Hydrol. Earth Syst. Sci. Discuss., 9, 11227–11266, 2012
www.hydrol-earth-syst-sci-discuss.net/9/11227/2012/
doi:10.5194/hessd-9-11227-2012

Hydrology and Earth System Sciences Discussions

Gains from modelling dependence of rainfall variables into a stochastic model: application of the copula approach at several sites

P. Cantet and P. Arnaud

IRSTEA, 3275 Route de Cézanne, CS 40061, 13182 Aix en Provence, France

Received: 30 August 2012 – Accepted: 4 September 2012 – Published: 2 October 2012

Correspondence to: P. Arnaud (patrick.arnaud@irstea.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Since the last decade, copulas have become more and more widespread in the construction of hydrological models. Unlike the multivariate statistics which are traditionally used, this tool enables scientists to model different dependence structures without drawbacks. The authors propose to apply copulas to improve the performance of an existing model. The hourly rainfall stochastic model SHYPRE is based on the simulation of descriptive variables. It generates long series of hourly rainfall and enables the estimation of distribution quantiles for different climates. The paper focuses on the relationship between two variables describing the rainfall signal. First, Kendall's tau is estimated on each of the 217 rain gauge stations in France, then the False Discovery Rate procedure is used to define stations for which the dependence is significant. Among three usual archimedean copulas, a unique 2-copula is chosen to model this dependence for any station. Modelling dependence leads to an obvious improvement in the reproduction of the standard and extreme statistics of maximum rainfall, especially for the sub-daily rainfall. An accuracy test for the extreme values shows the good asymptotic behaviour of the new rainfall generator version and the impacts of the copula choice on extreme quantile estimation.

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 storm duration, by assuming the two random variables independent and exponentially distributed. This paper stimulated much 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).

HESSD

9, 11227–11266, 2012

Modelling dependence of rainfall variables into a stochastic model

P. Cantet and P. Arnaud

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 and Papa (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 make it possible to model 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 then 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; Haberlandt et al., 2011) 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

HESSD
9, 11227–11266, 2012

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 aim of this paper is to present a practical framework which stochastically generates a dependence between the different rainstorm characteristics into a rainfall model already presented in (Cernesson et al., 1996; Arnaud and Lavabre, 1999, 2002; Arnaud et al., 2007). Like in Wu et al. (2006), the proposed model 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 based uniquely 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 (≥ 24 h) (Muller et al., 2009; Neppel et al., 2007).

However, the sub-daily rainfalls generated by the model do not properly respect observations on several sites, particularly for sites situated in the mountain landscape and near the Atlantic Ocean. In these regions, the coefficients of the Montana's laws[1] estimated from the simulated rainfalls are really different from the reality. It can be explained by a non-modelling dependence. To improve the generation of the sub-daily rainfall, the paper focuses on the application of the copula theory to generate correlated rainfall characteristics, especially the depth and duration of a rainstorm.

[1]In France, Montana's laws are widely used in applied hydrology providing a relationship between rainfall of different time steps. Different rainfall patterns occurring in France can be distinguished by the Montana coefficient.

HESSD
9, 11227–11266, 2012

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.

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. The selection threshold of 20 mm leads to the determination of a first parameter, the average number of events per year (NE), strongly variable 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.

HESSD
9, 11227–11266, 2012

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 (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. The objective of the model regionalization (realized on the whole French territory) led us to define the same theoretical law for a given 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 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

HESSD
9, 11227–11266, 2012

the precipitation regimes. Note that some of the 20 parameters either vary only 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. 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. To reduce the sampling effect on the simulated events, we chose to generate rainfall on periods which were a thousand times longer than the strongest return period which we want to determine. For example, a 100 yr-quantile is determined by generated hyetographs on a 100 000 yr simulation period. Quantiles can then be empirically estimated from these simulated times series without uncertainty due to the sampling variability.

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

HESSD
9, 11227–11266, 2012

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 km^2 everywhere in the French territory (Arnaud et al., 2006).

3 How to diagnose and model the dependence

The aim of this part is to introduce the mathematical tools used in the study of dependence between random variables. Only tools used in our study are clearly presented. For further information, see Nelsen (2006) and Genest and Favre (2007).

3.1 Measuring dependence: Kendall's tau

Classically, dependence is measured by correlation coefficients. The most well-known is Pearson's coefficient (R) used for example in a linear regression. It only characterizes a linear dependence between two variables. When the dependence is not linear, a correlation computed on ranks appears to be the best approach (Oakes, 1982) leading to the building of two other correlation coefficients: Spearman's rho and Kendall's tau. Only Kendall's tau (noted τ) is presented in this paper:

Suppose that a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is given from some pair (X, Y) of continuous variables. Here, R_i stands for the rank of X_i among $X_1, \ldots, X_n$, and S_i stands for the rank of Y_i among $Y_1, \ldots, Y_n$. The empirical version of Kendall's tau is given by:

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

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 $(X_i - X_j)(Y_i - Y_j) < 0$. The borderline case $(X_i - X_j)(Y_i - Y_j) = 0$ occurs with a probability zero under assumption that X and Y are continuous. The factor $n(n-1)/2$ corresponds to the number of pairs which are compared.

HESSD

9, 11227–11266, 2012

It is obvious that τ_n is a function of the ranks of the observations only, since $(X_i - X_j)(Y_i - Y_j) > 0$ if and only if $(R_i - R_j)(S_i - S_j) > 0$.

If X and Y are mutually independent, we have $\tau_n \approx 0$. The closer to $1\,|\tau_n|$ is, the stronger the dependence between two variables. If $\tau_n > 0$ (resp. < 0), the dependence is positive (resp. negative).

An independence test can be based on τ_n, since under H_0: independence between two variables,big this statistic is close to normal with zero mean and variance $2(2n+5)/(9n(n+1))$. For example, we can reject H_0 with a significance level (type I error) $\alpha = 5\,\%$ if $\sqrt{\frac{9n(n+1)}{2(2n+5)}}|\tau_n| > z_{\alpha/2} = 1.96$.

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

Traditionally, the pairwise dependence between variables has been described using classical families of multivariate distributions. The main limitation of this approach is that the individual behavior of the two variables must be characterized by the same parametric family of univariate distributions. The copula model, introduced by (Hoeffding, 1940; Sklar, 1959), is more and more widespread since it avoids this restriction.

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

A bidimensional copula, also called a 2-copula, is a two-place real function defined on $[0,1]\times[0,1] \longrightarrow [0,1]$ such as

1. $\forall u, v \in [0,1]$,
   $C(u,0) = 0$, $C(u,1) = u$, $C(0,v) = 0$, $C(1,v) = v$;

2. $\forall u_1, u_2, v_1, v_2 \in [0,1]$ such as $u_1 \leq u_2$ and $v_1 \leq v_2$,
   $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$

HESSD
9, 11227–11266, 2012

Modelling dependence of rainfall variables into a stochastic model

P. Cantet and P. Arnaud

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

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

where F_X and F_Y are the marginal functions and $C : [0,1] \times [0,1] \to [0,1]$ 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.

For a random sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ from some pair (X,Y), an *empirical copula* can be introduced, and is defined by

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(F_X(X_i)\leq u \cap F_Y(Y_i)\leq v)} \quad (3)$$

where $\mathbf{1}_{(.)}$ denotes the indicator function, F_X and F_Y are the marginal distributions of X and Y.

3.3 Estimation and choice of models

Modeling dependence between two random variables (X and Y) can be achieved by using some families of copulas. In this paper, we only considered 3 archimedean copulas: the Frank copula (Frank, 1979), the Clayton copula (Clayton, 1978), and the Gumbel copula (Gumbel, 1961). These copulas have been chosen because they have only one parameter and are easily applicable.

Like usual statistic laws, different methods are used to estimate copula parameters. Spearman's rho and Kendall's tau can be used as estimators since some analytic relations between these two quantities and the copula parameters exist (see Table 1 for Kendall's tau). A method based on the maximizing of the likelihood is also often used

HESSD
9, 11227–11266, 2012

(Genest et al., 1995). For other methods, see Joe (1997), Tsukahara (2005) and Genest et al. (2008a). In this study, we estimated the copula parameter with the Kendall's tau.

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. For the unidimensional law, several tests provide the best fitting to the observations, for example the Kolmogorov-Smirnov test. To test the suitability of copula models, the same principle can be used. For example, we can compare the empirical copula (defined in Eq. (3)) to a theoretical copula through the calculation of the Kolmogorov-Smirnov statistic or through a QQ-plot. In this way, Genest and Rivest (1993); Hillali (2001) proposed a test for the Archimedean copulas. Genest et al. (2008b) compared a lot of measures to choose the best copula. Genest and Rémillard (2008) use a bootstrap procedure for suitability testing. This test has been implemented in the "copula" package (Yan, 2007) from the language R (http://www.r-project.org/).

3.4 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:
   $Q_u(v) = \mathbb{P}\left(V \leq v | U=u\right) = \frac{\partial}{\partial u} C(u,v)$
   by setting $V = Q_u^{-1}(U^\circ)$, where $U^\circ \sim \mathcal{U}[0,1]$

HESSD
9, 11227–11266, 2012

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[2].

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 (already generated Arnaud et al., 2007).

3.5 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

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

4 Application into the rainfall generator: Depth/Duration dependence

The subject of this section is to apply the copula approach to the rainfall generator to simulate the relationship between the depth and duration of a rainstorm. This relationship is called further the *Depth/Duration dependence*. Only major storms are taken into account to study this dependence.

[2] 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 is possible in a relatively short time.

HESSD
9, 11227–11266, 2012

First, the data used in the study are briefly presented. Then, the mathematical tools, presented in Sect. 3, are applied to model this dependence. Finally, the impacts on the rainfall quantiles estimation are illustrated.

4.1 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 Cemagref; 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, providing the opportunity to see how the hourly rainfall models perform in highly diverse contexts. Arnaud et al. (2007) used the same stations and presented them in further details.

4.2 The *Depth/Duration dependence* model

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. 3.5 is applied to transform D to D^* without losing information.

4.2.1 Where is the *Depth/Duration dependence* significant?

First Kendall's tau between V and D^* is estimated on each of the 217 rain gauge stations. Then a False Discovery Rate (FDR) approach (see Appendix A) is used to

HESSD
9, 11227–11266, 2012

determine, from the 217 obtained p-values[3], the number of rejected null hypothesis, that is the number of stations for which the *Depth/Duration* independence hypothesis is rejected at a fixed significance level $\alpha = 0.05$ (type I error).

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. These results are climatologically consistent. Indeed, summer rainfall, especially in the continental climate, occur in rainy phenomena providing rainstorms with a strong intensity and a short duration (convective systems).

4.2.2 How can the *Depth/Duration dependence* be modelled?

The goal is to maintain a single model structure: only model parameters can distinguish the climate. Therefore only one copula should be used to model the *Depth/Duration dependence* for any station.

On each station where the dependence is significant (140 for Winter and 81 for Summer), the L^2-distance between the empirical copula and the 3 theoretical copulas is calculated and is ordered. The best copula, that is to say the copula whose distance is minimum, has the rank 1. Most of the time, the copula which is selected is the Frank copula (see Table 3). When another copula seems to be better (it often occurs when the numbers of storms is lower than 40), the Frank copula is always the second best copula, never the "worst" copula. Besides, the L^2-distance for the Frank copula

[3]the p-value is the probability of obtaining a result at least as large as the one actually observed, given that the null hypothesis is true. In our case, it corresponds to $\mathbb{P}(X > \text{tau})$ where $X \sim \mathcal{N}(0, 2(2n+5)/(9n(n+1)))$ with n the number of storms in the studied station.

is close to the L^2-distance for the best copula. Finally, there is no specific geographic localization where the Frank copula is not the best one.

This procedure is not a formal test, it only permits to define the best adapted copula among others according to a criterion. To assure the goodness-of-fit, the test presented by (Genest and Rémillard, 2008)[4] has been performed on each station where the dependence is significant. As for the independence test, the FDR procedure is applied on the p-values (140 for Winter and 81 for Summer). No null hypothesis – Frank copula is well adapted – is rejected at a fixed significance level $\alpha = 0.05$ for Winter and Summer. Note that, 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).

The Frank copula is chosen to model the Depth/Duration dependence for any station. The parameter θ of the copula is estimated with the inversion of Kendall's tau which is estimated on each station (as shown in Fig. 3). Therefore, the shifting from one climate to another is possible based uniquely on the parameter θ.

4.3 Impacts on the rainfall quantiles estimated by the generator

The Depth/Duration dependence modelling (by Frank and Gumbel copulas) has been implemented into the rainfall generator as shown in the Sect. 3.4. Simulations were performed on all 217 available stations and the performance of the new model was compared to the performance of the model that does not take into account the Depth/Duration dependence for all stations. The model is only tested in terms of reproduction of the maximum rainfall of an event. Testing autocorrelation, cross-validation or intermittency is not the subject of the paper.

First, the impact of the Depth/Duration dependence modelling is illustrated by the plotting of the frequency distributions for 1-h maximum rainfall for three stations (see Fig. 4). On these three stations, presented in Table 4, the effects are not high even if

[4] with the R function gofcopula of the "copula" package.

HESSD
9, 11227–11266, 2012

the positive dependence is significant. Nevertheless the new model allows a better estimation according to the observations. Indeed, without the dependence modelling, the generator seems to overestimate 1-h rainfall. Note that, the quantiles with the Gumbel copula model are not shown in Fig. 4 because they are very close to Frank's quantiles.

Then, we also compared the quantiles obtained from fitting an exponential law[5] to the observation samples (noted Q^T_{obs}) with quantiles from simulations of rainfall events (noted Q^T_{RG}) according to two criteria:

1. The relative error given by

$$\text{Error} = 100\,\frac{Q^T_{\text{RG}} - Q^T_{\text{obs}}}{Q^T_{\text{obs}}} \tag{5}$$

   is calculated on each 217 available stations. Its distribution is illustrated by a boxplot where whiskers corresponding to the 0.05 and 0.95 quantile (See Fig. 5).

2. The Nash criterion (Nash and Sutcliffe, 1970) given by

$$\text{Nash} = 1 - \frac{\sum_{i=1}^{n} (Q^T_{\text{obs}} - Q^T_{\text{RG}})^2}{\sum_{i=1}^{n} (Q^T_{\text{obs}} - \overline{Q^T_{\text{obs}}})^2} \tag{6}$$

   where $\overline{Q^T_{\text{obs}}} = \frac{1}{n}\sum_{i=1}^{n} Q^T_{\text{obs}}$ and n is the number of studied stations. It is widely considered that Nash ≥ 0.7 signifies that the two series are similar. Table 5 illustrated

[5] The exponential has been chosen because the estimation of its parameter is few influenced by the sampling in comparison to a Generalized Pareto Distribution. Besides, only quantiles with a return period $T \leq 10$ yr, for which the choice of the distribution leads to a little gap, are compared.

HESSD

9, 11227–11266, 2012

the value of the Nash criterion for $T = 2, 5, 10$ yr calculated on the 217 available stations coming from quantiles estimated by the two models.

Rainfall patterns can be distinguished according to the ratio between the short duration rainfall and long duration rainfall. Figure 6 illustrated the difference (like in Eq. (5)) between R^T_{obs} and R^T_{RG} where:

$$R^T_{\text{obs}}(D1, D2) = \frac{Q^T_{\text{obs}}(D1)}{Q^T_{\text{obs}}(D2)} \quad \text{and} \quad R^T_{\text{RG}}(D1, D2) = \frac{Q^T_{\text{RG}}(D1)}{Q^T_{\text{RG}}(D2)} \tag{7}$$

$D1$ or $D2$ being the duration of the maximum rainfall with $D1 = 1$ h or 6 h and $D2 = 6$ h or 24 h.

Results presented in Table 5 and Figs. 5 and 6 show an obvious gain of the copula used to reproduce hourly extreme rainfall. Indeed quantiles are globally more similar to observed data when dependence is modelled for both models (Gumbel or Frank). This improvement is due to a better grasp of the observed phenomena. Modelling the Depth/Duration dependence results in a more accurate plotting of rainfall quantiles, especially for the sub-daily maximum rainfall, enables us to generate different rainfall patterns. The copula choice in the Depth/Duration dependence modelling leads to little impact on the estimation of rainfall T-quantiles with $T \leq 10$ yr for any duration.

The previous part showed that quantiles estimated by the new models are similar to the quantiles coming from a fitting by an exponential for $T \leq 10$ yr. Dealing with (very) extreme values, finding a relevant accuracy test is not an easy task[6].

Arnaud et al. (2008) proposed a simple test which is also used in (Garavaglia et al., 2010). The purpose of this test is to count the number of stations where a given quantile (estimated by the tested method) is exceeded by the maximum observed rainfall. The distribution of the theoretical number of exceedances can be determined assuming the spatial independence of the observed records (See Appendix B). This test has been

[6]The choice of the distribution leads to a too high gap in the quantile estimation for $T \geq 10$ yr.

HESSD
9, 11227–11266, 2012

performed on 414 rain gauge stations[7]. Table 6 shows the number of stations where the T-quantiles are exceeded by the maximum observed rainfall.

The results of the accuracy test show that the model without dependence proposes overestimated extreme quantiles, especially for the sub-daily rainfalls. Unlike the models where the Depth/Duration dependence is modeled, the number of stations where the observed records exceed extreme quantiles is too small compared to the "theoretical" number. Modelling dependence allows the researchers to obtain extreme rainfall quantiles that are coherent with the accuracy test for the sub-daily rainfall. The choice of the copula (Frank versus Gumbel) affects the rainfall quantiles for duration $D > 1$ h, especially when the return period is high. Unlike the Frank copula, the Gumbel copula "amplifies" the dependence in extreme values leading to a clear underestimation for daily rainfall extreme quantiles. An extreme rainfall event is often constructed from a succession of heavy storms (persistence phenomenon in Arnaud et al., 2007). Generating too long durations for these heavy storms causes a decrease of the number of rainstorms occurring in 24 h leading to a decrease of daily rainfall depth.

5 Discussion and conclusions

In this paper, the copula approach is 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 desciption of the dependencies between many random variables, independently of their marginal distributions. It has been applied to generate a relation between the depth and the duration of a rainstorm.

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

[7] the 217 stations presented before + 207 other stations (generally these stations are only used to validate the regionalized model)

another is possible based uniquely on the model's parameters. To continue in the same way, only one copula has been chosen to model the same structure of dependence on any sites. A procedure has been performed to find the best adapted copula among others. For both seasons, the Frank copula appears to be the best for most of the sites, according to our criteria, and has been validated by a formal goodness-of-fit test. No influence seems to be exerted by the rain gauge locations over the dependence structure. Here, two seasons are distinguished into the generator calibration. A future version of the model will distinguish 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.

Taking dependence into account enables the researchers to improve the results of the rainfall model, especially in the sub-daily rainfall generation. The copula choice in the Depth/Duration dependence modelling (Gumbel or Frank) leads to minor impact on the estimation of rainfall T-quantiles with $T \leq 10$ yr for any duration. The two criteria used have shown that the proposed model could reproduce the standard statistics of maximum rainfall for all durations. Indeed biennial or decennial quantiles estimated by the model are close to those estimated by a fitting on observations: the relative errors are centred to 0 and do not exceed $\pm 20\,\%$ for $95\,\%$ of 217 available stations. Stations can be clustered according to three different types of climate: alpine, temperate and Mediterranean to test the model's performance in each type of climate. For the three climates, the Frank copula seems to be the best adapted and the two criteria (relative error and the Nash criterium) are approximately the same for each type of climate. However, simulated hourly rainfall in the alpine climate seems to be overestimated in the high elevation site. This overestimation can be caused by the fact that the snow can lead to inaccurate knowledge of the real rainfall intensity.

An accuracy test for the extreme values has shown the good asymptotic behaviour of the rainfall generator. The number of observed records which exceed a given quantile estimated by the proposed model is in the "theoretical" confidence interval for the Frank copula model while the Gumbel copula seems to model too long durations for the

HESSD
9, 11227–11266, 2012

heaviest storms leading an underestimation on the daily rainfall quantiles. As previously mentioned, this test should be performed in each climate but the authors think the number of stations is too small to apply it. However, the location of the stations where the quantiles have been exceeded can give an idea of the good representation of all types of climate. For example, the stations where the decennial quantile is exceeded are well distributed all over France. It is similarly for the centennial quantiles except for alpine sites where no centennial quantiles are exceeded. It can be explained by the relatively short observation periods (about 10 yr for the alpine stations).

To conclude, the proposed model can reproduce the standard and extreme statistics of maximum rainfall for any duration. Only one parameter has been added compared to the previous version. It has been regionalized and can provide rainfall quantiles on ungauged sites (Arnaud et al., 2006). The additional parameter can be explained by geographical variables and has been regionalized to apply the proposed version of the rainfall generator in the whole France leading to an improvement of the estimation of the hourly rainfall quantiles.

Note that, in this study, copulas are applied to model the Depth/Duration dependence but their application can be extended to many dependencies. In another study, the "persistence" phenomenon (dependence between the depth of rainstorms in a rainy event) introduced by (Arnaud et al., 2007), is modeled by an approach based on copulas, providing good results in extreme rainfall generation.

Appendix A

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 α_g of a multiple tests procedure. Assuming that K tests of a null hypothesis H_0 are achieved, the BH procedure is the following:

HESSD
9, 11227–11266, 2012

1. Let $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(K)}$ 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 $p_{(1)} \leq \ldots \leq p_{(m)}$ p-values; else reject no hypothesis.

Appendix B

An accuracy test for the extreme values

A procedure has proposed to test the pertinence of extreme quantiles estimated by a method. This test can be performed on many stations.

Let be NY^i the number of years of observation at the station i.

Let be NE^i the number of rainfall events observed during the NY^i years at the station i.

Let be $\overline{\mathrm{NE}^i} \overset{\mathrm{def}}{=} \frac{\mathrm{NE}^i}{\mathrm{NA}^i}$: the average number of events per year for the station i.

$X^i = \{X_j^i\}_{j=1\ldots\mathrm{NE}^i}$: the depth of the rainfall observed during the NE^i rainy events.

$X_\mathrm{S}^i = \max\{X_j^i\}_{j=1\ldots\mathrm{NE}^i}$: the maximum rainfall observed at the station i.

q_T^i: the "true" quantile with the return period T years at the station i

$\widehat{q_\mathrm{T}^i}$: the quantile with the return period T years at the station i estimated by the tested method.

N_sup: the number of stations, among N, where q_T^i is exceeded by X_S^i.

$\widehat{n_\mathrm{sup}}$: the number of stations, among N, where $\widehat{q_\mathrm{T}^i}$ is exceeded by X_S^i.

The goal is to define the theoretical law of N_sup.

HESSD
9, 11227–11266, 2012

Assuming that the X_j are independent, we have:

$$\mathbb{P}(X_S^i < x) = \prod_{j=1}^{\mathrm{NE}^i} \mathbb{P}(X_j^i < x)$$
$$= \left[\mathbb{P}(X_j^i < x)\right]^{\mathrm{NE}^i}$$

By definition, $\mathbb{P}(X_j^i < q_T^i) = 1 - \frac{1}{T \times \overline{\mathrm{NE}^i}}$. Therefore, we obtain $\mathbb{P}(X_S^i > q_T^i) = 1 - \left[1 - \frac{1}{T \times \overline{\mathrm{NE}^i}}\right]^{\mathrm{NE}^i}$ On each station, a Bernouilli draw is realized where the success probability $\mathbb{P}(X_S^i > q_T^i)$ depends on $\overline{\mathrm{NE}^i}$ and NE^i. Consequently, this success probability is different from a station to another: N_{sup} does not follow an usual binomial distribution.

The idea is to approach this distribution by a Monte-Carlo method.

When the observed records are considered independent between them, the following procedure is proposed to approximate the theoretical distribution of N_{sup}:

Let be N_{sim} a large number
for $k = 1 : N_{\mathrm{sim}}$ do
    $N_{\mathrm{sup}}^k = 0$
    for $i = 1 : N$ do
        $u \sim \mathcal{U}[0,1]$
        if $u < 1 - \left[1 - \frac{1}{T \times \overline{\mathrm{NE}^i}}\right]^{\mathrm{NE}^i}$ then $N_{\mathrm{sup}}^k = N_{\mathrm{sup}}^k + 1$
    end for
end for

N_{sim} values of the random variable N_{sup} have been performed to obtain $\widetilde{\Pi}(x)$ an approximation of $\Pi(x)$, the theoretical distribution of N_{sup}. Then, it is easy to estimate $\mathbb{P}(N_{\mathrm{sup}} = \widehat{n_{\mathrm{sup}}})$ which can correspond to the p-value in a classical test. In this paper, the

HESSD
9, 11227–11266, 2012

construction of a confidence interval I_{conf} of N_{sup} has been chosen. If $\widehat{n_{sup}} \in I_{conf}$ then the quantile $\widehat{q^i_T}$ estimated by the method appears to be correct according to the data. If $\widehat{n_{sup}}$ is too large (respectively small), the method seems to underestimate (respectively overestimate) rainfall quantiles.

To obtain the independence between the observed records (a strong hypothesis to construct $\widetilde{\Pi}(x)$), we propose to delete the stations where their records occur on the same day (only one station is kept, randomly chosen). Thus, the confidence interval I_{conf} can differ according to the rainfall duration.

References

Adams, B. and Papa, F.: Urban Stormwater Management Planning with Analytical Probabilistic Models, John Wiley & Sons, New York, 2000. 11229

Arnaud, P.: Modèle de prédetermination de crues basé sur la simulation stochastique des pluies horaires, Ph.D. thesis, Université Montpellier II, Montpellier, 1997. 11231

Arnaud, P. and Lavabre, J.: Using a stochastic model for generating hourly hyetographs to study extreme rainfalls, Hydrolog. Sci. J., 44, 433–446, 1999. 11230

Arnaud, P. and Lavabre, J.: Coupled rainfall model and discharge model for flood frequency estimation., Water Resour. Res., 38, 1075–1085, 2002. 11230

Arnaud, P. and Lavabre, J.: Estimation de l'aléa pluvial en France métropolitaine, Editions QUAE, Versailles, 2010. 11231

Arnaud, P., Lavabre, J., Sol, B., and Desouches, C.: Cartographie de l'aléa pluviographique de la France, La Houille Blanche, 5, 102–111, 2006. 11230, 11234, 11246

Arnaud, P., Fine, J., and Lavabre, J.: An hourly rainfall generation model adapted to all types of climate., Atmos. Res., 85, 230–242, 2007. 11230, 11231, 11232, 11233, 11238, 11239, 11244, 11246, 11261

Arnaud, P., Lavabre, J., Sol, B., and Desouches, C.: Régionalisation d'un générateur de pluies horaires sur la France métropolitaine pour la connaissance de l'aléa pluviographique/Regionalization of an hourly rainfall generating model over metropolitan France for flood hazard estimation, Hydrolog. Sci. J., 53, 34–47, 2008. 11243

Bacchi, B., Becciu, G., and Kottegoda, N.: Bivariate exponential model applied to intensities and durations of extreme rainfall, J. Hydrol., 155, 225–236, doi:10.1016/0022-1694(94)90166-X, 1994. 11229

Balistrocchi, M. and Bacchi, B.: Modelling the statistical dependence of rainfall event variables through copula functions, Hydrol. Earth Syst. Sci., 15, 1959–1977, doi:10.5194/hess-15-1959-2011, 2011. 11230

Bárdossy, A. and Pegram, G. G. S.: Copula based multisite model for daily precipitation simulation, Hydrol. Earth Syst. Sci., 13, 2299-2314, doi:10.5194/hess-13-2299-2009, 2009. 11229

Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Stat. Soc. B Met., 57, 289–300, 1995. 11246

Cantet, P., Bacro, J., and Arnaud, P.: Using a rainfall stochastic generator to detect trends in extreme rainfall, Stoch. Env. Res. Risk A., 25, 429–441, 2011. 11233

Carvajal, C., Peyras, L., and Arnaud, P., Boissier, D., and Royet, P.: Probabilistic Modeling of Floodwater Level for Dam Reservoirs, J. Hydrol. Eng., 14, 223–232, 2009. 11233

Cernesson, F.: Modèle simple de prédetermination des crues de fréquences courantes à rares sur petits bassins versants méditéranéens., Ph.D. thesis, Université Montpellier II, Montpellier, 1993. 11231

Cernesson, F., Lavabre, J., and Masson, J.: Stochastic model for generating hourly hyetograph, Atmos. Res., 42, 149–161, 1996. 11230, 11231, 11232

Clayton, D.: A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, Biometrika, 65, 141–151, 1978. 11236

Córdova, J. R. and Rodríguez-Iturbe, I.: On the probabilistic structure of storm surface runoff, Water Resour. Res, 21, 755–763, doi:10.1029/WR021i005p00755, 1985. 11228

De Michele, C. and Salvadori, G.: A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas, J. Geophys. Res., 108, 4067, doi:10.1029/2002JD002534, 2003. 11229

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R.: Bivariate statistical approach to spillway design flood, J. Hydraul. Eng. ASCE, 10, 50–57, 2005. 11229

Denuit, M. and Lambert, P.: Constraints on concordance measures in bivariate discrete data, J. Multivariate Anal., 93, 40–57, 2005. 11238

Devroye, L.: Non-Uniform Random Variate Generation, Springer-Verlag, New York, 1986. 11237

Díaz-Granados, M. A., Valdes, J. B., and Bras, R. L.: A physically based flood frequency distribution, Water Resour. Res., 20, 995–1002, doi:10.1029/WR020i007p00995, 1984. 11228
Eagleson, P.: Dynamics of flood frequency, Water Resour. Res., 8, 878–898, 1972. 11228
Eagleson, P.: Climate, soil, and vegetation. 2. The distribution of annual precipitation derived from observed storm sequences, Water Resour. Res., 14, 713–721, 1978a. 11228
Eagleson, P.: Climate, soil, and vegetation. 5. A derived distribution of storm surface runoff, Water Resour. Res., 14, 741–748, 1978b. 11228
Eagleson, P.: Climate, soil, and vegetation. 7. A derived distribution of annual water yield, Water Resour. Res., 14, 765–776, 1978c. 11228
Embrechts, P., Lindskog, F., and Mcneil, A.: Modelling dependence with copulas and applications to risk management, Handbook of Heavy Tail Distributions in Finance, North Holland, Amsterdam, 324–384, 2003. 11238
Favre, A., El Adlouni, S., Perreault, L., Thiemonge, N., and Bobee, B.: Multivariate hydrological frequency analysis using copulas, Water Resour. Res., 40, W01101, 12PP, 2004. 11229
Fine, J. and Lavabre, J.: Synthèse des débits de crue sur l'Ile de la Réunion. Phase I : la pluviométrie. Éléments de régionalisation du générateur de pluie., Tech. rep., Cemagref, Aix en Provence, 2002. 11232
Fouchier, C.: 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, vol. 9, p. 02843, Copernicus, Gottingen, 2007. 11234
Frank, M.: On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$, Aeq. Math., 19, 194–226, 1979. 11236
Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garçon, R., and Bernardara, P.: Introducing a rainfall compound distribution model based on weather patterns sub-sampling, Hydrol. Earth Syst. Sci., 14, 951-964, doi:10.5194/hess-14-951-2010, 2010. 11243, 11245
Gargouri-Ellouze, E. and Chebchoub, A.: 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, Hydrolog. Sci. J., 53, 802–817, 2008. 11229
Genest, C. and Favre, A.: Everything you always wanted to know about copula modeling but were afraid to ask, J. Hydrol. Eng., 12, 347–368, doi:10.1061/(ASCE)1084-0699(2007)12:4(347), 2007. 11234

HESSD
9, 11227–11266, 2012

Modelling dependence of rainfall variables into a stochastic model

P. Cantet and P. Arnaud

Genest, C. and Rémillard, B.: Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models, Annales de l'Institut Henri Poincaré-Probabilités et Statistiques, 44, 1096–1127, 2008. 11237, 11241

Genest, C. and Rivest, L.: Statistical inference procedures for bivariate Archimedean copulas, J. Am. Stat. Assoc., 88, 1034-1043, 1993. 11237

Genest, C., Ghoudi, K., and Rivest, L.: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika, 82, 543–552, 1995. 11237

Genest, C., Masiello, E., and Tribouley, K.: Estimating copula densities through wavelets, Insur. Math. Econ., 44, 170–181, 2008a. 11237

Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, Insur. Math. Econ., 44, 199–213, 2008b. 11237

Ghosh, S.: Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula, Hydrol. Process., 24, 3558–3567, 2010. 11229

Goel, N. K., Kurothe, R. S., Mathur, B. S., and Vogel, R. M.: A derived flood frequency distribution for correlated rainfall intensity and duration, J. Hydrol., 228, 56–67, doi:10.1016/S0022-1694(00)00145-1, 2000. 11229

Gumbel, E.: Bivariate logistic distributions, J. Am. Stat. Assoc., 56, 335–349, 1961. 11236

Guo, Y. and Adams, B. J.: An analytical probabilistic approach to sizing flood control detention facilities, Water Resour. Res, 35, 2457–2468, doi:10.1029/1999WR900125, 1999. 11228

Haberlandt, U., Ebner von Eschenbach, A.-D., and Buchwald, I.: A space-time hybrid hourly rainfall model for derived flood frequency analysis, Hydrol. Earth Syst. Sci., 12, 1353–1367, doi:10.5194/hess-12-1353-2008, 2008. 11229

Haberlandt, U., Hundecha, Y., Pahlow, M., and Schumann, A. H.: Rainfall Generators for Application in Flood Studies, Flood Risk Assessment and Management, edited by: Schumann, A. H., Springer The Netherlands, 117–147, doi:10.1007/978-90-481-9917-4_7, 2011. 11229

Hillali, Y.: Test d'ajustement d'une loi bidimensionnelle; Application a des donnees climatologiques, Rev. Stat. Appl., 49, 79–96, 2001. 11237

Hoeffding, W.: Maszstabinvariante Korrelationstheorie, Schrijtfenr. Math. Inst. Inst., Angew. Math. Univ. Berlin, 5, 179–233, 1940. 11229, 11235

Javelle, P., Fouchier, C., Arnaud, P., and Lavabre, J.: Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations, J. Hydrol., 394, 267–274, 2010. 11234

Joe, H.: Multivariate models and dependence concepts, Chapman & Hall/CRC, London, 1997. 11237

Kurothe, R. S., Goel, N. K., and Mathur, B. S.: Derived flood frequency distribution for negatively correlated rainfall intensity and duration, Water Resour. Res., 33, 2103–2107, 1997. 11229

Li, J. Y. and Adams, B. J.: Probabilistic models for analysis of urban runoff control systems, J. Environ. Eng., 126, 3, 217–224, doi:10.1061/(ASCE)0733-9372(2000)126:3(217), 2000. 11228

McNeil, A.: Sampling nested Archimedean copulas, J. Stat. Comput. Sim., 78, 567–581, 2008. 11238

Muller, A., Arnaud, P., Lang, M., and Lavabre, J.: Uncertainties of extreme rainfall quantiles estimated by a stochastic rainfall model and by a generalized Pareto distribution, Hydrolog. Sci., 54, 417–429, 2009. 11230, 11231

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, 1970. 11242

Nelsen, R.: An introduction to copulas, Springer Verlag, New York, 2006. 11229, 11234

Neppel, L., Arnaud, P., and Lavabre, J.: Connaissance régionale des pluies extrêmes. Comparaison de deux approches appliquées en milieu méditéranéens, C. R. Geoscience, 339, 820–830, 2007. 11230, 11231

Oakes, D.: A model for association in bivariate survival data, J. Roy. Stat. Soc. B, 44, 414–422, 1982. 11234

Salvadori, G. and De Michele, C.: Frequency analysis via copulas: Theoretical aspects and applications to hydrological events, Water Resour. Res., 40, W12511, doi:10.1029/2004WR003133, 2004. 11229

Salvadori, G. and De Michele, C.: Statistical characterization of temporal structure of storms, Adv. Water Resour., 29, 827–842, 2006. 11229

Salvadori, G., de Michele, C., Kottegoda, N., and Rosso, R.: Extremes in Nature: An Approach Using Copulas, Springer, Dordrecht, The Netherlands, 2007. 11229

Salvadori, G., De Michele, C., and Durante, F.: On the return period and design in a multivariate framework, Hydrol. Earth Syst. Sci., 15, 3293–3305, doi:10.5194/hess-15-3293-2011, 2011. 11229

Singh, K. and Singh, V. P.: Derivation of bivariate probability density functions with exponential marginals, Stoch. Hydrol. Hydraul., 5, 55–68, doi:10.1007/BF01544178, 1991. 11229

HESSD

9, 11227–11266, 2012

Sklar, A.: Fonctions de répartition á n dimensions et leurs marges, Publ. Inst., Statistics Univ. Paris, 8, 229–231, 1959. 11229, 11235, 11236

Tsukahara, H.: Semiparametric estimation in copula models, Can. J. Stat., 357–375, 2005. 11237

Vandenberghe, S., Verhoest, N., and De Baets, B.: 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, doi:10.1029/2009WR007857, 2010. 11229

Vandenberghe, S., Verhoest, N., Onof, C., and De Baets, B.: A comparative copula-based bivariate frequency analysis of observed and simulated storm events: A case study on Bartlett-Lewis modeled rainfall, Water Resour. Res., 47, W07529, doi:10.1029/2009WR008388, 2011. 11229

Whelan, N.: Sampling from Archimedean copulas, Quant. Financ., 4, 339–352, 2004. 11237

Wu, S., Tung, Y., and Yang, J.: Stochastic generation of hourly rainstorm events, Stoch. Env. Res. Risk A., 21, 195–212, 2006. 11230

Yan, J.: Enjoy the Joy of Copulas: With a Package copula, J. Stat. Softw., 21, 1–21, http://www.jstatsoft.org/v21/i04/, 2007. 11237

Zhang, L. and Singh, V.: Bivariate rainfall frequency distributions using Archimedean copulas, J. Hydrol., 332, 93–109, 2007. 11229

Table 1. Definition of the four archimedean copulas with their parameter (θ) space, and an expression for the population value of Kendall's tau (τ).

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

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD
9, 11227–11266, 2012

Modelling dependence of rainfall variables into a stochastic model

P. Cantet and P. Arnaud

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

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

HESSD
9, 11227–11266, 2012

Table 3. The number of stations where a copula is associated with a rank. The rank 1 corresponds to the smaller L^2-distance between the empirical copula and the theoretical copula.

Season	Rank	Frank	Gumbel	Clayton	Total
Winter	Rank 1	67	44	29	140
	Rank 2	73	44	23	140
	Rank 3	0	52	88	140
Summer	Rank 1	45	17	19	81
	Rank 2	36	22	23	81
	Rank 3	0	42	39	81

HESSD
9, 11227–11266, 2012

Table 4. Information on the 3 stations illustrated in Fig. 4.

Station	X(m)	Y(m)	Altitude (m)	Location	τ in Winter	τ in Summer
4809	705.0	1917.4	913	Barre des Cévennes	0.31	0.23
3809	875.05	2044.5	1700	La Scia	0	0.29
38VI	853.6	2013.4	1050	Villard-de-Lans	0.23	0.22

HESSD
9, 11227–11266, 2012

Table 5. Value of the Nash criterion between Q^T_{obs} and Q^T_{RG} (Eq. 6) for 1-hour, 6-hours and 24-hours maximal rainfall at the 217 rain-gauge stations studied with $T = 2, 5, 10$ yr.

Rainfall duration	Return period	Without dependence	Frank Copula	Gumbel Copula
1 h	$T = 2$ yr	0.7	0.92	0.92
	$T = 5$ yr	0.52	0.86	0.89
	$T = 10$ yr	0.35	0.77	0.83
6 h	$T = 2$ yr	0.97	0.98	0.97
	$T = 5$ yr	0.95	0.96	0.94
	$T = 10$ yr	0.92	0.94	0.92
24 h	$T = 2$ yr	0.96	0.95	0.95
	$T = 5$ yr	0.96	0.95	0.94
	$T = 10$ yr	0.95	0.95	0.94

HESSD
9, 11227–11266, 2012

Table 6. Number of stations where the maximum observed rainfall exceed the T-quantiles estimated by the 3 models with $T = 10, 100, 1000$ yr. The range of the theoretical number is estimated by the 0.025 and 0.975 quantiles of the sample generated by Monte-Carlo simulations (See Appendix B).

Return Period	Rainfall duration	"Theoretical"	Without Dependence	Frank Copula	Gumbel Copula
$T = 10$ yr	1 h	[208, 232]	180	215	216
	6 h	[183, 205]	175	190	198
	24 h	[160, 183]	157	164	179
$T = 100$ yr	1 h	[31, 50]	20	40	42
	6 h	[29, 50]	32	45	55
	24 h	[24, 39]	18	24	46
$T = 1000$ yr	1 h	[1, 7]	1	4	4
	6 h	[1, 7]	4	6	12
	24 h	[1, 7]	0	2	13

HESSD
9, 11227–11266, 2012

Modelling dependence of rainfall variables into a stochastic model

P. Cantet and P. Arnaud

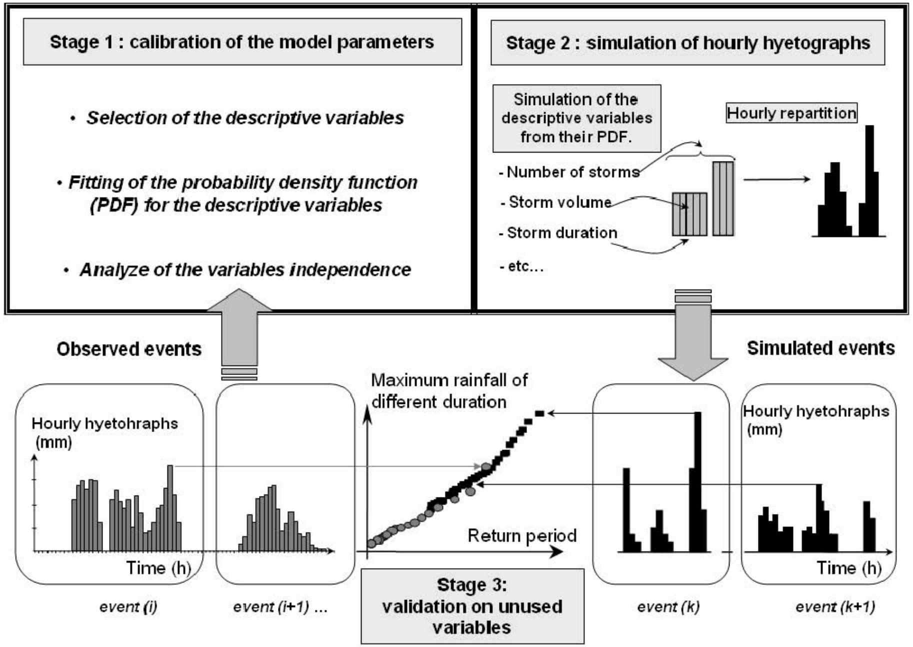

Fig. 1. Principle of the hourly rainfall generator. Figure from (Arnaud et al., 2007).

HESSD
9, 11227–11266, 2012

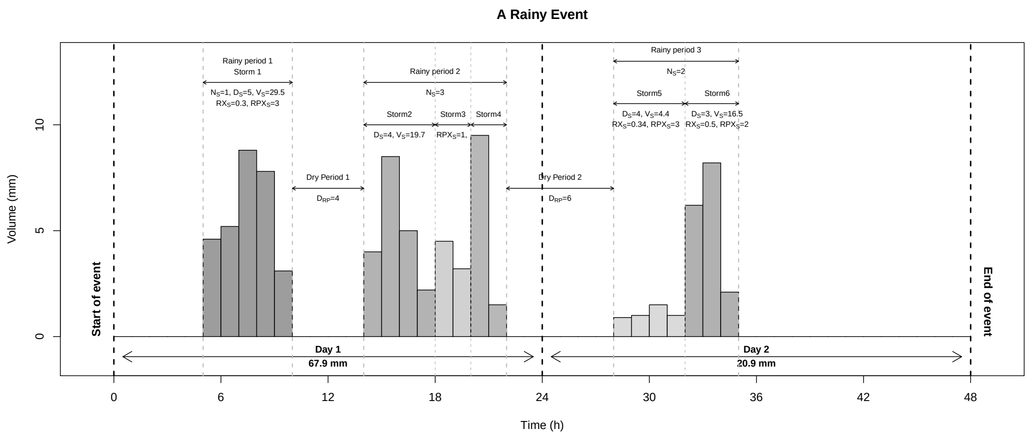

Fig. 2. Illustration of a rainy event (88.8 mm in 2 days) with 3 rainy periods ($N_{RP} = 3$) and 6 storms (including 2 "major" storms: Storm1 and Storm2).

HESSD
9, 11227–11266, 2012

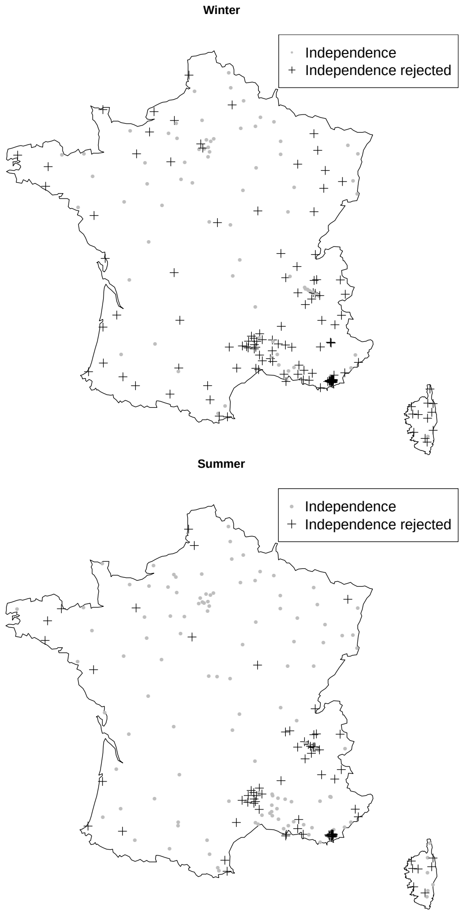

Fig. 3. Results of the independence test ($\alpha = 0.05$) based on the Kendall's tau between the depth and duration of major storm tested on each 217 rain gauge stations with the False Discovery Rate approach.

HESSD
9, 11227–11266, 2012

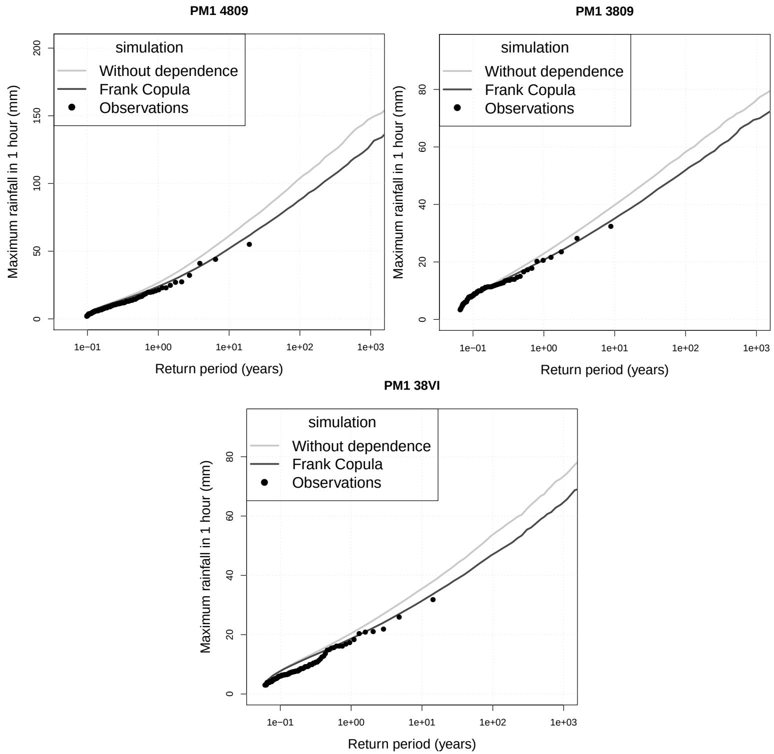

Fig. 4. Frequency distribution for 1-h maximum rainfall observed (points) and drawn from simulated hyetographs for both models (lines).

HESSD
9, 11227–11266, 2012

HESSD
9, 11227–11266, 2012

Modelling dependence of rainfall variables into a stochastic model

P. Cantet and P. Arnaud

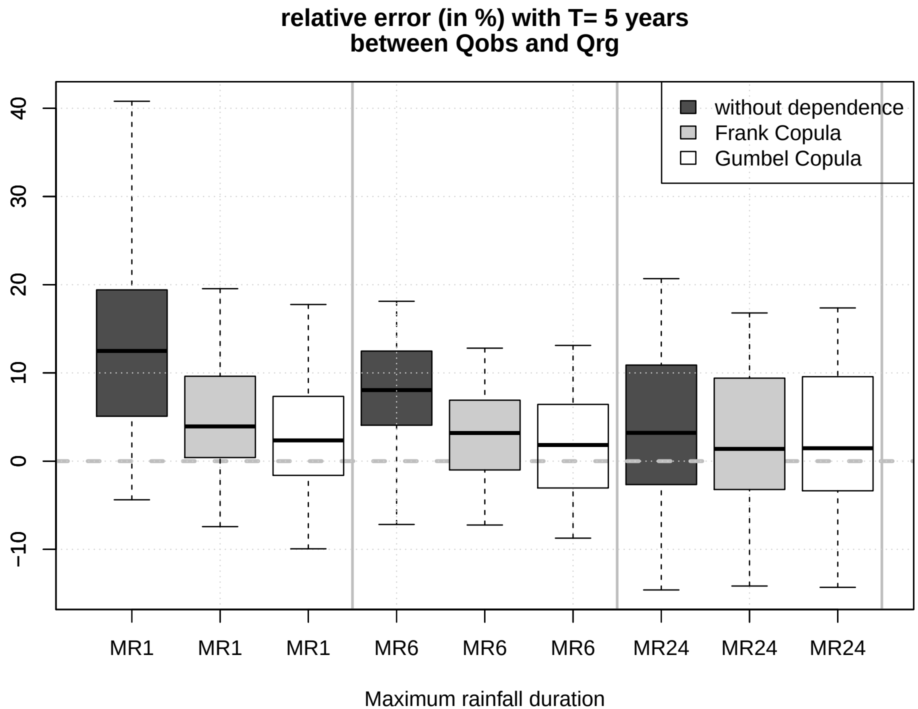

Fig. 5. Boxplots of relative errors (in %) between Q^{10}_{obs} and Q^{10}_{RG} (Eq. 5) according to the choice of the model (independence in dark grey, Frank copula in light grey, and Gumbel copula in white) for the maximal rainfall in 1 h (MR1), 6 h (MR6) or 24 h (MR24).

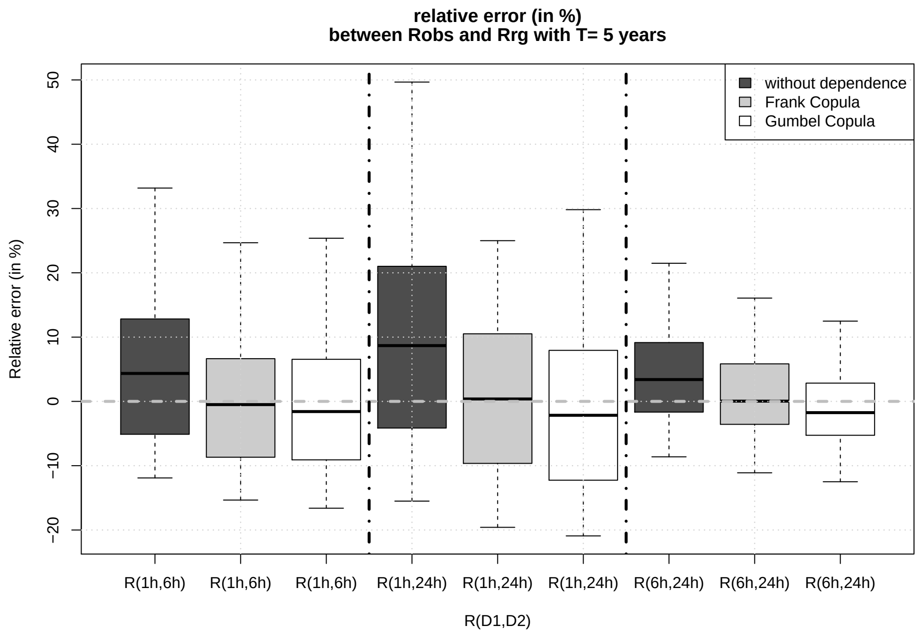

Fig. 6. Boxplots of relative errors (in %) between R^5_{obs} and R^5_{RG} (Eq. 7) according to the choice of the model (independence in dark grey, Frank copula in light grey, and Gumbel copula in white) for the ratio MR1/MR6 (R(1 h, 6 h)), MR1/MR24 (R(1 h, 24 h)) or MR6/MR24 (R(6 h, 24 h)).

HESSD
9, 11227–11266, 2012