A 3-copula function application for design hyetograph analysis

SALVATORE GRIMALDI[1], FRANCESCO SERINALDI[2], FRANCESCO NAPOLITANO[2] & LUCIO UBERTINI[1,2]

1 *National Research Council, CNR-IRPI, Via Madonna Alta, I-126 06124 Perugia, Italy*
salvatore.grimaldi@irpi.cnr.it

2 *Department of Hydraulics, Transportation and Highways, University of Rome "La Sapienza", Via Eudossiana 18, I-00184 Rome, Italy*

Abstract A design hyetograph is a synthetic rainfall temporal pattern associated with a return period. Usually it is determined by means of statistical analysis of observed rainfall mean intensity through intensity–duration–frequency (IDF) curves. The other characteristics of a rainfall event, such as the peak, total depth and duration, are found indirectly throughout several phases of hydrological analysis, and trigger suitable work assumptions. The aim of this paper is to apply a multivariate approach in order to analyse jointly observed data of rainfall critical depth, maximum intensity and total depth. In particular, bivariate analysis of intensity–total depth conditioning to critical depth is developed using a 3-copula function to define the trivariate joint distribution function. Following the proposed procedure, once a design return period and related critical depth are selected, it is possible to determine, in a probabilistic way, the intensity and total depth, without advancing *a priori* hypotheses on the design hyetograph pattern. In a case study, the results obtained with the proposed procedure are compared with those deduced from standard design hyetographs usually applied in practical hydrological applications.

Key words copula; multivariate analysis; rainfall; synthetic hyetographs

INTRODUCTION

The event critical depth for a given duration and the design hyetograph are the two most important quantities of interest about rainfall for design and management purposes. The first is the maximum amount per unit area that can fall in a specified time interval with defined probability, and is given by intensity–frequency–duration (IDF) curves (Chow *et al.*, 1988). These curves are defined by a univariate statistical analysis and show the relation between mean critical intensity and duration, for an assigned value of return period. The second, the design hyetograph, is the temporal pattern of critical depth.

Approaches for defining design hyetographs can be synthesized into the following four categories (Chow *et al.*, 1988; Pilgrim & Cordery, 1975; Veneziano & Villani, 1999; Thompson, 2002):

1. definition of a simple geometric pattern relative to a single point on IDF curves, that is hyetographs whose depth is critical only for the chosen duration,
2. use of patterns in accordance with whole IDF curve,
3. use of standardized profiles obtained from rainfall records, and
4. use of patterns given by stochastic rainfall simulation.

In the following, examples of the four categories are described.

Type 1 The rectangular hyetograph, whose duration equals the basin concentration time and constant intensity equals a value achieved by IDF curves, belongs to this type. This hyetograph, very simple and for this reason widely used, systematically underestimates both rainfall intensity and depth (Veneziano & Villani, 1999). To overcome this limit, hyetographs with different profiles have been proposed. Among these, for instance, there is the so called Sifalda (1973) hyetograph, obtained from rainstorm records and characterized by a central rectangular part and two tails that take into account non critical amounts both previous to and after the critical one. The hyetograph suggested by Desbordes (1978) is similar to the above but it has a triangular profile in the central part. Finally, there is the simple, triangular-shaped hyetograph proposed by Yen & Chow (1980) and Chow *et al.* (1988).

Type 2 The most well-known examples of this category are the "alternating block" hyetograph and the so-called Chicago hyetograph introduced by Keifer & Chu (1957). In the first, rainfall intensity changes discontinuously whereas in the second it is continuous. As shown by Keifer & Chu (1957), the Chicago hyetograph is defined by considering critical intensities for every duration. Consequently the return period is longer, relative to the total depth, than that of the related IDF.

Type 3 This approach yields temporal patterns based on rainfall records. The precipitation event is adimensionalized by dividing time by total duration and cumulative depth by the total depth. Temporal patterns obtained through this approach differ from the previous methods, and yield complete events and not only the critical part. For instance, Huff's hyetograph (Chow *et al.*, 1988) is one of those derived by standardization. The advantage of these hyetographs is that they are related to rainstorm profiles observed in the area taken into account. Their limit is the high variability of observed patterns. In order to obtain stable profiles wide samples are needed.

Type 4 This approach reproduces the natural rainfall variability, simulating the event by stochastic models that allow one to overcome the limits of the previous methods. Research into this approach has been performed by Koutsoyiannis & Foufoula-Georgiou (1993). Finally, Zarris *et al.* (1998) proposed a mixed method that uses stochastic models in order to disaggregate the rainfall depth, deduced by IDF curves for a given duration and return time, at a shorter time scale. Using this technique one obtains a series of random independent storm patterns with the same duration and depth, that can be used as input to rainfall–runoff models.

Usually, hyetographs involve the following quantities: the *return period*, $T_r = 1/\left(1 - P\left(Z_{t=d}\right)\right)$ where $P(Z_{t=d})$ is the non exceeding probability of the critical depth for the critical duration d, suggested by design hypotheses; the *duration*, deduced by the basin concentration time, defined as "the time of flow from the (hydrologically) farthest point of the watershed to the outlet" (Chow *et al.*, 1988; Olivera & Stolpa, 2003), which is equivalent to the time when all the basin contributes to the runoff at the outlet; the *total rainfall depth*; the *maximum rainfall intensity* or *peak*; the *peak time*, that is the peak position in the rainfall pattern.

The maximum intensity and the peak position within the event are determined by both the analysis of record profiles (e.g. Huff, Pilgrim and Cordery hyetographs) and the mathematical properties of the pattern equations (e.g. Chicago hyetograph).

Keeping in mind how the IDF curves are built, it is useful to remember that the critical depth is not the total amount of an actual event, but the highest rainfall pulse (relative to the selected return period and critical duration) within a higher duration event, that usually is not critical for the other durations. As previously stressed, the patterns of the Sifalda, Desbordes and Chicago hyetographs try to consider a non critical depth of the storm, whose critical pulse is just a part, adding a rainfall contribution before and after the critical one. Then, there are two types of depth: a *critical depth* derived from IDF relationship and an *event total depth*. These quantities coincide in rectangular and triangular hyetographs, whereas, more realistically, they differ in the Sifalda, Desbordes and Chicago design storms.

So, it is possible to distinguish two types of quantities. Those related to the IDF relationship (duration, mean intensity and depth) usually so-called *critical*, and those (duration, peak intensity and depth) that characterize the actual storm pattern. As shown, the latter are derived from the former through either conceptual or empirical methods. However, with this approach, the design storms have the same conventional rarity degree of the IDF relationship employed.

This paper investigates the possibility of carrying out a multivariate statistical analysis of critical depth, maximum intensity and total depth related to the rainfall observations, in order to obtain the information necessary to derive the design storm directly from observed time series. For this purpose the *copula* function, a remarkably flexible tool for constructing multivariate cumulative distribution functions (cdfs) coupling arbitrary marginal (De Michele & Salvadori, 2003; Cherubini *et al.*, 2004; Favre *et al.*, 2004), is employed.

The proposed methodology is characterized by the use of the trivariate copula to define the joint cdf of total depth, maximum intensity and critical depth, the bivariate cdf of maximum intensity and total depth conditioned to critical depth, and the corresponding probability density functions (pdfs). The procedure is applied to a case study and the results are compared to those obtained applying the Sifalda and Triangular hyetographs.

THE 3-COPULA APPROACH

The aim of the paper is to derive event quantities useful for defining design hyetographs from critical ones probabilistically, instead of using empirical or conceptual methods as described in the Introduction.

In this work, critical duration and return period are regarded as design parameters defined *a priori*. The critical depth is the information directly achieved through the IDF relationship, whereas the mean intensity can be viewed as derived information because it is the ratio between the critical depth and critical duration. Overlooking duration and peak position, attention is focused on total depth (X, mm), peak intensity (Y, mm h^{-1}) and critical depth (Z, mm). Since these three quantities show several aspects of the same physical phenomenon they should be mutually correlated. As shown in the next section, the values of Kendall's τ rank correlation coefficient (Nelsen, 1999) confirms this hypothesis and shows that the dependence is usually positive and not to be overlooked. Thus, the behaviour of the three variables has to be studied jointly by a trivariate joint cdf $H(x,y,z)$ which contains all of the statistical information. $H(x,y,z)$ is built by means of a copula function, briefly described in "Appendix".

Once the return period and a critical duration are chosen, the total depth and maximum peak that occur with a given joint probability are determined, related to critical depth deduced from IDF curve. For this purpose the knowledge of joint cdf $H(x,y|z)$ and pdf $h(x,y|z)$ of total depth X and peak intensity Y conditioning to critical depth Z are needed. Then, using equation (A1) (see Appendix) their expressions can be written as follows:

$$\begin{aligned} H(x,y|z) &= P[X \le x, Y \le y | Z = z] = \frac{1}{f_Z(z)} \times \frac{\partial H(x,y,z)}{\partial z} \\ &= \frac{1}{f_Z(z)} \times \frac{\partial C(F_X(x), F_Y(y), F_Z(z))}{\partial F_Z(z)} \times \frac{dF_Z(z)}{dz} = \frac{\partial C(u,v,w)}{\partial w} = C(u,v|w) \end{aligned} \tag{1}$$

$$h\,(x,y|z) = \frac{\partial^2 H(x,y|z)}{\partial x \partial y} = \frac{h(x,y,z)}{f_Z(z)} = f_X(x) \times f_Y(y) \times C_{uvw}(u,v,w) \tag{2}$$

where $C_{uvw}(u,v,w) = \dfrac{\partial^3}{\partial u \partial v \partial w} C(u,v,w)$ is the copula density.

By fixing the z_0 value (obtained from IDF curves) and hence $w_0 = F_Z(z_0)$, the values of the total depth X and the peak Y, that occur with highest joint frequency relative to the chosen critical depth $Z = z_0$, are those that maximize function $h(x,y|z_0)$. In addition, the pairs total depth–peak with a given joint probability for a fixed critical depth can be defined. For a more complete description of the procedure (see Grimaldi & Serinaldi, 2005).

CASE STUDY

Copula and marginal cdfs assessment

The proposed procedure has been applied to half-hourly rainfall data from 1995 to 2001 from ten raingauges located in Umbria (Italy). The first step of the record analysis is to identify rainfall events in the time series. Following De Michele & Salvadori (2003), two events are judged independent if separated by a 7-h dry period (Koutsoyiannis & Foufoula-Georgiou, 1993). The heaviest events, in terms of critical depth, for each year, are selected giving seven annual maximum events relative to critical depth for every site. Generally, the ten events selected for each year belong to different periods of the year. So they seem not to be a consequence of the same contemporaneous weather phenomenon, and so they can be regarded, as a first approximation, as independent. On the other hand, since the sites show similar climate features, we can assume that the ten series belong to the same population, so a 70-year sample is available. Of course, these assumptions are approximate and imperfect but acceptable in order to test the model. So, for each event, peak (30-min maximum depth), total depth and critical depth related to the 2-h critical duration are defined. As shown by the τ correlation matrix (Table 1) the three quantities are positively correlated. This condition is necessary to apply the theory stated above.

The K-S test shows that critical depth, total depth and peak intensity have respectively Gumbel, log-normal and Gumbel distributions, whose parameters are illustrated in Table 2.

Table 1 Kendall's τ correlation matrix of critical depth, peak intensity and total depth.

Kendall's τ	Critical depth	Peak intensity	Total depth
Critical depth	1.000	0.504	0.492
Peak intensity	0.504	1.000	0.193
Total depth	0.492	0.193	1.000

Table 2 Parameters of marginal cdfs.

Marginal cdfs	Scale parameters	Position parameters
Gumbel (critical depth)	0.107	30.275
Lognormal (total depth)	0.435	3.840
Gumbel (peak intensity)	0.083	30.506

As explained in the Appendix, applying the CML method, the dependence parameters α of several 3-copula are assessed. Among these, the best copula is:

$$C(u,v,w)=\left(1+\frac{\left[(1+u)^{-\alpha}-1\right]\cdot\left[(1+v)^{-\alpha}-1\right]\cdot\left[(1+w)^{-\alpha}-1\right]}{\left(2^{-\alpha}-1\right)^{2}}\right)^{-1/\alpha}-1 \tag{3}$$

As previously described, the definitions of the conditioning joint pdf and cdf are needed. Applying equation (1) with the expression of copula (3) one obtains:

$$H(x,y\,|\,z)= \tag{4}$$

$$\left(1+\frac{\left((1+u)^{-\alpha}-1\right)\left((1+v)^{-\alpha}-1\right)\left((1+w)^{-\alpha}-1\right)}{\left(2^{-\alpha}-1\right)^{2}}\right)^{-\frac{1+\alpha}{\alpha}}\left(-\frac{1}{\alpha}\frac{\left((1+u)^{-\alpha}-1\right)\left((1+v)^{-\alpha}-1\right)(1+w)^{-\alpha-1}(-\alpha)}{\left(2^{-\alpha}-1\right)^{2}}\right)$$

The bivariate pdf $h(x,y|z)$ (whose expression is left out) conditioning to a critical depth value with a 5-year return period (P = 0.8, z_0 = 44.26 mm) is shown in Fig. 1(a). Figure 1(a) also shows a sample of 5000 pairs intensity–total depth conditioning to critical depth generated by the conditional copula method (Cherubini *et al.*, 2004; Favre *et al.*, 2004). Figure 1(b) shows a contour plot of $H(x,y|z)$. This function allows one to define all possible pairs of intensity–total depth conditioning to the value of critical depth chosen in this case study for every fixed conditional joint probability.

Comparison with standard design hyetograph methods

As described in the Introduction, a fixed critical depth relative to a chosen duration, the total volume and peak of the event are deduced by a synthetic hyetograph. The methods considered here are the Triangular and the Sifalda hyetographs. In the first one the temporal pattern of rainfall is a triangle with the peak equal to $2i_c$, with i_c critical intensity, and the duration equal to concentration time d. In the second one the pattern is characterized by a central rectangle and two triangles. The rectangle has intensity equal to i_c and duration equal to d, the first triangle has an intensity increasing from $0.0675i_c$ to $0.45i_c$ with duration equal to d and the second triangle has duration $2d$ and intensity decreasing from $0.45i_c$ to $0.09i_c$.

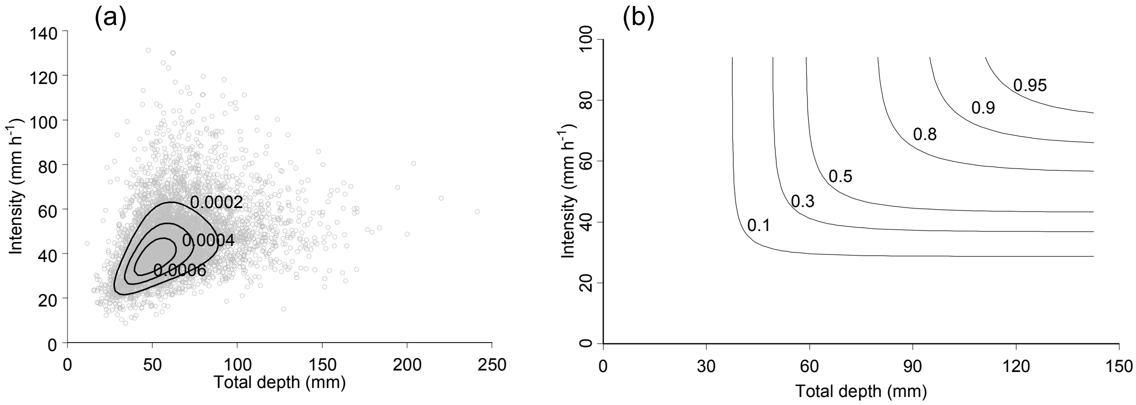

Fig. 1 (a) Contour plot of pdf $h(x,y|z)$ with *critical depth* z_0 = 44.26 mm (T_r = 5 years). Grey points are the 5000 pairs simulated with the model estimated. (b) Contour plot of cdf $H(x,y|z)$ with *critical depth* z_0 = 44.26 mm (T_r = 5 years).

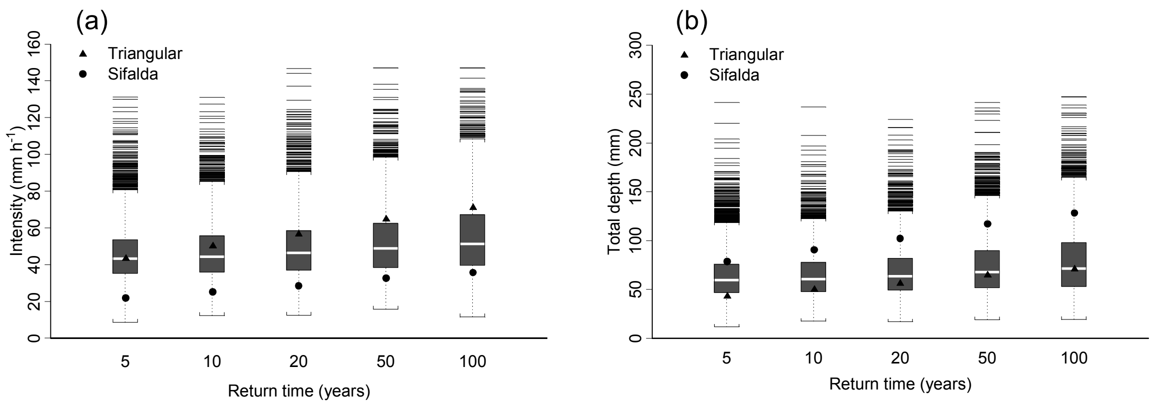

Fig. 2 Box-plot of: (a) intensity and (b) total depth, relative to the 5000 pairs generated considering five different return time values. The triangle and the circle represent respectively the pairs of Triangular and Sifalda hyetographs.

In Figs 2 and 3 the results, obtained using the standard approaches, are compared with one obtained using the copula method. In fact the availability of the distribution function conditioning to the critical depth and of a point generation algorithm allows us to create potential samples of volume–peak pairs. Therefore, 5000 pairs were simulated conditioned to the critical depth deduced from intensity–duration curves for duration equal to 2 h and for 5, 10, 20, 50, and 100-year return periods. The box-plots of volumes and peaks obtained are shown in Fig. 2(a) and (b). In the same graphs are overlapped the pairs obtained using the Triangular and Sifalda methods. As expected, the first method provides peak values higher than the second one and *vice versa* for the volume. However, both pairs of values are located near the boxes and far from the extreme values of both quantities.

Another way to judge the results is provided by Fig. 3(a,b,c). For three return time values (5, 20, 100-year), all points (generated, Sifalda and Triangular) are shown on a plot of intensity–volume. The position of the two standard methods is near to the centre of mass of the joint density. The value, probably low, of the joint return time of the two pairs, obtained from standard methods, is confirmed in Fig. 3(d,e,f). The two points are always lower than the 2-year joint return time.

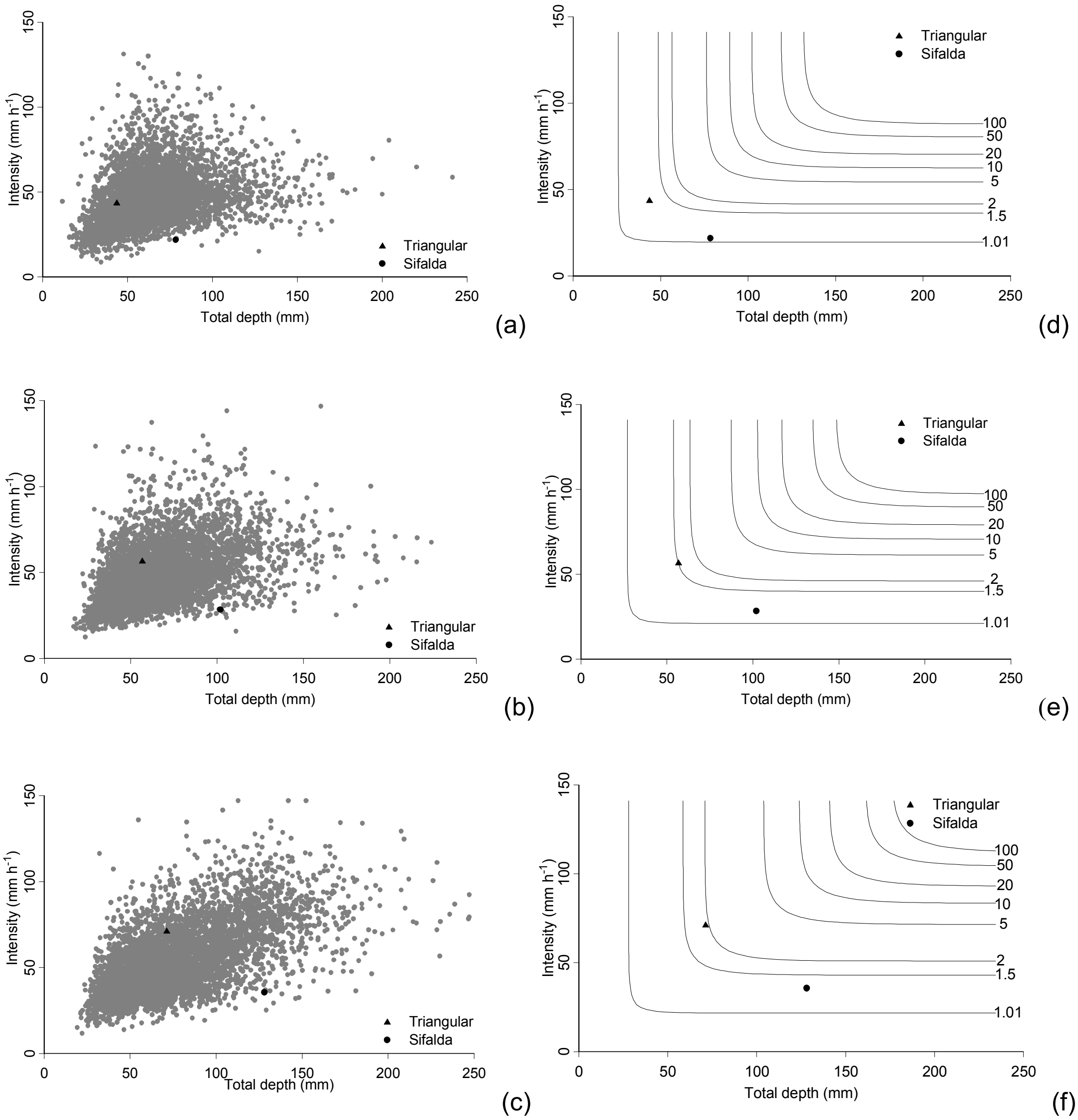

Fig. 3 (a, b, c) The 5000 generated pairs conditioned to three different critical depths: (a) 5-year return time, (b) 20-year return time, (c) 100-year return time. The triangle and the circle represent respectively the pairs of Triangular and Sifalda hyetograph. (d, e, f) Contour plot of joint return time derived from cdf $H(x,y|z)$ with *critical depth* z_0 relative to $T_r = 5$, 20, 100 years.

Therefore the comparison allows us to highlight that, for this case study, the values of peak and volume deduced from the critical depth via standard methods are not cautious because it could be common to have at the same time a pair with higher values of peak and volume.

CONCLUSIONS

In this paper a procedure to obtain information about the peak and total depth of design hyetographs is described. Using copula theory, the trivariate cdf critical depth–total depth–peak intensity has been determined. Knowing the joint probability it is possible to define bivariate pdf and cdf conditioning to critical depth. Choice of the best fitting copula has be carried out by comparing several 3-copulas. Comparing actual values and synthetic ones highlights the goodness of the estimated trivariate model. The joint cdf conditioned to critical depth allows one to generate pairs of intensity and total depth that show reasonable values. A comparison of the results obtained with the Triangular and the Sifalda synthetic hyetograph is shown. The standard method results provide peak–volume pairs with a low (1–2 year) joint return time.

APPENDIX: COPULA FUNCTION

A copula is a multivariate cdf defined in the unit cube $[0,1]^n$ with standard uniform marginals (for more mathematical details see Joe (1997) and Nelsen (1999)). A very important result in copula theory, named the Sklar theorem (Nelsen, 1999), assures, under some conditions, that it is possible to write a n-dimensional cdf and in particular a 3-dimensional cdf H with margins F_X, F_Y and F_Z as follows:

$$H(x, y, z) = C(F_X(x), F_Y(y), F_Z(z)) = C(u, v, w) \tag{A1}$$

where $F_X(x) = u$, $F_Y(y) = v$ and $F_Z(z) = w$.

In hydrological applications (De Michele & Salvadori, 2003; Favre *et al.*, 2004), copulas belonging to the symmetric Archimedean class are used, since they involve suitable properties (Nelsen, 1999). In these copulas, dependence among the variables is synthesized through an α parameter (e.g. see equation (3)).

Using the copula method, parameters of the marginal cdfs and the α parameter of the copula can be separately estimated. In particular, α can be assessed by the Canonical Maximum Likelihood (CML) method (Cherubini *et al.*, 2004). Usually, several Archimedean families are taken into account. After estimation of parameter α for each family, in order to choose which copula fits the data adequately, a test can be performed by using the cdf K_C of $t = C(u,v,w)$ (Barbe *et al.*, 1996). Besides the dependence structure of Archimedean copulas, C is uniquely determined by the function K_C (Genest & Rivest, 1993). Since $t \sim K_C$ and K_C is continuous, then the random variable $K_C \sim U(0,1)$. So, an estimated copula fits the sample when a *QQ*-plot K_C *vs* standard uniform quantiles has a roughly linear pattern. The goodness of fit can be evaluated by p-values related to a Kolmogorov-Smirnov test.

REFERENCES

Barbe, P., Genest, C. Ghoudi, K. & Rémillard, B. (1996) On Kendall's process. *J. Multivariate Analysis* **58**, 197–229.

Cherubini, U., Luciano, E. & Vecchiato, W. (2004) *Copulas Method in Finance*. John Wiley & Sons, Ltd, Chichester, UK.

Chow, V. T., Maidment, D. R. & Mays, L. W. (1988) *Applied Hydrology*. McGraw-Hill, New York, USA.

De Michele, C. & Salvadori, G. (2003) A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res*. **108**(D2), 4067, doi:10.1029/2002JD002534.

Desbordes, M. (1978) Urban runoff and design storm modeling. Paper presented at International Conference on Urban Storm Drainage, International Association for Hydraulic Research, Southampton, England.

Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N. & Bobée B. (2004) Multivariate hydrological frequency analysis using copulas. *Water Resour. Res*. **40**, W01101, doi:10.1029/2003WR002456.

Genest, C. & Rivest, L. (1993) Statistical inference procedures for bivariate Archimedean copulas. *J. Am. Statistical Assoc.* **88**, 1034–1043.

Grimaldi, S. & Serinaldi, F. (2005) Design hyetograph analysis with 3-copula function. Manuscript in review for *Hydrol. Sci. J.*

Joe, H. (1997) *Multivariate Models and Dependence Concept*. Chapman and Hall, New York, USA.

Keifer, C. J. & Chu, H. H. (1957) Synthetic storm pattern for drainage design. *J. Hydraulics Div*. **83**(1332), 1–25.

Koutsoyiannis, D. & Foufoula-Georgiou, E. (1993) A scaling model of a storm hyetograph. *Water Resour. Res*. **29**(7), 2345–2361.

Nelsen, R. B. (1999) *An Introduction to Copulas*. Lecture Notes in Statistics 139, Springer-Verlag, New York, USA.

Olivera, F. & Stolpa, D. (2003) *Effect of the Storm Hyetograph Duration and Shape on the Watershed Response*. Proceedings of the 82nd Annual Meeting of the Transportation Research Board (TRB). Washington DC, USA.

Pilgrim, D. H. & Cordery, I. (1975) Rainfall temporal patterns for design floods. *J. Hydraul. Div.* **101**, 81–95.

Sifalda, V. (1973) Development of a design rainfall for the calculation of channels. *GWF Wasser/Abwasser* **114**, 435–440.

Thompson, D. B. (2002) Storm hyetographs for Texas. Literature Review for TxDOT project 0-4194, 1–25. Texas Tech University, Lubbock, Texas, USA.

Veneziano, D. & Villani, P. (1999) Best linear unbiased design hyetograph. *Water Resour. Res*. **35**(9), 2725–2738.

Yen, B. C. & Chow, V. T. (1980) Design hyetograph for small drainage structures. *J. Hydraul. Div.* **106**, 1055–1076.

Zarris, D., Koutsoyiannis, D. & Karavokiros, G. (1998) A simple stochastic rainfall disaggregation scheme for urban drainage modelling. In: *Proceedings of the Fourth International Conference on Developments in Urban Drainage Modelling* (ed. by D. Butler & C. Maksimovic), 85–92. International Association of Water Quality, International Association of Hydraulic Research, UNESCO, and Imperial College, London, UK.