavabre, J. (2007). An hourly rainfall generation model adapted to all types of climate. *Atmospheric Research*, 85, Issue 2:230–242.
- Arnaud, P. and Lavabre, J. (1999). Using a stochastic model for generating hourly hyetographs to study extreme rainfalls. *Hydrological Sciences Journal*, 44(3):433–446.
- Arnaud, P. and Lavabre, J. (2002). Coupled rainfall model and discharge model for flood frequency estimation. *Water Resources Research*, 38:1075–1085.
- Arnaud, P. and Lavabre, J. (2010). *Estimation de l'aléa pluvial en France métropolitaine*. Editions QUAE.
- Arnaud, P., Lavabre, J., Sol, B., and Desouches, C. (2006). Cartographie de l'aléa pluviographique de la France. *La Houille Blanche*, 5:102–111.
- Bacchi, B., Becciu, G., and Kottegod, N. (1994). Bivariate exponential model applied to intensities and durations of extreme rainfall. *Journal of Hydrology*, 155(1-2):225–236.
- Balistracchi, M. and Bacchi, B. (2011). Modelling the statistical dependence of rainfall event variables through copula functions. *Hydrology and Earth System Sciences*, 15:1959–1977.
- Bárdossy, A., Pegram, G., et al. (2009). Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences*, 13(12):2299.
- Ben Ghorbal, N., Genest, C., and Nešlehová, J. (2009). On the ghoudi, khoudraji, and rivest test for extreme-value dependence. *Canadian Journal of Statistics*, 37(4):534–552.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57:289–300.



- Bougadis, J. and Adamowski, K. (2006). Scaling model of a rainfall intensity-duration-frequency relationship. *Hydrological Processes*, 20(17):3747–3757.
- Cantet, P. (2009). *Impacts du changement climatique sur les pluies extrêmes par l'utilisation d'un générateur stochastique de pluies*. PhD thesis, Université de Montpellier II.
- Cantet, P., Bacro, J., and Arnaud, P. (2011). Using a rainfall stochastic generator to detect trends in extreme rainfall. *Stochastic Environmental Research and Risk Assessment*, Vol. 25; Issue 3:429–441.
- Carreau, J., Neppel, L. Neppel, L. P., and Cantet, P. (2013). Extreme rainfall frequency analysis in the south of france : Comparisons of three regional methods. *Journal de la Société Française de Statistique*, 154(2):119–138.
- Carvajal, C., Peyras, L. ans Arnaud, P., Boissier, D., and Royet, P. (2009). Probabilistic modeling of floodwater level for dam reservoirs. *Journal of Hydrologic Engineering*, 14(3):223–232.
- Cernesson, F. (1993). *Modèle simple de prédetermination des crues de fréquences courantes à rares sur petits bassins versants méditerranéens*. PhD thesis, Université Montpellier II.
- Cernesson, F., Lavabre, J., and Masson, J. (1996). Stochastic model for generating hourly hyetograph. *Atmospheric Research*, 42, n°1-4:149–161.
- Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.
- Coles, S. (2001). *An introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, Heidelberg, Germany.
- Córdova, J. R. and Rodríguez-Iturbe, I. (1985). On the probabilistic structure of storm surface runoff. *Water Resour. Res*, 21(5):755–763.
- De Michele, C. and Salvadori, G. (2003). A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. *Journal of geophysical research*, 108(D 2):15–15.
- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R. (2005). Bivariate statistical approach to spillway design flood. *J. Hydrol. Engng ASCE*, 10(1):50–57.
- Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1):40–57.
- Devroye, L. (1986). *Non-uniform random variate generation*. Springer-Verlag, New York.
- Díaz-Granados, M. A., Valdes, J. B., and Bras, R. L. (1984). A physically based flood frequency distribution. *Water Resour. Res*, 20(7):995–1002.
- Eagleson, P. (1972). Dynamics of flood frequency. *Water Resour. Res*, 8 (4):878–898.
- Eagleson, P. (1978a). Climate, soil, and vegetation. 2. the distribution of annual precipitation derived from observed storm sequences. *Water Resour. Res*, 14(5):713–721.
- Eagleson, P. (1978b). Climate, soil, and vegetation. 5. a derived distribution of storm surface runoff. *Water Resour. Res*, 14(5):741–748.
- Eagleson, P. (1978c). Climate, soil, and vegetation. 7. a derived distribution of annual water yield. *Water Resour. Res*, 14(5):765–776.

- Embrechts, P., Lindskog, F., and Mcneil, A. (2003). Modelling dependence with copulas and applications to risk management. *Handbook of Heavy Tail Distributions in Finance*, pages 324–384.
- Favre, A., El Adlouni, S., Perreault, L., Thiémonge, N., and Bobee, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1).
- Fine, J. and Lavabre, J. (2002). Synthèse des débits de crue sur l'île de la réunion. phase i : la pluviométrie. éléments de régionalisation du générateur de pluie. Technical report, Cemagref.
- Fouchier, C. (2007). Aiga: an operational tool for flood warning in southern france. principle and performances on mediterranean flash floods. In *Geophysical Research Abstracts of EGU, Vienne, 15-20 avril 2007*, volume 9, page 02843.
- Frank, M. (1979). On the simultaneous associativity of  $F(x, y)$  and  $x+y - F(x, y)$ . *Aequationes Mathematicae*, 19(1):194–226.
- Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garçon, R., and Bernardara, P. (2010). Introducing a rainfall compound distribution model based on weather patterns sub-sampling. *Hydrology and Earth System Sciences*, 14:951–964.
- Gargouri-Ellouze, E. and Chebchoub, A. (2008). Modelisation de la structure de dependance hauteur-duree d'evenements pluvieux par la copule de Gumbel/Modelling the dependence structure of rainfall depth and duration by Gumbel's copula. *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, 53(4):802–817.
- Genest, C. and Favre, A. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12:347.
- Genest, C., Ghoudi, K., and Rivest, L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.
- Genest, C., Masiello, E., and Tribouley, K. (2008a). Estimating copula densities through wavelets. *Insurance Mathematics and Economics*.
- Genest, C., Nešlehová, J., and Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, 64(4):811–834.
- Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, 44(6):1096–1127.
- Genest, C., Rémillard, B., and Beaudoin, D. (2008b). Goodness-of-fit tests for copulas: A review and a power study. *Insurance Mathematics and Economics*, 44:199–213.
- Ghosh, S. (2010). Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula. *Hydrological Processes*, 24(24):3558–3567.
- Goel, N. K., Kurothe, R. S., Mathur, B. S., and Vogel, R. M. (2000). A derived flood frequency distribution for correlated rainfall intensity and duration. *Journal of Hydrology*, 228(1-2):56–67.
- Gräler, B., van den Berg, M., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph

- estimation. *Hydrology and Earth System Sciences*, 17(4):1281–1296.
- Gumbel, E. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association*, 56(294):335–349.
- Guo, Y. and Adams, B. J. (1999). An analytical probabilistic approach to sizing flood control detention facilities. *Water Resour. Res.*, 35(8):2457–2468.
- Haberlandt, U., Ebner von Eschenbach, A., and Buchwald, I. (2008). A space-time hybrid hourly rainfall model for derived flood frequency analysis. *Hydrol. Earth Syst. Sci.*, 12:1353–1367.
- Hoeffding, W. (1940). Masstabinvariante korrelationstheorie. *Schrijftfen Math. Inst. Inst. Angew. Math. Univ. Berlin*, 5:179–233.
- Ivan Kojadinovic and Jun Yan (2010). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34(9):1–20.
- Javelle, P., Fouchier, C., Arnaud, P., and Lavabre, J. (2010). Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations. *Journal of Hydrology*, 394(2):267–274.
- Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of multivariate analysis*, 46:262–282.
- Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall/CRC.
- Kojadinovic, I. and Yan, J. (2012). A non-parametric test of exchangeability for extreme-value and left-tail decreasing bivariate copulas. *Scandinavian Journal of Statistics*, 39(3):480–496.
- Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1):79–88.
- Kurothe, R. S., Goel, N. K., and Mathur, B. S. (1997). Derived flood frequency distribution for negatively correlated rainfall intensity and duration. *Water Resour. Res.*, 33(9):2103–2107.
- Li, J. Y. and Adams, B. J. (2000). Probabilistic models for analysis of urban runoff control systems. *Journal of Environmental Engineering*, 126(3).
- Mc Neil, A. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581.
- Muller, A., Arnaud, P., Lang, M., and Lavabre, J. (2009). Uncertainties of extreme rainfall quantiles estimated by a stochastic rainfall model and by a generalized pareto distribution. *Hydrological Sciences*, 54 : 3:417–429.
- Nelsen, R. (2006). *An introduction to copulas*. Springer Verlag.
- Neppel, L., Arnaud, P., and Lavabre, J. (2007). Connaissance régionale des pluies extrêmes. comparaison de deux approches appliquées en milieu méditerranéens. *C. R. Geoscience*, 339:820–830.
- Poulin, A., Huard, D., Favre, A., and Pugin, S. (2007). Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, 12:394–403.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Foundation for Statistical Computing, Vienna, Austria.

- 
- Renard, B., Kochanek, K., Lang, M., Garavaglia, F., Paquet, E., Neppel, L., Carreau, J., Arnaud, P., Aubert, Y., Borchi, F., Soubeyroux, J.-M., J. S., Veysseire, J.-M., Sauquet, E., Cipriani, T., and Auffray, A. (2013). Data-based comparison of frequency analysis methods: a general framework. *Water Resources Research*, Accepted for publication.
- Salvadori, G. and De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(12):W12511.
- Salvadori, G. and De Michele, C. (2006). Statistical characterization of temporal structure of storms. *Advances in Water Resources*, 29(6):827–842.
- Salvadori, G., De Michele, C., and Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15(11):3293.
- Salvadori, G., de Michele, C., Kottegoda, N., and Rosso, R. (2007). *Extremes in Nature: An Approach Using Copulas*. Springer.
- Singh, K. and Singh, V. P. (1991). Derivation of bivariate probability density functions with exponential marginals. *Stoch. Hydrol. Hydraul.*, 5(1):55–68.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. *Statistics Univ. Paris*, 8:229–231.
- Stephenson, A. G. (2002). evd: Extreme value distributions. *R News*, 2(2):0.
- Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics/La revue canadienne de statistique*, 33(3):357–375.
- Vandenberghe, S., Verhoest, N., and De Baets, B. (2010). Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall. *Water Resour. Res.*, 46:W01512.
- Vandenberghe, S., Verhoest, N., Onof, C., and 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 Resources Research*, 47(7):W07529.
- Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative finance*, 4(3):339–352.
- Wu, S., Tung, Y., and Yang, J. (2006). Stochastic generation of hourly rainstorm events. *Stochastic Environmental Research and Risk Assessment*, 21(2):195–212.
- Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):1–21.
- Zhang, L. and Singh, V. (2007). Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology*, 332(1-2):93–109.