

A stochastic design rainfall generator based on copulas and mass curves

S. Vandenberghe¹, N. E. C. Verhoest¹, E. Buyse¹, and B. De Baets²

¹Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, 9000 Ghent, Belgium

²Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, 9000 Ghent, Belgium

Received: 1 June 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 22 June 2010

Revised: 22 November 2010 – Accepted: 23 November 2010 – Published: 3 December 2010

Abstract. The use of design storms can be very useful in many hydrological and hydraulic practices. In this study, the concept of a copula-based secondary return period in combination with the concept of mass curves is used to generate point-scale design storms. The analysis is based on storms selected from the 105 year rainfall time series with a 10 min resolution, measured at Uccle, Belgium. In first instance, bivariate copulas and secondary return periods are explained, together with a focus on which couple of storm variables is of highest interest for the analysis and a discussion of how the results might be affected by the goodness-of-fit of the copula. Subsequently, the fitted copula is used to sample storms with a predefined secondary return period for which characteristic variables such as storm duration and total storm depth can be derived. In order to construct design storms with a realistic storm structure, mass curves of 1st, 2nd, 3rd and 4th quartile storms are developed. An analysis shows that the assumption of independence between the secondary return period and the internal storm structure could be made. Based on the mass curves, a technique is developed to randomly generate an intrastorm structure. The coupling of both techniques eventually results in a methodology for stochastic design storm generation. Finally, its practical usefulness for design studies is illustrated based on the generation of a set of statistically identical design storm and rainfall-runoff modelling.

1 Introduction

Engineers involved in the design of hydraulic structures in river systems are often confronted with a lack of available data regarding the phenomenon under study, e.g. peak discharges at a specific point in a catchment. As rainfall data are often readily available, they often serve as an indispensable source of information for the further analysis. A variety of point rainfall data products can be used in such design studies: the historical time series, a synthetically generated time series, intensity-duration-frequency (IDF) relations and design storms.

A method for generating design storms should be flexible enough to entail the variability of several rainfall characteristics. To that end, design storms are mostly characterized by a specific return period, a rainfall volume or intensity, and a duration, which are related to the extremity of the storm, and a temporal rainfall pattern or an internal storm structure (Chow et al., 1988).

The internal storm structure (Bernardara et al., 2007; Thompson et al., 2002; Koutsoyiannis, 1994, 1993) is often the most important characteristic and several methods exist for its characterization (see e.g., Prodanovic and Simonovic, 2004; Veneziano, 1999; Chow et al., 1988; Pilgrim and Cordery, 1975, and references therein): the use of an arbitrary (symmetrical) pattern in combination with an average intensity derived from the IDF curves, the construction of a pattern out of the complete IDF curves, using simulations of stochastic models, or by using standardized profiles (summation curves) derived from an empirical probabilistic analysis of the rainfall. The method of Huff (1967), which falls in the latter category, will be used in this study. Mass curves, often referred to as Huff curves, are representations of the normalized time versus the normalized cumulative storm depth since the beginning of a



Correspondence to: S. Vandenberghe
(sander.vandenberghe@ugent.be)

storm. At regular moments in the storm, e.g. when partitioning a storm in 20 identical time intervals at every 5% of the total storm duration, the empirical distribution of the normalized cumulative storm depth is evaluated. To obtain the Huff curves, often the 10%, 50% and 90% percentiles of that distribution are then visualized. Furthermore, a classification of storms into quartile groups can be made, according to which quarter of the storms received the largest amount of rainfall.

Besides the internal storm structure, also the duration and mean storm intensity or storm depth should be calculated for a specific design storm. In practice, this is often done by fixing a specific design return period. Subsequently, a certain storm duration is fixed, which could be based on the concentration time of the catchment under study, and the corresponding mean storm intensity or storm depth for that design return period is then retrieved from previously established IDF relations. In this study, the recently proposed framework for a multivariate copula-based frequency analysis (Salvadori et al., 2007; Salvadori and De Michele, 2004) is used to establish a direct link between a physical storm duration, its depth or mean intensity, and the corresponding return period. The work of Ellouze et al. (2009), Gargouri-Ellouze and Chebchoub (2008), Kao and Govindaraju (2008) and Grimaldi and Serinaldi (2006) already considered the use of copulas in the modelling of design storms.

All analyses in this work are based on a 105 year rainfall time series with a 10 min resolution, measured at Uccle (Brussels), Belgium. These rainfall measurements have already been the subject of several studies, generally focusing on either the construction of IDF relations (Willems, 2000) or, more recently, on climate change and the detection of trends in rainfall observations (Ntegeka and Willems, 2008; De Jongh et al., 2006). In the context of design storms, the development of composite storms for Flanders based on the Uccle rainfall has been of great importance for Flemish water managers and engineers involved in urban drainage or river design (Vaes and Berlamont, 2000).

This work proposes a stochastic design rainfall generator by combining the traditional concept of Huff curves for the analysis of the internal storm structure with the concept of a copula-based secondary return period of a storm. In the following, firstly a copula-based secondary return period is assigned to each Uccle storm and some remarks on the choice of the couple of storm variables and the importance of a good fit are made (Sect. 2). Secondly, the internal storm structure of the observed storms and its independence of the return period is analyzed and an algorithm for a random internal storm structure generation is proposed (Sect. 3). Section 4 then illustrates how the design storm generator could be used for water management purposes. Finally, conclusions and recommendations for future research are given in Sect. 5.

2 Assignment of return periods to Uccle storms

2.1 Storm selection

A first step is the delineation of individual and statistically independent storms in the 105 year time series of 10 min rainfall. Therefore, a minimal dry period, or critical dry duration, should be defined. All dry periods which are shorter than this criterion are considered to belong to the same storm (Bonta and Rao, 1988). Here, a 24-h dry period criterion is chosen, based on the method of Restrepo-Posada and Eagleson (1982), which assumes that the arrival times of statistically independent storms follow a Poisson distribution. However, from a design perspective, one could consider to use another independence criterion dictated by the application in which the storms are to be used, e.g. storms should be separated by a dry period at least as long as the concentration time of the catchment under study. For reasons of consistency, the same storm selection procedure as applied by Vandenberghe et al. (2010a,b) is used.

For each storm, several variables are calculated, such as the total storm depth D (mm), the storm duration W (h) and the mean storm intensity I (mm/h). In order to be able to analyze Huff curves (see Sect. 3), for each storm, the rainfall depth in every 10 min interval is made cumulative and the time lapse within the storm is expressed as a fraction of the total duration of the storm. Then cumulative rainfall depths in each 5% interval of the storm duration, which do not correspond with the normalized 10-min intervals, are assigned using a linear interpolation of the cumulative 10-min rainfall amounts. Subsequently, each storm is assigned to a specific quartile group, depending on which quarter of the storm has the highest rainfall depth, where storms with equal rainfall amounts in different quarters are not further considered. Because of the assumption of stationarity, the further analysis is based on a seasonal division of storms: winter is defined as December, January and February. March, April and May make up the spring season. June, July and August are then considered as the summer months, while September, October and November are assigned to autumn. Eventually, this results in 1777 winter, 1652 spring, 1647 summer and 1697 autumn storms (Table 1).

2.2 Copulas and secondary return periods

In order to assign a return period to each storm in the data set, bivariate copulas and secondary return periods are used (Salvadori et al., 2007; Salvadori and De Michele, 2004; Salvadori, 2004). This means that a storm is described by two variables, which can be dependent, and that a copula is used to construct their bivariate cumulative distribution function, which in its turn is used to calculate specific probabilities of occurrence of each storm. Section 2.2.1 shortly explains

Table 1. Number of storms, considering different seasons and quartile-groups.

Season	Quartile group				
	1st	2nd	3rd	4th	All
Year	2662	1397	1237	1477	6773
Winter	664	335	324	454	1777
Spring	659	356	307	330	1652
Summer	679	327	299	342	1647
Autumn	660	379	307	351	1697

the concepts of a copula and the secondary return period. Section 2.2.2 provides some reflections on which couple of storm variables is best used in the present context, whereas Sect. 2.2.3 illustrates the impact of the quality of the copula fit on the interpretation of the secondary return period.

2.2.1 Concepts

A bivariate copula C models the dependence structure between two random variables X and Y , which can in the present context be thought of as the total storm depth D , the mean storm intensity I or the storm duration W . It is a function that couples the marginal cumulative distribution functions (CDFs) $F_X(x)$ and $F_Y(y)$ in order to construct the bivariate CDF $F_{X,Y}(X, Y)$, as expressed by the theorem of Sklar (1959):

$$\mathbb{P}(X \leq x, Y \leq y) = F_{X,Y}(x, y) = C(F_X(x), F_Y(y)) \tag{1}$$

$$= C_{UV}(u, v)$$

A bivariate copula is thus a bivariate cumulative distribution function with uniform marginals U and V on the unit interval. One can also determine U and V non-parametrically as the normalized ranks (Genest and Favre, 2007). With the use of the inverse marginal CDFs the corresponding x and y for specific values u and v can be calculated: $x = F_X^{-1}(u)$ and $y = F_Y^{-1}(v)$.

In order to come to the definition of the secondary return period, let us consider a storm for which either X or Y (or both) exceed a respective threshold x and y . This is called the OR-case: $\{X > x \text{ or } Y > y\}$ (Salvadori et al., 2007). In a marginal-free context this event is given as $\{U > u \text{ or } V > v\}$. The probability of occurrence of a storm in the OR-case can then be calculated as follows:

$$\mathbb{P}(F_X(x) > u, F_Y(y) > v) = 1 - C(u, v) \tag{2}$$

With the knowledge of the mean storm interarrival time ω_T , the so-called primary return period in the OR-case T_{OR} can then be calculated:

$$T_{OR} = \frac{\omega_T}{1 - C_{UV}(u, v)} = \frac{\omega_T}{1 - t} \tag{3}$$

It should be clear that different combinations of u and v can result in the same T_{OR} , as long as they have the same probability t .

Eventually, the secondary return period was proposed as being very useful for design purposes (Salvadori and De Michele, 2004; Salvadori, 2004; Salvadori et al., 2007). Unfortunately, this type of return period did not yet find its way to engineering applications, except for some considerations in the work of Vandenberghe et al. (2010b).

In engineering applications, one usually chooses a design storm with a certain (primary) return period for which the design should hold. Consider now the storms in the OR-case. By fixing a certain design return period T_{OR}^* , a certain level t^* of the copula is fixed. In Fig. 1 this is indicated with the thick contour line (for $t=0.4$). A storm that lies on this curve has a return period T_{OR}^* and is called a critical storm (Salvadori et al., 2007). In Fig. 1 S^* is such a critical storm and is defined by the critical thresholds u^* and v^* . A more extreme storm with a higher return period, and thus on a higher contour line, is called a super-critical storm, e.g. S_1^+ and S_2^+ . On the other hand, the storms S_1^- and S_2^- have lower return periods and are called sub-critical storms. The secondary return period is now defined as the average time between the occurrence of two super-critical storms and is expressed as follows:

$$T_{SEC} = \frac{\omega_T}{1 - K_C(t^*)} = \frac{\omega_T}{K_C(t^*)} \tag{4}$$

The function K_C is the distribution function of the random variable $Z=C(U, V)$, i.e. $K_C(z)=\mathbb{P}\{Z \leq z\}$. For Archimedean copulas this function can easily be calculated:

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)} \quad \text{with } 0 < t \leq 1 \tag{5}$$

where $\varphi'(t^+)$ is the right derivative of the additive generator φ (Nelsen, 2006).

Similar to the equation used for calculating the primary return period (Eq. 3), the denominator in Eq. (4) expresses a probability, as $K_C(t^*)$ can be interpreted as the probability that a super-critical storm will occur at any realisation of a storm (Salvadori et al., 2007). This probability mass is given in dark grey in Fig. 1a.

In the calculation of the primary return period in the OR-case, one uses the probability that X or/and Y will exceed a respective threshold x and y . This probability can be expressed as $1 - C_{UV}(u, v)$ (see Eq. 3). However, this probability is not the same as the probability of the occurrence of a super-critical storm that is used for the calculation of the secondary return period, i.e. $1 - K_C(t)$. Figure 1b gives the area in dark grey on which the probability mass, i.e. $1 - C(u^*, v^*)$, is calculated. The primary return period T_{OR} will then be the average time between the occurrence of two successive storms in this region, which is defined by the critical thresholds u^* and v^* . It is obvious that the probability mass corresponding to the dark grey area in

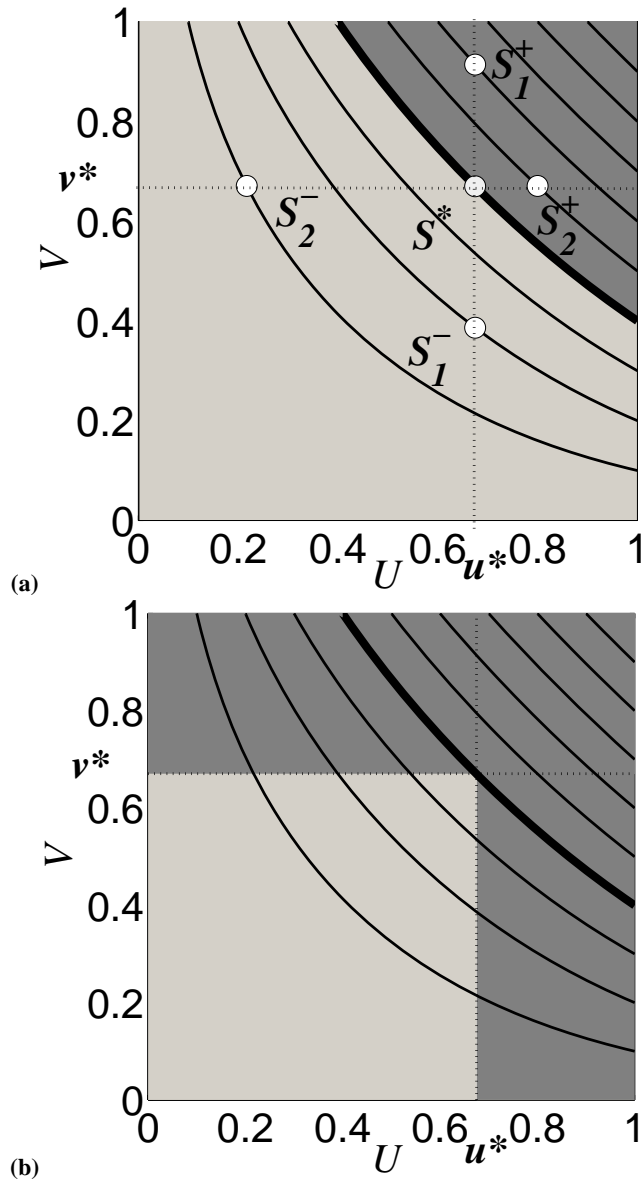


Fig. 1. Illustration of sub-, super- and critical storms (a). Difference in probability used for secondary and primary return periods (dark gray area (a) vs. (b)). (After Salvadori et al., 2007).

Fig. 1a is smaller than the probability mass of the dark grey area in Fig. 1b, for the same critical thresholds. In practice, it is advised to use this probability of super-critical events, as these are in fact dangerous for the design, and hence the secondary return period is a more realistic concept. Note that the function K_C has recently been applied in the work of Kao and Govindaraju (2010) as it is a compelling univariate summary of multivariate information.

An additional comment should be made with regard to the seasonal analysis as carried out in this paper. To calculate the return period (primary and secondary) of e.g. summer

storms one should incorporate the probability of occurrence of a summer storm. This probability can easily be calculated by dividing the number of summer storms by the total number of storms (see Table 1, i.e. $1647/6773=0.24$). For winter, spring and autumn storms this is 0.26, 0.24 and 0.25, respectively. Intuitively, one should expect a probability of 0.25 for each season in a climate without dry and wet regimes. When a different copula is fitted for each season, the probability of occurrence of an extreme event given by the denominators in Eqs. (3) and (4) should thus be multiplied by the probability of occurrence of a storm in the considered season. The mean interarrival time is that of the storms in the considered season. The general equation for calculating a seasonal return period T_s is then:

$$T_s = \frac{\omega_{T,s}}{\mathbb{P}_{\text{seasonal storm}} \times \mathbb{P}_{\text{survival}}} \quad (6)$$

where $\omega_{T,s}$ is the mean interarrival time of the storms in the specific season, $\mathbb{P}_{\text{seasonal storm}}$ the probability of occurrence of a storm in that season and $\mathbb{P}_{\text{survival}}$ the survival probability as given by the denominators of Eqs. (3) and (4).

On the contrary, when the probability of occurrence of a storm in a specific season would not be taken into account, the length of the data set should be considered to be $105/4=26.25$ years, which of course would influence the further interpretation of the results (i.e., the most extreme summer storm within the 105 year series would be assigned a return period of 26.25 year instead of 105 year).

2.2.2 Choice of the couple to fit a copula

To calculate the primary and secondary return periods of each storm, a copula should be fitted to the dependence structure between two rainfall characteristics on a seasonal basis. Three different couples of storm variables can be used: (I, W) , (W, D) and (I, D) . a way to visualize this dependence structure is by making a normalized rank scatter plot, which is in fact the support of the empirical copula (Genest and Favre, 2007). Figure 2 shows these plots for the three couples of summer storm variables. As the secondary return period focuses on storms in the upper-right quadrant of the copula, it only makes sense to consider two random variables that are positively associated. Therefore, it would be an irrational choice to consider the couple (I, W) for the further analysis.

In what follows, the couple (W, D) will be considered for several reasons. Firstly, for (W, D) more storms are located in the upper-right quadrant compared to the other couples, which will contribute to the reliability of the further analysis. Secondly, it still allows for the calculation of I and resembles the approach of traditional frequency analyses where the distribution of rainfall amounts for different aggregation levels is considered (Willems, 2000). Thirdly, no problems with asymmetrical empirical copulas occur for this couple (Vandenberghe et al., 2010a).

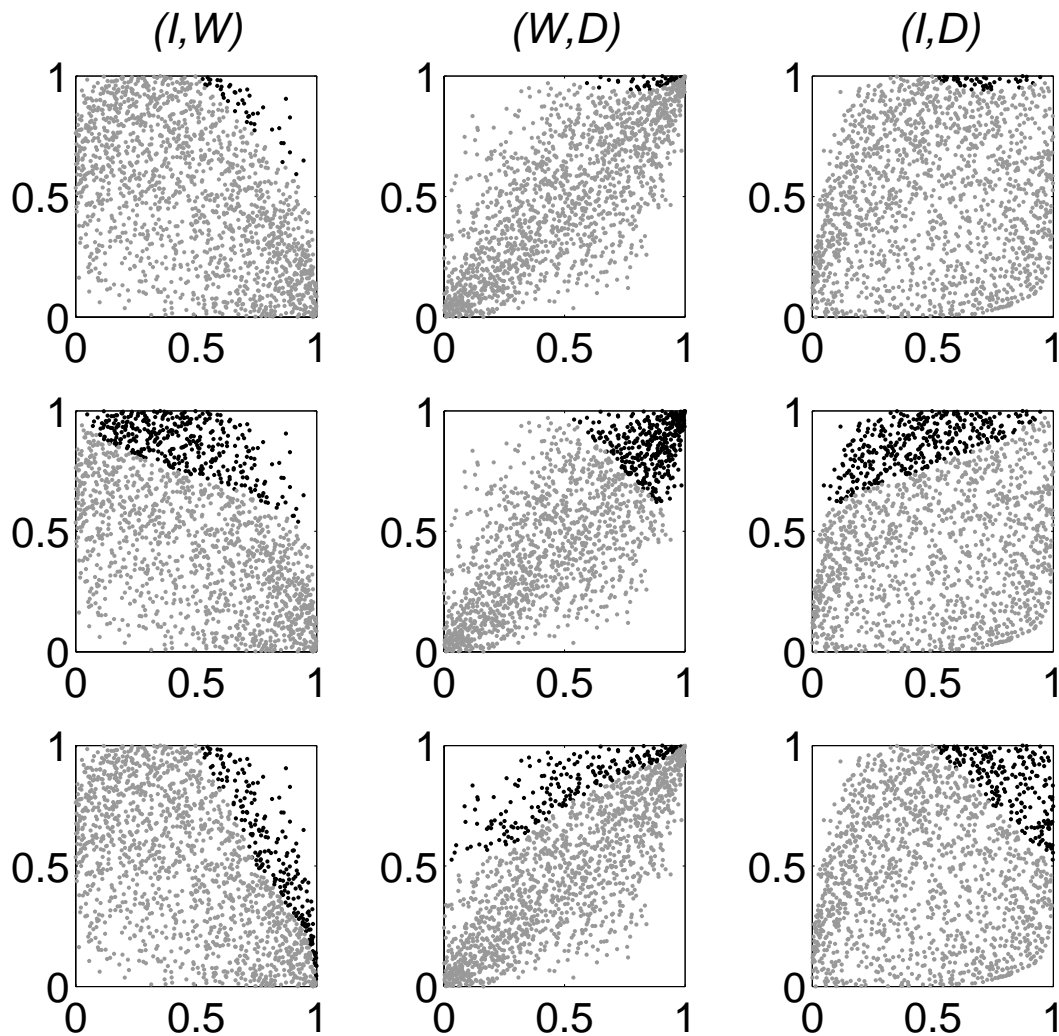


Fig. 2. Normalized rank scatter plots of three different couples of storm variables (per column). Each row shows the same plots, with different storms highlighted within each row. In the first row, the extreme storms with respect to I and W are selected (i.e., storms found in the upper-right corner of the (I, W) plane), and are indicated in black in each of the plots of the first row. For the second and the third row, extreme storms with respect to, respectively (W, D) and (I, D) are selected, and marked in black.

The A12 copula family (Nelsen, 2006) is considered here for modelling the dependence between W and D . The ability of this copula family to model the dependence between rainfall characteristics has been studied by Vandenberghe et al. (2010a). The A12 copula family is given by

$$C_{UV}(u, v) = \left(1 + \left[(u^{-1} - 1)^\theta + (v^{-1} - 1)^\theta \right]^{1/\theta} \right)^{-1} \quad (7)$$

where θ is a copula parameter taking values in the range $[1, \infty[$. The corresponding values of Kendall's tau τ_K belong to $[\frac{1}{3}, 1]$.

Table 2 gives detailed information on the seasonal fits of the A12 copula family, based on the inversion of Kendall's tau τ_K . For more details on the fitting process and the goodness-of-fit evaluation, the reader is referred to Vandenberghe et al. (2010a). As stated before, seasons

are treated separately to comply with the assumption of non-stationarity, which is needed because of the underlying copula theory. The 95% confidence intervals (Table 2) of the estimated parameters for winter, spring and summer do not overlap, indicating considerable differences between the parameters. The interval for the autumn storms lies somewhat in between the interval of the winter and spring storms.

The performance of the test statistics S_n and T_n is studied by Genest et al. (2009) (therein denoted by $S_n^{(K)}$ and $T_n^{(K)}$). The S_n statistic is considered to be very powerful, especially in the case of Archimedean copulas, such as the A12 copula family. It should be noted that the reported p-values of the S_n statistics indicate that the null hypothesis, which states that the fitted A12 copula is an adequate model for the data, should be rejected for the winter and autumn

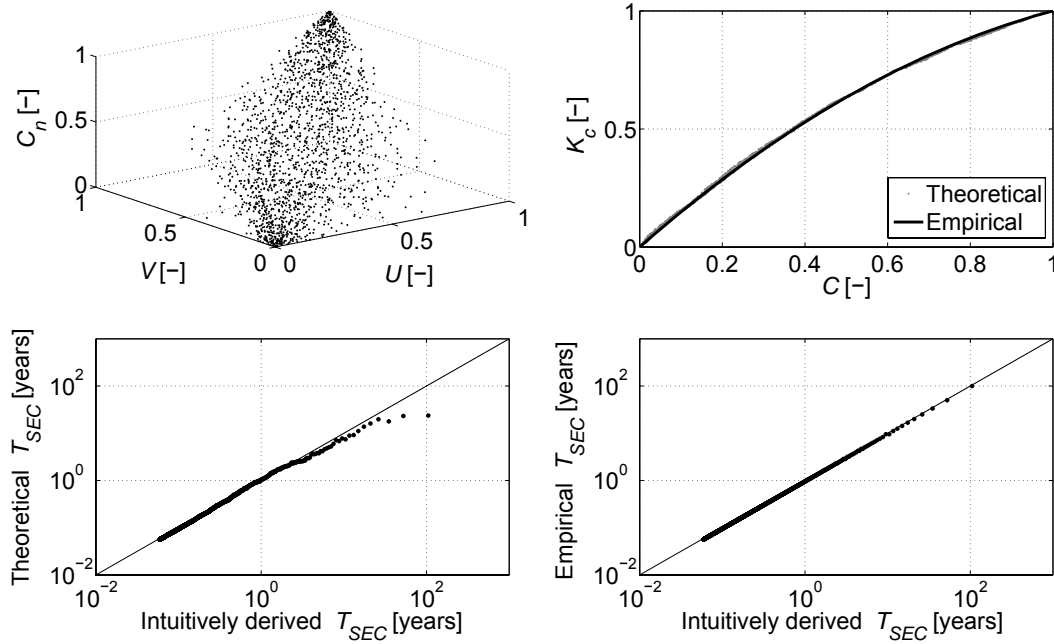


Fig. 3. Importance of a good fit for the interpretation of the secondary return period, considering the winter storms.

Table 2. Kendall’s tau τ_K and the estimated A12 copula parameter θ with 95% confidence bounds (lower bound LB and upper bound UB) and corresponding S_n and T_n goodness-of-fit measures with their p-values for the storms per season.

	Winter	Spring	Summer	Autumn
τ_K	0.6420	0.6068	0.5695	0.6252
θ	1.8622	1.6953	1.5485	1.7786
LB	1.7955	1.6309	1.4767	1.7036
UB	1.9289	1.7597	1.6202	1.8536
RMSE	0.0055	0.0043	0.0044	0.0064
S_n	0.1453	0.0842	0.0907	0.2719
$p(S_n)$	0.0019	0.0640	0.0648	0.0000
T_n	0.7834	0.7885	0.8989	1.0088
$p(T_n)$	0.0461	0.0543	0.0192	0.0019

storms when considering a significance level of 1% (i.e., the p-values are smaller than 0.01). For the spring and summer storms, this hypothesis cannot be rejected. Because of the large data sample the performed tests are relatively severe. Therefore, the p-values for the spring and summer storms are in fact very promising as they would become even larger when only subsamples of the data were to be considered. The fits of the winter and autumn storms are more doubtful, and their complications will be considered in Sect. 2.2.3. Nevertheless, it should be made clear that there is no need to restrict to the A12 copula family. Any practitioner should choose an adequate copula family for the data at hand. This adequacy should be evaluated by both visual and statistical goodness-of-fit evaluation methods. Because the

K_C function is important in the calculation of the secondary return period, a good correspondence between the empirical and theoretical K_C function is critical, and several methods exist for this evaluation (Vandenberghe et al., 2010a; Genest and Favre, 2007). For all Archimedean copula families, the calculation of the K_C function is fairly simple.

For the calculation of the mean storm interarrival time ω_T , one needs to consider that the original data set of 105 years is split up into four periods of each 26.25 years because of the seasonal division. For a specific season, ω_T is thus calculated as 26.25 years divided by the total number of storms in that season. The primary and secondary return period can then be calculated: ω_T (in years) divided by the product of the probability of occurrence of storm in a specific season and the respective probabilities as given in Fig. 1, which are easily calculated with the fitted copulas (see also Sect. 2.2.3).

It should be noted that the overall distribution function of (W, D) is considered here. It would also be interesting to consider e.g. annual maxima, which can be defined in several ways in a multidimensional setting (Kao and Govindaraju, 2007), in combination with extreme value copulas. In combination with marginal extreme value distributions of these annual maxima, a multivariate extreme value distribution would be obtained (Salvadori et al., 2007). Also, with an appropriate construction of the extreme value copula, the asymmetry of the data could be taken into account (Durante and Salvadori, 2009). This will be considered in future research.

2.2.3 Importance of a good fit

In this section, two different ways of calculating the secondary return period are matched with the intuitive interpretation of the secondary return period. On the one hand, T_{SEC} can be calculated theoretically with Eq. (4). For the A12 copula family, which is an Archimedean copula family, K_C can be calculated as follows:

$$K_C(t) = \frac{t(\theta + 1 - t)}{\theta} \quad (8)$$

Thus for each storm, the corresponding level t can be calculated, after which T_{SEC} can be obtained easily. On the other hand, T_{SEC} can also be calculated empirically. Therefore, the empirical copula C_n is considered, which can be defined as follows (Genest and Favre, 2007):

$$C_n(u, v) = \frac{1}{n} \sum_{k=1}^n \mathbf{I}\left(\frac{R_k^W}{n+1} \leq u, \frac{R_k^D}{n+1} \leq v\right) \quad (9)$$

where R_k^W and R_k^D are the ranks of W and D , n is the number of data points and $\mathbf{I}(A)$ is the indicator function of the set A . If for all n storms the corresponding empirical copula values $c_i = C_n(u_i, v_i)$, with $i=1, \dots, n$, are now treated as random values of the variable C_n , these values can be ranked and the corresponding empirical CDF can be calculated. This empirical CDF is in fact the empirical function K_{C_n} :

$$K_{C_n}(i) = \frac{R_i^{C_n}}{n+1} \quad (10)$$

where $R_i^{C_n}$ is the rank of the empirical copula value c_i .

The secondary return period can intuitively be derived as the mean interarrival time of super-critical storms. For a specific season, the empirical copula is constructed. The storm with the largest empirical copula value (the highest point in a 3-D-representation of the empirical copula) is thus the most extreme storm in the considered season out of a data set of 105 years. Subsequently, its intuitively derived return period is thus 105 years. For the second highest point in the empirical copula, the intuitively derived secondary return period of the corresponding storm is 52.5 (=105/2) years, and so on.

Figure 3 shows an evaluation considering the winter storms. Note that according to the goodness-of-fit test based on S_n in Sect. 2.2.2, the fitted A12 copula was rejected for the winter storms. The upper-left panel shows the empirical copula C_n . The upper-right panel shows a good visual correspondence between the theoretical function K_C and the empirical function K_{C_n} . The lower-left and lower-right panel illustrate the correspondence between the intuitively derived secondary return period in abscissa and, respectively, the theoretical and empirical secondary return period in ordinate, plotted in a double logarithmic scale. The full line corresponds with the 1:1 line. It is clear that the theoretical T_{SEC} underestimates the intuitively derived

return period, whereas the empirical T_{SEC} almost perfectly matches it. This points out that one should always be very cautious when interpreting theoretically calculated secondary return periods: very small shortcomings in the fitted copula (which may e.g. not accurately enough model the tail dependence) can already induce large deviations from what would intuitively be expected. In the case of summer storms, for which the fitted A12 copula was not rejected, a much better correspondence between the intuitively derived, the empirical and the theoretical return period exists (not shown here). More research on this topic is certainly recommended, however it is beyond the scope of setting the methodology for generating design storms as presented in this paper.

2.3 Marginal distribution functions

To be able to perform the transformation from U and V to W and D , the marginal cumulative distribution functions of W and D need to be known. For the further analysis in this paper, a kernel smoothed version of the empirical cumulative distribution functions will be used for this purpose (provided by the Matlab distribution fitting toolbox). As no extrapolations of W and D out of the available data range will be made throughout this paper, these non-parametrical CDFs provide sufficient information. When extrapolations would be needed, marginal distribution functions that are able to accurately model the tail behaviour of W and D should be used instead.

3 Analysis of Huff curves

Each storm in the time series is classified as a first, second, third or fourth quartile storm, based on the occurrence of the largest rainfall amount in, respectively the first [0, 0.25], second [0.25, 0.50], third [0.50, 0.75] or fourth [0.75, 1] quarter of the total storm duration (see also Sect. 2.1). If such classification cannot be made (e.g. a storm shorter than four 10-min intervals) or the maximum is observed in two or more quartiles, then the storm is removed from further analyses. The construction of Huff curves is then based on the distribution of normalized cumulative rainfall amounts in 5% time intervals of the normalized storm duration. Here, the 10%, 50% and 90% percentile curves are analyzed, considering storms of different quartile groups and with specific secondary return periods.

Figure 4 shows different Huff curves which are constructed considering all storms in a specific season and quartile group, regardless of their return period. The number of storms considered is given in Table 1.

3.1 Independence of return periods

In order to study the influence of the extremity of a storm on its internal storm structure, Huff curves can be constructed

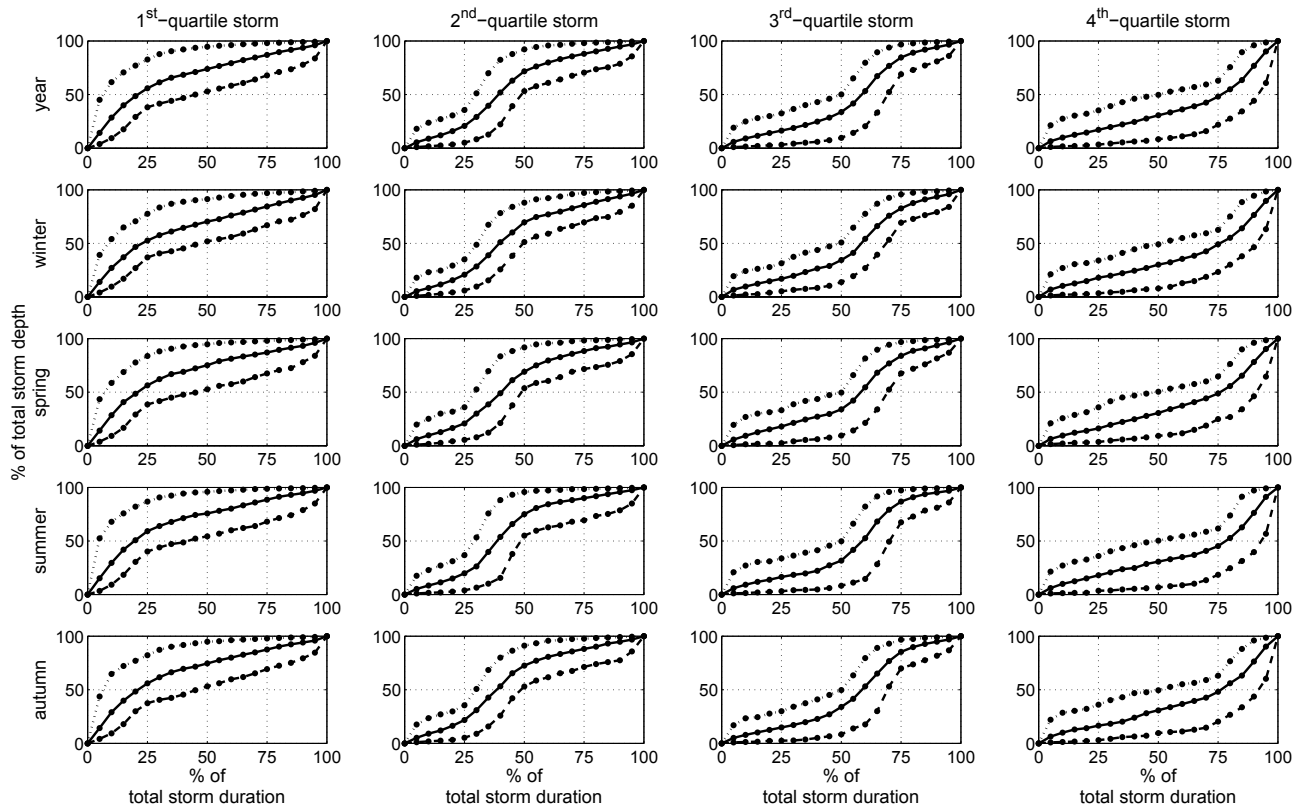


Fig. 4. Huff curves for different quartile storms and different seasons. The 10%, 50% and 90% percentile curves are given by, respectively the dashed, full and dotted lines.

Table 3. Statistical comparison of Huff curves for different thresholds of T_{SEC} (0.04, 0.08, 0.12, 0.16, 0.20 and 0.24 year) and different seasons. The p-values for the different non-parametrical Anderson-Darling 6-sample tests are given. Significant differences of Huff curves at a significance level of 1% are indicated in bold.

	1st quartile			2nd quartile			3rd quartile			4th quartile		
	Percentage of storm duration											
	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%
Year	0.00	0.00	0.00	0.00	0.00	0.48	0.68	0.00	0.48	0.34	0.00	0.00
Winter	0.00	0.02	0.74	0.07	0.83	0.97	0.83	0.15	0.87	0.89	0.30	0.03
Spring	0.00	0.00	0.93	0.68	0.55	0.93	0.98	0.03	0.93	0.97	0.11	0.22
Summer	0.00	0.00	0.07	0.10	0.63	0.90	0.91	0.85	0.97	0.92	0.34	0.27
Autumn	0.00	0.00	0.24	0.28	0.63	0.91	0.94	0.01	0.92	0.68	0.34	0.08

considering only storms having a return period larger than a specific threshold. To obtain reliable Huff curves a sufficient number of storms is needed. Therefore, these thresholds on the secondary return period should not be too large, as most secondary return periods are small (see Fig. 3). Figure 5 shows Huff curves for third quartile storms in summer, with thresholds on the secondary return period of 0.04, 0.08, 0.12, 0.16, 0.20 and 0.24 year. Note that these thresholds are not extreme, however, 68% of all 3rd quartile summer storms have a secondary return period smaller than 0.24 year (88

days). For these thresholds, respectively 299, 259, 185, 147, 117 and 97 storms are considered. As the percentile curves are very similar, this might indicate the independence of the internal storm structure and the extremity of a storm.

As Huff curves are constructed based on empirical distribution functions of the normalized cumulative storm depth at a specific percentage of the total storm duration, from which the 10%, 50% and 90% percentiles are calculated, a statistical test could be used to assess whether these distributions differ significantly when different

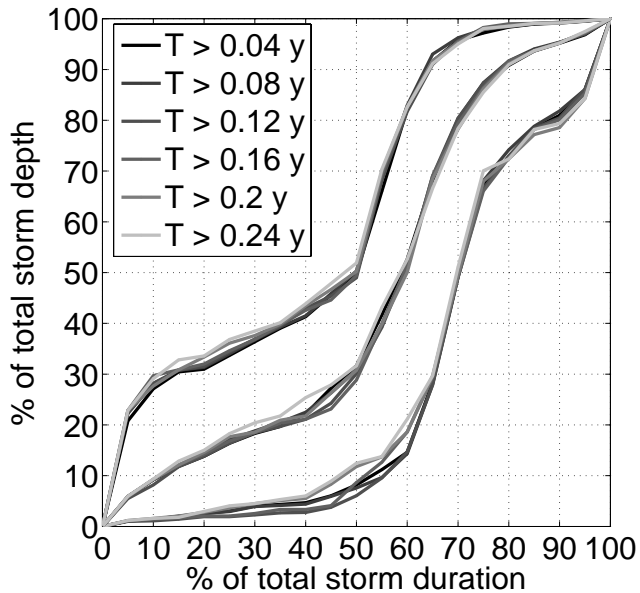


Fig. 5. Comparison of Huff curves, considering 3rd quartile summer storms for different thresholds on the secondary return period T .

thresholds of the secondary return period are considered. To assess whether the total Huff curves are identical, this statistical test can be performed at several percentages of the total storm duration. For example, for testing the significance of the differences between the six Huff curves considered in Fig. 5 obtained with six different thresholds of T_{SEC} , three different non-parametrical Anderson-Darling (AD) six-sample tests are considered to compare the empirical distribution functions of the normalized cumulative storm depth at 25%, 50% and 75% of the total storm duration. Figure 6 shows that per quartile group (columns) and at three specific moments in the storm (rows) six different cumulative distribution functions of the normalized cumulative storm depth (considering six thresholds for the return period) can always be compared and seem to be quite similar. The AD test is also able to account for ties, which is preferable as e.g. the most extreme storms are six times considered for the analysis of all six Huff curves. When the three resulting p-values are larger than the significance level of 1%, the null hypothesis of equal Huff curves is not rejected. These tests can then be performed for all configurations of seasons and quartile storms and the results are given in Table 3. The internal storm structure, expressed by Huff curves, of 2nd, 3rd and 4th quartile storms on the one hand are not influenced by the extremity of the storms in winter, spring, summer and autumn. On the other hand, 1st quartile storms seem to be influenced. When the storms are considered, regardless of their season, significant differences are present for all quartile storms.

It should be noted that the comparison as proposed here is somewhat subjective and could have a different outcome

using other thresholds of the return period (resulting in different sample sizes) and other percentages of the total storm duration at which the empirical distributions are compared (i.e., 25%, 50% and 75%). Nevertheless, the independence between the return period of a storm and its internal storm structure is assumed for the further analysis in this study in which storms will be treated seasonally.

3.2 Random generation of the internal storm structure

In order to obtain a random design storm generator, an algorithm for the random generation of a cumulative internal storm structure is developed. In first instance, the cumulative storm depths at the 25%, 50% and 75% of the total storm duration are randomly generated, i.e. the normalized cumulative storm depths in each quarter of the storm. This generation is constrained in three ways. Firstly, the cumulative storm depths are uniformly selected between the 10% and 90% percentile Huff curves. These bounds are subjective and can of course be altered by the user. The uniform selection assures that the whole range of possible internal storm structures will be covered. Secondly, the cumulative storm depths may not decrease in time. Thirdly, to assure that the design storm will respect the desired quartile, the maximum increase in cumulative storm depth should occur in that chosen quartile. Once the cumulative storm depths are chosen for each quartile, the rainfall in each quartile is further refined to each 5% time interval, based on the same Huff curves. Therefore, a random generator again uniformly selects cumulative rainfall depths that fall within the 10% to 90% percentile curves, assuring that the total preset cumulative rainfall depth during the chosen quartile, as determined before, is obtained. Again, the algorithm is forced to respect the non-decreasing nature of the cumulative rainfall depth in time. In other words, if the normalized cumulative storm depth in the preceding 5% time interval is higher than the lower bound given by the 10% percentile curve in the considered 5% time interval, then this preceding normalized cumulative storm depth forms the lower bound.

Figure 7 shows the outcome of the random generation of such a cumulative internal storm structure, together with the 10% and 90% percentile curves which serve as boundaries in the random generation.

4 Practical use of the random design storm generator

4.1 Simulation of one design storm

The developed algorithm for the generation of an internal storm structure and the concept of the copula-based secondary return period can now be used to generate design storms. Firstly, a secondary return period should be defined. With this return period the corresponding copula level t can be defined. Then, a random couple (u, v) can be selected having the predefined probability of occurrence, i.e. on the

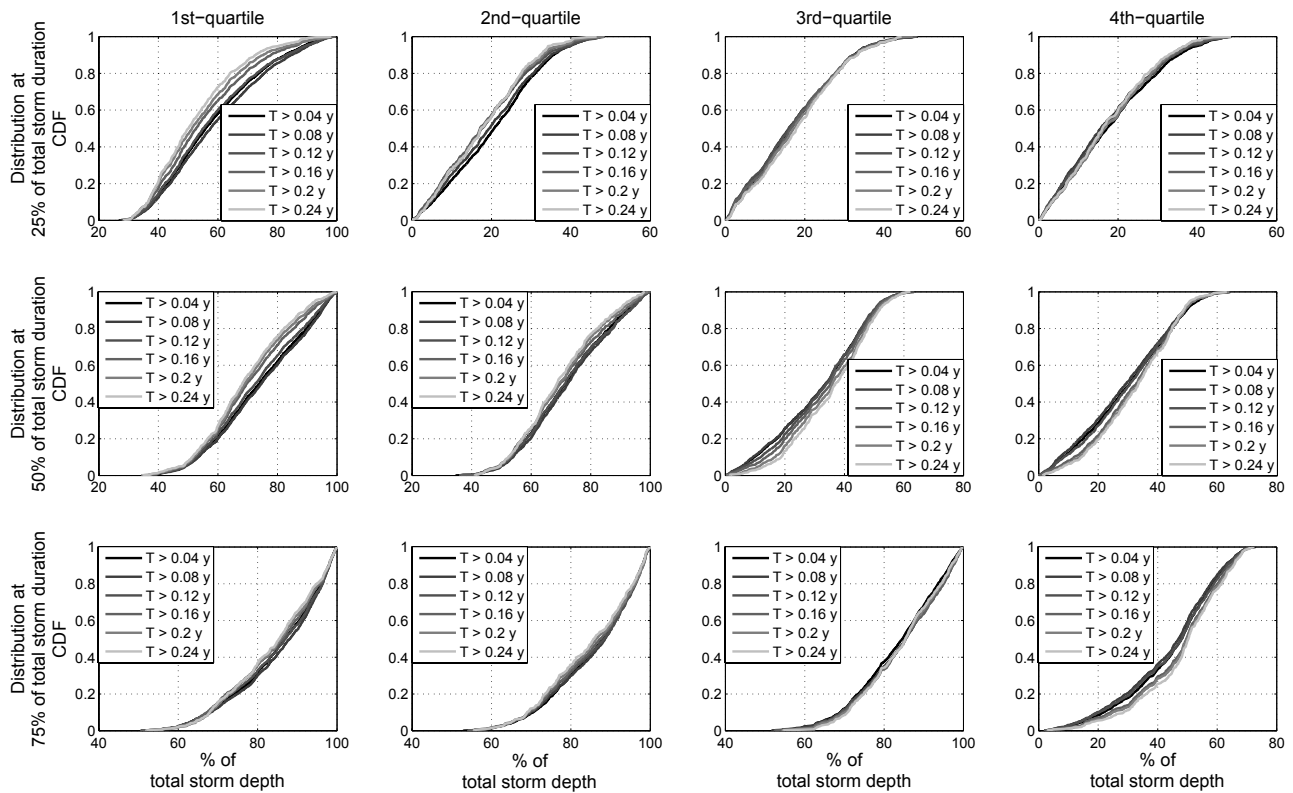


Fig. 6. Comparison of the empirical distribution functions (CDFs) of the percentage of total storm volume (normalized cumulative storm volume) at 25%, 50% and 75% of the total storm duration, considering summer storms for different thresholds on the secondary return period T .

t -contour line of the copula. With the use of the inverse of the marginal cumulative distribution functions, (u, v) is then transformed to values for, respectively the storm duration W and the total storm depth D . Then a random dimensionless internal storm structure is generated based on the Huff curves, which is superimposed on W and D .

For example, a 2nd quartile storm in winter with a secondary return period of 0.4 year corresponds to a copula level of 0.7618 and a random couple $(u, v) = (0.8439, 0.8047)$. Then, non-parametrical distribution functions are fitted to W and D considering all winter storms (see Sect. 2.3), regardless of their return period. By using the inverse CDFs, the couple (u, v) results in $W = 87.4$ h and $D = 18.6$ mm. Imposing a randomly generated dimensionless internal storm structure on these values results in the design storm given in Fig. 8.

4.2 Random generation of a set of design storms: a practical example

This section explains one possible way in which the stochastic design storm generator could be used in practice.

In hydraulic design studies, one often uses a hydrological model to generate a discharge input for a hydraulic model

based on a point rainfall series. Very often, only one extreme storm event is used because of computational reasons. This storm event is most likely a historical storm that caused problems within the river system. Through a simulation with a hydrological model, a discharge event corresponding to this historical storm can be obtained. However, there is no information on the uncertainty of this simulated discharge. When several statistically identical extreme storm events could be used, a set of hydrological discharges would be generated, yielding a distribution of the simulated discharges at a specific point of interest in the catchment under study. In this way, one could design a hydraulic structure in such way that it has only 1% of failure for storms with similar extreme statistics as the specific historical storm event that caused problems in the past.

The usefulness of the random generation of a set of statistically identical storm events is demonstrated here at the level of the hydrological model. The probability distributed moisture (PDM) model (Moore, 2007) is considered, as it is a rainfall-runoff model which is widely used in Flemish water management practices (Ferket et al., 2010; Cabus, 2008). The model is used with the calibrated parameters for the catchment of the Demer river (area: 96.64 km², Belgium). First of all, a specific storm is selected from

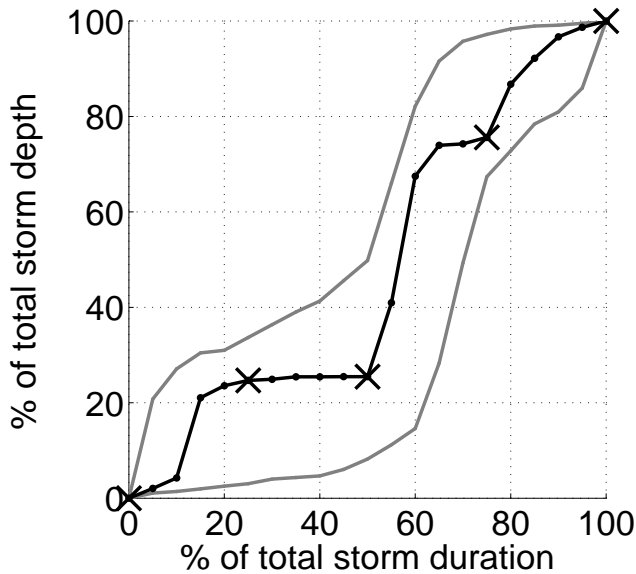


Fig. 7. A cumulative randomly generated storm structure of a third quartile storm in summer. For each 5% time interval, a percentage of the total storm depth is assigned in the range of the 10% and 90% percentiles. The grey lines form the 10% and 90% percentile curves of the Huff curves. The points marked by x are the randomly chosen cumulative storm depths at 25%, 50% and 75% of the storm duration. The black line is the random generation using the 5% interval data.

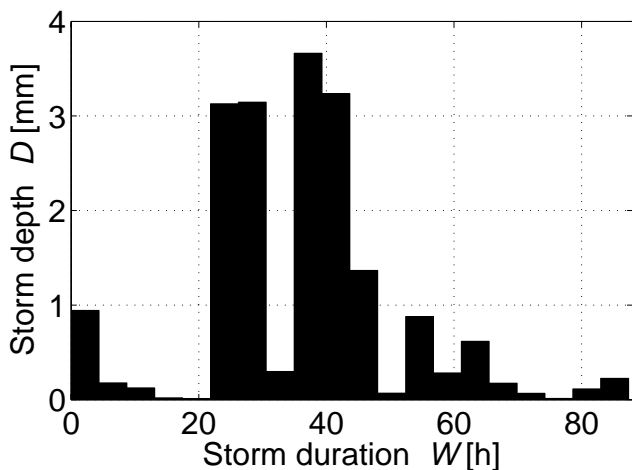


Fig. 8. A 2nd quartile storm in winter with a secondary return period of 0.4 year.

the historical data set. The storm starting on 6 July 1980, more specifically a 2nd quartile summer storm, is chosen and has a secondary return period of 26.96 year. Based on this storm, several statistically identical storms can be generated with the method as described above. We considered 10 000 randomly generated design storms. To be sure that the antecedent soil moisture conditions of the catchment are the

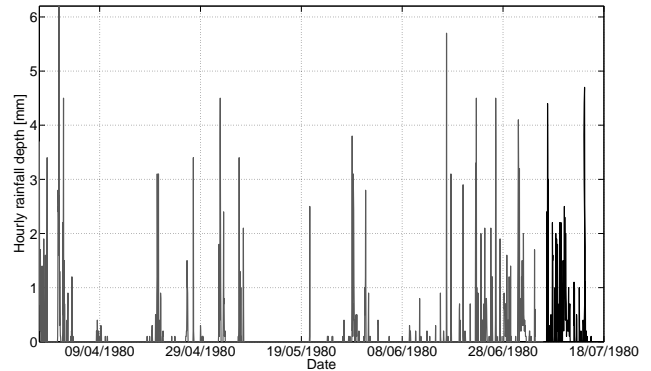


Fig. 9. The rainfall time series used in the rainfall-runoff modelling: the grey storms determine the antecedent soil moisture conditions and the black storm is the historical storm which is replaced by 10 000 randomly generated design storms.

same for each storm of the set, the same number of preceding historical storms, i.e. the preceding 100 days, is used in the rainfall-runoff simulations. Figure 9 shows the series of rainfall used.

Subsequently, this series forces the PDM model yielding a discharge series from which the last discharge event corresponding to the selected historical storm is selected (indicated in black in Fig. 9). For this discharge event, the maximum peak discharge Q_{max} is then calculated: $8.12 \text{ m}^3/\text{s}$. However, it would be interesting to know what the distribution of Q_{max} looks like, in order to obtain information on its uncertainty. Therefore, the PDM model is re-run 10 000 times with the rainfall series of Fig. 9 as input in which the historical storm considered is replaced by a randomly generated 2nd quartile summer storm having the same secondary return period. Again, the discharge event corresponding to the design storm is selected and Q_{max} is calculated.

Figure 10 shows the resulting distribution of Q_{max} , which gives an indication of the uncertainty of the discharge and could provide valuable information for a further hydraulic study. The peak discharge generated by the historical storm corresponds with a cumulative probability of 33%. This means that in 66% of the cases, a statistically identical design storm generates a larger peak discharge than what is observed historically. Furthermore, the distribution of Q_{max} has a non-Gaussian form. This results from the fact that, depending upon the properties of the catchment, the discharge does not behave linear with respect to the rainfall (see a discussion in Willems, 2000). Because of this non-linearity of the system, discharge will not have the same frequency of occurrence as the rainfall event. It should thus be noted that it would make no sense to perform a frequency analysis of the discharge and define a return period of a specific discharge event and then to generate storms with the same return period to use this information in a design study.

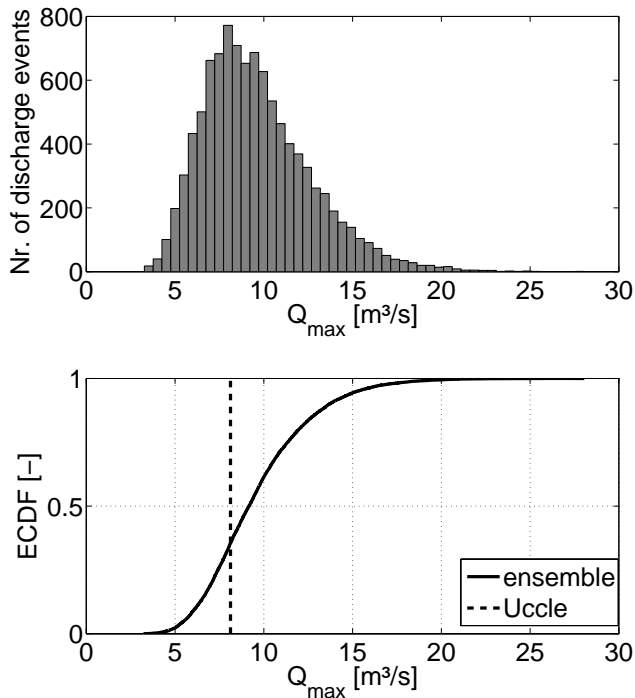


Fig. 10. The set of randomly generated design storms is able to give an idea of the distribution of the maximum peak discharge Q_{\max} . The maximum peak discharge for the historical 2nd-quartile summer storm on which the set is based clearly falls within the range of the distribution.

One way of validating the generator which is practically useful, is to evaluate whether the observed maximum discharge falls within the range of maximum discharges, generated by the set of design storms. Therefore the same analysis is conducted considering the forty most extreme observed summer storms one by one, having a secondary return period ranging from 97.2 to 3.1 years. For each observed storm, the position of the observed maximum discharge is evaluated with the range of maximum discharges. In other words, 40 times an evaluation as the one displayed in Figure 10 is made. To be sure that the generator does not have a bias, all observed discharges should clearly fall within the range of the generated set of maximum discharges, which is in fact the case for all 40 storm events. Also, the position of the observed maximum discharge within the range of the generated maximum discharges should vary uniformly. In other words, the empirical probability of having a maximum discharge (in the generated set) smaller than or equal to the observed maximum discharge should uniformly vary between 0 and 1. Figure 11 shows that the variation of this probability is in fact uniform.

The above only concerns the hydrological part of a design study. To come to an eventual design of hydraulic structures, 10 000 hydrological discharge time series should

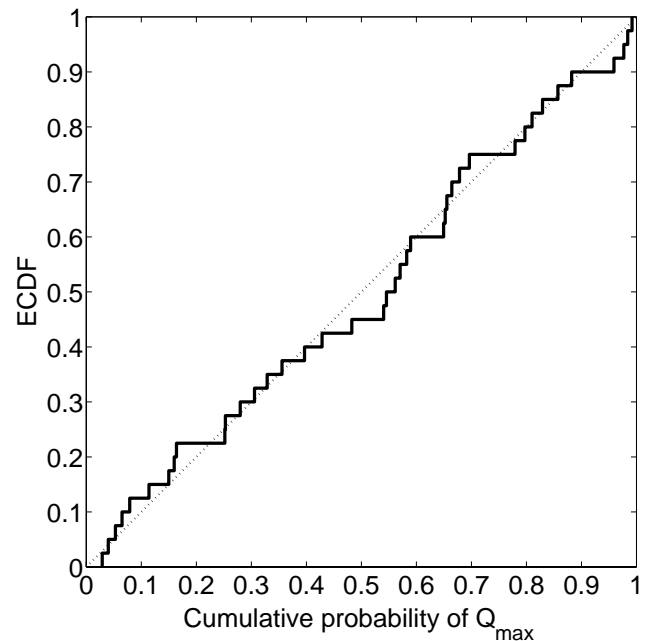


Fig. 11. The empirical cumulative distribution function (ECDF; full line) of the cumulative probability of Q_{\max} for the forty most extreme summer storm events corresponds with a uniform one (dotted line).

be routed through a hydraulic model. This allows then for an evaluation of the distribution of hydraulic discharges or water levels at a specific hydraulic structure. The latter could then be designed having a predefined probability of failure for a specific historical storm event.

5 Conclusions

This paper demonstrated that by combining the copula-based concept of a secondary return period together with a random internal storm structure generation based on Huff curves, a fairly simple stochastic design storm generator can be constructed. Its potential for obtaining information on the uncertainty of the maximum discharges in a rainfall-runoff modelling context has been illustrated in a comprehensible way.

The proposed generator is conceptually different from more traditional approaches. Firstly, it is storm based, i.e. storms are selected out of a time series based on an independence criterion. Traditional approaches mostly do not consider storms, but only consider “aggregation levels”, which are less interesting from a physical point of view. Secondly, a bivariate frequency analysis is conducted in which the dependence between storm depth and storm duration is explicitly incorporated. In traditional approaches, a univariate frequency analysis is mostly used and an Intensity-Duration-Frequency (IDF) curve is the main tool to link intensities, durations (in fact aggregation levels) and

return periods. However, this association between intensity, duration and return period is quite artificial as it does not address a real storm. Furthermore, it is not possible to find different combinations of intensity and duration that result in the same return period with this univariate approach.

However, some issues should be considered when applying the generator. First of all, although the internal structure of a storm is incorporated explicitly, it can be a rather average one. If a storm lasts very long (e.g. 60 h), then the generator as proposed here still ends up with a coarse internal storm structure, i.e. 3 h resolution (60 h divided by 20 different time intervals). However, the choice of the resolution of the Huff curves is user defined and greatly depends upon the resolution of the time series at hand. Additionally, the generator in its present form is not able to incorporate dry spells within a storm, which of course occur frequently in observed storms. By taking into account the percentage of dry spells in the total storm duration, this could be overcome. Of course, the existence of very long storms and the occurrence of dry spells is highly affected by the independence criterion, used for the selection of storms. A shorter criterion (e.g. 1 h instead of 24 h) would result in shorter storms, with less dry spells. The choice of this criterion should actually be inspired by the nature of the application, e.g. the concentration time of a watershed. Secondly, all generated storm structures are chosen uniformly between the 10% and 90% Huff curves. Ideally, one could take into account the real distribution of the cumulative storm depths at specific time intervals. The latter would be more useful when the storm generator should be used within the context of a point rainfall model. The choice of the bounds is also subjective, and is a decision to be made by the user. Thirdly, the performance of the generator relies largely on the skills of the practitioner with regard to copula modelling. It should be clear that there is no restriction to the use of the A12 copula family, but that a family should be chosen that provides the best goodness-of-fit (statistically and visually). The choice of a wrong copula family and bad fit can easily lead to an erroneous analysis of the uncertainty on extreme discharges.

Finally, it should be noted that this generator focuses on a point scale although the spatial aspect of rainfall is very often important as well. Several studies have already studied to some extent a copula-based spatial rainfall simulation and characterization (Bárdossy and Pegram, 2009; Serinaldi, 2009; AghaKouchak et al., 2010; Serinaldi, 2009; Villarini et al., 2008). Besides the incorporation of spatial rainfall characteristics, a design storm generator should ideally also address the antecedent soil moisture conditions, as these greatly influence the generated runoff in a catchment.

Acknowledgements. The comments of two anonymous referees and the issues raised in the interactive discussion are greatly acknowledged. The authors would further like to thank the Royal Meteorological Institute (RMI) for providing the extensive data set of Uccle rainfall. They also acknowledge Bruno Samain for the fruitful discussions concerning the PDM modelling exercise. The first author is a doctoral research fellow of the Research Foundation Flanders (FWO).

Edited by: S. Grimaldi

References

- AghaKouchak, A., Bárdossy, A., and Habib, E.: Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula, *Adv. Water Resour.*, 33, 624–634, doi:10.1016/j.advwatres.2010.02.010, 2010.
- 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.
- Bernardara, P., De Michele, C., and Rosso, R.: A simple model of rain in time: an alternating renewal process of wet and dry states with a fractional (non-Gaussian) rain intensity, *Atmos. Res.*, 84, 291–301, 2007.
- Bonta, J. V. and Rao, A. R.: Factors affecting the identification of independent storm events, *J. Hydrol.*, 98, 275–293, 1988.
- Cabus, P.: River flow prediction through rainfall-runoff modelling with a probability-distributed model (PDM) in Flanders, Belgium, *Agric. Water Manage.*, 95, 859–868, doi:10.1016/j.agwat.2008.02.013, 2008.
- Chow, V. T., Maidment, D. R., and Mays, L. W.: *Applied hydrology*, McGraw-Hill, 1988.
- De Jongh, I. L. M., Verhoest, N. E. C., and De Troch, F. P.: Analysis of a 105-year time series of precipitation at Uccle, Belgium, *Int. J. Climatol.*, 26, 2023–2039, doi:10.1002/joc.1352, 2006.
- Durante, F. and Salvadori, G.: On the construction of multivariate extreme value models via copulas, *Environmetrics*, 21, 143–161, doi:10.1002/env.988, 2009.
- Ellouze, M., Abida, H., and Safi, R.: A triangular model for the generation of synthetic hyetographs, *Hydrol. Sci. J.-J. Sci. Hydrol.*, 54, 287–299, 2009.
- Ferret, B. V. A., Samain, B., and Pauwels, V. R. N.: Internal validation of conceptual rainfall-runoff models using baseflow separation, *J. Hydrol.*, 381, 158–173, doi:10.1016/j.jhydrol.2009.11.038, 2010.
- Gargouri-Ellouze, E. and Chebchoub, A.: Modelling the dependence structure of rainfall depth and duration by Gumbel's copula, *Hydrol. Sci. J.-J. Sci. Hydrol.*, 53, 802–817, 2008.
- Genest, C. and Favre, A. C.: Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, 12, 347–368, 2007.
- Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, *Insurance: Mathematics and Economics*, 44, 199–213, doi:10.1016/j.insmatheco.2007.10.005, 2009.
- Grimaldi, S. and Serinaldi, F.: Design hyetograph analysis with 3-copula function, *Hydrol. Sci. J.-J. Sci. Hydrol.*, 51, 223–238, 2006.

- Huff, F. A.: Time distribution of rainfall in heavy storms, *Water Resour. Res.*, 3, 1007–1019, 1967.
- Kao, S. C. and Govindaraju, R. S.: A bivariate frequency analysis of extreme rainfall with implications for design, *J. Geophys. Res.*, 112, doi:10.1029/2007JD008522, 2007.
- Kao, S. C. and Govindaraju, R. S.: Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas, *Water Resour. Res.*, 44, doi:10.1029/2007WR006261, 2008.
- Kao, S.-C. and Govindaraju, R. S.: A copula-based joint deficit index for droughts, *J. Hydrol.*, 380, 121–134, doi:10.1016/j.jhydrol.2009.10.029, 2010.
- Koutsoyiannis, D.: A scaling model of a storm hyetograph, *Water Resour. Res.*, 29, 2345–2361, 1993.
- Koutsoyiannis, D.: A stochastic disaggregation method for design storm and flood synthesis, *J. Hydrol.*, 156, 193–225, 1994.
- Moore, R. J.: The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, 11, 483–499, 2007.
- Nelsen, R. B.: *An Introduction to Copulas*, Springer, 2006.
- Ntegeka, V. and Willems, P.: Trends and multidecadal oscillations in rainfall extremes, based on a more than 100-year time series of 10 min intensities at Uccle, Belgium, *Water Resour. Res.*, 44, 1–15, 2008.
- Pilgrim, D. H. and Cordery, I.: Rainfall temporal patterns for design floods, *Journal of the Hydraulics Division*, HY1, 81–95, 1975.
- Prodanovic, P. and Simonovic, S. P.: Generation of synthetic design storms for the Upper Thames River basin, *Assessment of Water Resources Risk and Vulnerability to Changing Climatic Conditions 5*, CFCAS Project, report, 20 pages, 2004.
- Restrepo-Posada, P. and Eagleson, P.: Identification of independent rainstorms, *J. Hydrol.*, 55, 303–319, 1982.
- Salvadori, G.: Bivariate return periods via 2-copulas, *Statistical Methodology*, 1, 129–144, 2004.
- Salvadori, G. and De Michele, C.: Frequency analysis via copulas: theoretical aspects and applications to hydrological events, *Water Resour. Res.*, 40, doi:10.1029/2004WR003133, 2004.
- Salvadori, G., De Michele, C., Kottegoda, N., and Rosso, R.: *Extremes in Nature, An Approach using Copulas*, Springer, 2007.
- Serinaldi, F.: Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective, *Stoch. Environ. Res. Risk Assess.*, 23, 677–693, doi:10.1007/s00477-008-0249-z, 2009.
- Serinaldi, F.: A multisite daily rainfall generator driven by bivariate copula-based mixed distributions, *J. Geophys. Res.-Atmos.*, 114, D10103, doi:10.1029/2008JD011258, 2009.
- Sklar, A.: Fonctions de répartition à n dimensions et leurs marges., *Publ. Inst. Statist. Univ. Paris*, 8, 229–231, 1959.
- Thompson, D. B., Asquith, W. H., Cleveland, T. G., and Fang, X.: Storm hyetographs for Texas, Literature review for TxDOT Project 0-4194, Texas Tech University, 2002.
- Vaes, G. and Berlamont, J.: Selection of appropriate short rainfall series for design of combined sewer systems, *Proceedings of First International Conference on Urban Drainage on Internet*, Hydroinform, Czech Republic, 2000.
- Vandenberghe, S., Verhoest, N. E. C., 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, 2010a.
- Vandenberghe, S., Verhoest, N. E. C., 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, submitted to *Water Resour. Res.*, 2010b.
- Veneziano, D.: Best linear unbiased design hyetograph, *Water Resour. Res.*, 35, 2725–2738, 1999.
- Villarini, G., Serinaldi, F., and Krajewski, W. F.: Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches, *Adv. Water Resour.*, 31, 1674–1686, doi:10.1016/j.advwatres.2008.08.002, 2008.
- Willems, P.: Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types., *J. Hydrol.*, 233, 189–205, 2000.