Hydrol. Earth Syst. Sci. Discuss., 8, 429–481, 2011
www.hydrol-earth-syst-sci-discuss.net/8/429/2011/
doi:10.5194/hessd-8-429-2011

Hydrology and Earth System Sciences Discussions

Modelling the statistical dependence of rainfall event variables by a trivariate copula function

M. Balistrocchi and B. Bacchi

Hydraulic Engineering Research Group, Department of Civil Engineering Architecture Land and Environment DICATA, University of Brescia, 25123 Brescia, Italy

Received: 18 December 2010 – Accepted: 26 December 2010 – Published: 18 January 2011

Correspondence to: M. Balistrocchi (matteo.balistrocchi@unibs.it)

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

Abstract

In many hydrological models, such as those derived by analytical probabilistic methods, the precipitation stochastic process is represented by means of individual storm random variables which are supposed to be independent of each other. However, several proposals were advanced to develop joint probability distributions able to account for the observed statistical dependence. The traditional technique of the multivariate statistics is nevertheless affected by several drawbacks, whose most evident issue is the unavoidable subordination of the dependence structure assessment to the marginal distribution fitting. Conversely, the copula approach can overcome this limitation, by splitting the problem in two distinct items. Furthermore, goodness-of-fit tests were recently made available and a significant improvement in the function selection reliability has been achieved. Herein a trivariate probability distribution of the rainfall event volume, the wet weather duration and the interevent time is proposed and verified by test statistics with regard to three long time series recorded in different Italian climates. The function was developed by applying a mixing technique to bivariate copulas, which were formerly obtained by analyzing the random variables in pairs. A unique probabilistic model seems to be suitable for representing the dependence structure, despite the sensitivity shown by the dependence parameters towards the threshold utilized in the procedure for extracting the independent events. The joint probability function was finally developed by adopting a Weibull model for the marginal distributions.

1 Introduction

The simplest stochastic models utilized for the representation of the precipitation point process usually employ individual event random variables, that consist in the rainfall volume, or the average intensity, the wet weather duration and the inter arrival time, whose probability functions have to fitted according to recorded time series (Beven, 2001). However, more complex formulations, as the cluster models (Cox and Isham, 1980; Foufoula-Georgiu and Guttorp, 1987; Waymire et al., 1984), equally require the calibration of some distributions of the event characteristics.

The first hydrological application of the event based statistics is due to Eagleson (1972), who derived the peak flow rate frequency starting from those featuring the average intensity and the storm duration, by assuming the two random variables independent and exponentially distributed. Since this seminal paper was published, these hypotheses have been assumed in a number of other works aimed at various purposes. Focusing only on those exploiting the derived distribution theory, probability functions were developed for hydrological dependent variables such as the runoff volume (Chan and Brass, 1979; Eagleson, 1978b), the annual precipitation (Eagleson, 1978a), the annual water yield (Eagleson, 1978c), the runoff peak discharge in urban catchments (Guo and Adams, 1998) and in natural watersheds (Córdova and Rodríguez-Iturbe, 1985; Díaz-Granados et al., 1984), the flood peak discharge routed by a detention reservoir (Guo and Adams, 1999) and the pollution load and the runoff volume associated with sewer overflows (Li and Adams, 2000).

Despite the remarkable results of the cited works, a significant association between the volume, or the intensity, and the duration usually arises from observed data statistic. Eagleson himself (1970) underlines the strong correlation that features the low intensities and the long durations, as well as the high depths and the short durations. So, the independence assumption should be essentially regarded as an operative hypothesis for simplifying the manipulation of the model equations and sometimes allowing the analytical integration of the derived distributions. Nevertheless, some authors (Adams and Papa, 2000; Seto, 1984), who compared analytical models derived by assuming both dependent and independent rainfall characteristics to continuous simulations, obtained better performances and more conservative results by neglecting the association among the random variables. The reason for this occurrence has not been clearly understood yet and it might be attributed to the improper selection of the joint probability models from which the derived distributions are obtained.

HESSD
8, 429–481, 2011

Furthermore, the exponential function does not always satisfactorily fit the sample distributions and distinct marginal probability functions may be needed for the three variables.

The development of multivariate probability functions, which account for these aspects by the traditional methods, represents nevertheless a very troublesome task. The most important constraint lies in the practical need to adopt marginal distributions belonging to the same probability family. The joint probability function must therefore be a direct multivariate generalization of these marginals, whose parameters also rule the estimation of the dependence ones. Examples of this can be found in Singh and Singh (1991), Bacchi et al. (1994), Kurothe et al. (1997) and Goel et al. (2000).

A remarkable advance in multivariate statistics has been however attained by means of copula functions (Nelsen, 2006), that represent an opportunity to remove these modelling drawbacks and enhance the accuracy of the rainfall probabilistic model. The approach allows to split the inference problem in two distinct issues: the dependence structure assessment and the marginal distribution fitting. Consequently, even complex marginal functions can be implemented and the estimation of their parameters does not affect the association analysis. In addition, a greater effectiveness in the model selection procedure has been achieved by the development and the validation of several goodness-of-fit tests.

In this work, three Italian rainfall time series recorded in locations subject to diverse precipitation regimes were examined to extract samples of individual storm characteristics: the rainfall volume, the wet weather duration and the interevent dry weather period. Hence, the dependence structure and the sensitivity with respect to the separation criteria were assessed. This problem was faced by analysing the involved random variables in pairs, in order to distinguish the different dependence strengths that can rule the associations among the three variables. The joint probability model was finally completed by assuming for the marginal distributions the more suitable Weibull probability functions. Test statistics were conducted to evaluate the goodness-of-fit of the proposed function to the observed samples, both for the copula functions and the marginal distributions.

2 Individual event variables

The preliminary elaboration of the continuous time series, which is required for making the precipitation data suitable for the statistical analysis, must lead to the identification of independent occurrences of the rainfall stochastic process. In the probabilistic framework such an issue is met by segregating the continuous record into individual events by applying two discretization thresholds: an interevent time definition (IETD) and a volume depth. The first one represents the minimum dry weather period needed to assume that two subsequent rain pulses are independent. Therefore, if two following hyetographs are detached by a dry weather period shorter than the IETD, they are aggregated into a unique event, whose duration is computed from the beginning of the first one and the end of the latter one, and whose total depth is given by the sum of the two contributes.

The second one corresponds to the minimum rainfall depth that must be exceeded in order to have an "appraisable" rainfall. When this condition is not satisfied, the event is suppressed and the corresponding wet period is assumed to be rainless and its duration is joined to the adjacent dry weather ones.

The choice of the threshold values strongly affects the statistical properties of the derived samples (Bacchi et al., 2008) and so it must be carried out very carefully. A lot of works were dedicated to such an argument (Restrepo-Posada and Eagleson, 1982; Bonta and Rao, 1988), exclusively focusing on the rainfall time series itself. For hydrological application purposes, a more effective criterion is to relate the procedure to the physical characteristics of the derived runoff process, rather than to the meteorological one (Balistrocchi et al., 2009). So, the IETD should be estimated, depending on the hydrologic-hydraulic system that is analysed, as the minimum time necessary to avoid the overlapping of the hydrographs generated by two subsequent storms. Furthermore, the volume threshold can be reasonably identified with the initial abstraction (IA) of the catchment hydrological losses (see for example Chow et al., 1988); thus, only the runoff producing rainfalls are taken into account. An individual rainfall event is therefore fully defined by simple random variables such as the total rainfall volume v and the wet weather duration t and associated with an antecedent dry period of duration a.

2.1 Precipitation time series

From the beginning of the sixties the operative department of the national hydrological service (SII) has been, as a matter of fact, dismissed. As a consequence, the collection of digital records, as long as needed for reliable statistical analyses, has been made very problematic. Despite this difficulty, a certain number of extended Italian rainfall time series were available for this study. However, in this paper only the results of those listed in Table 1 are presented. The reason lies in the matching behaviour that was shown by the precipitations belonging to the same meteorological regime. So, the three series can be considered as representative cases of their corresponding climates.

The series are constituted by sub-hourly observations, which were collected during periods longer than ten years by the raingauges of Brescia Pastori (Po Valley Alpine foothill), Parma University (northern Apennine foothill) and San Mauro Naples (Tyrrhenian coast of the Campania). Although they all belong to the Italian Peninsula, their climates are quite dissimilar.

In the traditional classification proposed by Bandini (1931), shown in Fig. 1, the typical Italian rainfall regimes vary from the continental one (with a summer maximum and a winter minimum) to the maritime one (a winter maximum and a summer minimum) and the majority of the territory exhibits an intermediate pluviometry between these two opposites.

The Brescia series is associated with the sub Alpine climate, which interests the northern portion of the Po Valley and the foothill areas of the tri-Veneto region. Parma lays instead in a transition region between the sub-coastal Alpine to sub Apennine climates, that distinguishes the southern part of the Po River catchment and the coastal area of the tri-Veneto. These intermediate regimes show two maximums and two minimums and the range between the extremes is moderate, since it does not exceed 100% of the monthly annual mean.

In the first case, the main maximum usually occurs in autumn and the secondary one in spring, while the main dry season is in winter; the second case is characterized by a main maximum in autumn and two equal size minimums in summer and in winter. On the contrary, the Naples series originates from the maritime regime, which features Sicily, Sardinia and a large part of the Ionic and Tyrrhenian coasts of Southern Italy. Under this boundary regime a single winter maximum and a summer minimum occur; so, the maximum range is sensibly greater than those typical of the intermediate climates and it can rise up to 200% of the monthly annual mean (Moisello, 1985).

The average annual rainfall depths that are assessed by using the selected time series evidence comparable values, since they are included between the 800 mm and the 1000 mm boundaries. Such a quantity actually exhibits a considerable spatial fluctuation in the Italian peninsula, because it rises over 3000 mm in the Carnic Alps but usually does not exceed 600 mm in the Sicilian inland. However, the estimates illustrated in Table 1 satisfactorily match those expected for the corresponding locations (Bandini, 1931).

2.2 Problem formalization

The dependence analysis carried out by way of the copula approach must rely on uniform random vectors (Joe, 1997), that are gathered from the original ones by means of the Probability Integral Transformation (PIT). In the set of equations written below, the quantities required for this study are defined:

$$\begin{cases} x = P_V(v) \\ y = P_T(t) \\ z = P_A(a) \end{cases} \quad \text{with} \quad x, y, z \in I = [0, 1], \tag{1}$$

being the P_V, P_T and P_A the cumulative distribution functions (cdfs) of the original variables v, t and a, at which correspond the x, y and z dimensionless variables.

Two main advantages arise by employing such transformations: (i) it is easy to prove that they have the same distribution and that this distribution is uniform, (ii) their population is constrained inside the unitary interval I=[0,1]. The joint distribution of the original random variables is linked to the copula function by the fundamental Sklar theorem (1959). In our trivariate case, if P_{VTA} is the joint probability function of the rainfall event variables having marginals P_V, P_T and P_A, it allows to state the equality

$$P_{VTA} = C_{XYZ}[P_V(v), P_T(t), P_A(a)]. \tag{2}$$

The function C_{XYZ}: $I^3 \rightarrow I$ is the copula that, in practice, constitutes the joint probability function of the uniform random variables and defines the dependence structure. The Sklar theorem ensures that such a function exists and, if the marginals are continuous, it is unique. So, the inference problem of the joint probability function P_{VTA} from the samples derived by the discretization procedure can be separated in the assessment of the copula C_{XYZ} and of the three univariate distributions P_V, P_T and P_A.

2.3 Pseudo-observation evaluation

When the sample data are considered, the cdfs in Eq. (1) must be approximated. That can be done by exploiting the plotting positions F_V, F_D and F_A. Herein, according to the standard Weibull formulation, they were expressed as a function of the ranks $R(.)$, associated with the random vector $\{\hat{v}_i, \hat{t}_i, \hat{a}_i\}$ of dimension n, as shown in Eq. (3) where $\hat{x}_i$, $\hat{y}_i$ and $\hat{z}_i$ are usually called pseudo-observations.

$$\begin{cases} \hat{x}_i = F_V = \frac{R(\hat{v}_i)}{n+1} \\ \hat{y}_i = F_T = \frac{R(\hat{t}_i)}{n+1} \quad \text{with} \quad i = 1, \cdots, n \\ \hat{z}_i = F_A = \frac{R(\hat{a}_i)}{n+1} \end{cases} \tag{3}$$

A proper estimate of such quantities is essential in the copula framework but is affected by the presence of ties, that could make the rank assignment uncertain. In the

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

context of the rainfall time series statistics, the occurrence of a repeated value essentially arises from the discrete nature of the records, from which the event variables are extracted. Indeed, it involves both the volume and the durations, because the rainfall is registered by using a constant time interval, as a multiple of the minimum depth detectable by the raingauge.

Therefore, if a preliminary data processing is not introduced, the sample is constituted by a huge number of minor events distinguished by the lower resolution values both for the volume and the duration (Vandenberghe et al., 2010). Indeed, the initial discretization required by the analytical-probabilistic approach acts positively on the ties, thanks to its aggregation and filter effects. Following this procedure, the single variables, especially the wet weather duration, may still present several repeated values. However, the occurrence of a rainfall having contemporaneously the same characteristics, that is the real concern in the copula assessment perspective, is very rare.

The performed statistical analyses revealed indeed that the suppression of the ties does not lead to an appreciable change in the measure of the dependence strength. The ranks in Eq. (3) were then calculated by applying the most common formulation accounting for all the repeated values, as indicated for the volume in the expression:

$$R(v_i) = \sum_{j=1}^{n} \mathbf{1}\left(v_j \leq v_i\right) \quad \text{with} \quad i = 1, \cdots, n; \tag{4}$$

in which the $\mathbf{1}(.)$ denotes the indicator function; the adaptations for the other variables are obvious.

3 Association measure analyses

The measures of the association degree which relate the uniform variables combined in pairs, $\{\hat{x}_i,\hat{y}_i\}$, $\{\hat{y}_i,\hat{z}_i\}$ and $\{\hat{x}_i,\hat{z}_i\}$, constitute the first step in the assessment of the overall

HESSD
8, 429–481, 2011

dependence structure. The use of rank correlation coefficients is very convenient inside the copula framework. The general avails of such measures indeed consist in their scale invariant property and non parametric nature. Furthermore, a more practical advantage is the possibility to easily relate them to the parameters of the most common copula functions because, in contrast to the usual Pearson linear correlation measure, they are purely a function of the dependence structure. So, although they represent necessary conditions to discriminate if a pair of random variables is independent or not, as well as the traditional Pearson correlation coefficient, they can significantly address the association function selection. In fact, close to zero values suggest the adoption of the independence copula, while some copula families only suit samples which show positive association.

The rank correlation coefficients herein utilized are those defined by Kendall and by Spearman. The Kendall coefficient τ_k (Kendall, 1938) estimates the difference between the probability of concordance and the probability of discordance for the pairs belonging to a bivariate random vector. The sample version $\hat{\tau}_k$ of the Kendall coefficient is written in the ratio

$$\hat{\tau}_k = \frac{c-d}{c+d}; \tag{5}$$

where c is the number of concordant data pairs, while d is the number of the discordant ones.

The Spearman coefficient ρ_s (Kruskal, 1958) also relies on the concordance concept, but its population version has a more complex interpretation. Nevertheless, from the copula point of view, it can be understood like a scalar measure of the average distance between the underlying copula, that rules the random process, and the independence one. Considering for example the bivariate random vector $\{\hat{x}_i, \hat{y}_i\}$, the estimate $\hat{\rho}_s$ of the Spearman coefficient is given by the relationship

$$\hat{\rho}_s = 1 - \frac{6\sum_{i=1}^{n}[R(x_i) - R(y_i)]^2}{n^3 - n}. \tag{6}$$

HESSD

8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

Following the scale invariant property of these association measures, the computation does not change if the natural data or the pseudo-observations are employed in the estimators Eqs. (5) and (6), because the PIT is based on a strictly increasing function.

The sensitivity of the rank correlation measures with respect to the independent event separation criteria was assessed by varying the two discretization thresholds within quite extended intervals, which were set in consideration of the potential hydrological applications. Hence, the minimum and the maximum IETD values were set in 3 h, corresponding to the time of concentration of a medium urban catchment, and in 96 h, that could be representative of larger drainage networks (Balistrocchi et al., 2009). The IA values were instead set between 1 mm and 10 mm, according to the catchment initial abstractions that can be reasonably assumed in urban and natural catchments, respectively.

3.1 Rainfall volume and wet weather duration pair

When the rainfall volume and the wet weather duration are coupled, the Kendall measure exhibits substantially similar trends in the three locations, as the continuous lines in the graphs of Fig. 2 show. The coefficient always assumes positive values from 0.25 to 0.55, evidencing quite a strong concordance between the two variables. Although this range is relatively narrow, they tend to moderately increase with the minimum interevent time but to decrease with the volume threshold. In fact, the weaker dependence occurs when the minimum IETD and the maximum IA are set, while the stronger one corresponds to the opposed circumstance.

The first behaviour can be understood by considering the aggregation effect of an increasing IETD on the rainfall pulses: the longer the minimum interevent time, the greater the wet weather durations and the rainfall depths which characterize the isolated storms are. In addition, the number of independent events decreases. As a consequence, the pseudo-observation variability is diminished towards a more concordant set of pairs. Such a mechanism is more evident for the lower IETD increments, because of the many storm pulses separated by small interevent periods existing in the

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |
HESSD
8, 429–481, 2011
Modelling the statistical dependence
M. Balistrocchi and B. Bacchi

series. Further increases of the higher values determine minor effects and the curves practically tend to assume a constant value.

The second one can only be explained by recognizing that the suppressed low depth events are mainly featured by short durations. Hence, those that have high depths but short durations, and form discordant pairs when they are coupled with the major storms, remain in the sample and decrease the whole dependence strength.

The correlation analysis performed by way of the Spearman concordance measure, whose trends are drawn in dashed lines in Fig. 2, leads to analogous outcomes. The curves of the ρ_s coefficient are shifted upward with respect to the Kendall ones and the values are enclosed in the interval 0.35 and 0.65. Thus, in every series the pseudo-observations seem to outline dependence structures that are quite different from those due to a couple of independent variables and the average distance between them is comparable in the three climates.

These results match the well known empirical evidence, because they confirm the existence of a non negligible statistical association relating the total volume and the duration of an individual storm, according to which the larger precipitation volumes are tendentially generated by the longer wet weather durations. Moreover, its strength demonstrates to be largely independent of the raingauge location. Therefore, the precipitation regime does not seem to represent a constituent factor in the assessment of this kind of dependence structure, unlike the discretization parameters whose selection determines an appreciable variation in the rank correlation estimate.

3.2 Wet weather duration and interevent period pair

The Kendall and Spearman coefficients estimated for the couple given by the wet weather duration and the interevent period led to the identification of a negligible association in every analysed climate. As the Kendall coefficient curves in graphs a) and b) in Fig. 3 show, the coefficient curves of the Brescia and the Parma time series oscillate around the null value in the very small interval of −0.10 and 0.10. The Naples data, graph c) in Fig. 3, are instead characterized by a negative association, which however

HESSD
8, 429–481, 2011

does not exceed the value of −0.15 for the largest IETD. Such a discordant dependence matches with the known behaviour of the mediterranean climate, in which short wet weather durations are separated by extended interevent times. The weakness of the relationship can be caused by the fact that this occurrence is only characteristic of certain periods, like the summer, when usually very long dry weather periods are interrupted by intense short storms.

In all three climates, the ρ_s coefficient brings to analogous considerations, because its curves are quite overlapped to the τ_k ones and substantially follow their trend. Despite the peculiarity of the Naples precipitation, the analysis outlines on the whole a random process due to the coupling of two independent variables, in which the absence of a detectable variability with respect to IETD and IA allows to state that their choice is largely trivial.

3.3 Rainfall volume and interevent period pair

When the rainfall volume and the interevent period are joined in pairs, insignificant association degrees are found for all of the series. The curves of the Kendall and Spearman coefficients, that are drawn in Fig. 4, reveal close to zero values for the majority of the discretization parameter combinations. Only when very high IETD and IA values are set, a small discordant dependence can be noticed in the Parma and in the Naples series. Such a finding does not seem to be relevant and can be explained by considering a huge decrease in the sample size that generates fictitious dependence strength enhancements.

4 Dependence structure assessment

The association analysis reveals a non homogeneous behaviour, in which only the volume-duration couple shows an appreciable dependence degree fairly affected by the IETD and IA values. On the contrary, the couples of the interevent period appear to be ruled by the independence copula, regardless of the setting of the event separation procedure. The development of a trivariate copula under such conditions therefore requires the use of a function able to suit this very asymmetric dependence structure.

One of the most helpful advantages of the copula approach is indeed the availability of a number of parametric copula families, whose members can be defined in any multivariate case. The majority of models that have found practical application is however fully defined by only one parameter, which is univocally determined by the dependence degree. In our situation the dependence is expected to be low when the three random variables are joined together. So, a one-parameter trivariate copula would be close to the independence one and the observed concordance between the event volume and the wet weather duration would not be correctly modelled. Moreover, in a multivariate framework having a dimension greater than two, the estimate of the dependence measure is controversial because its definition and estimation are ambiguous and complex (Salvadori et al., 2007).

A more convenient way is still given by the pairwise analysis, that can exploit various methods for constructing copulas of higher dimension by mixing together also different bivariate functions; examples are given in Grimaldi and Serinaldi (2006) and Salvadori and De Michele (2006). This technique appeared to be particularly appealing in this application and was therefore utilized in the following development: firstly one-parameter bivariate copulas were selected and fitted for the three random vectors $\{x_i, y_i\}$, $\{y_i, z_i\}$ and $\{x_i, z_i\}$, successively statistical tests were performed for assessing the goodness-of-fit and finally the trivariate function was constructed.

4.1 Copula function fitting

The selection of the most suitable copulas relies on the empirical copulas C_n, whose general expression is written as follows (Ruymgaart, 1973)

$$C_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{\mathbf{u}}_i \leq \mathbf{u}) \quad \text{with} \quad \hat{\mathbf{u}}_i \in I^p; \tag{7}$$

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

where $\mathbf{u}$ is a p-dimension random vector, $\hat{\mathbf{u}}_i$ are the corresponding pseudo-observations, $\mathbf{1}(.)$ denotes the indicator function and n is the sample dimension; analogous formulations can be given for the other two couples.

This function computes the frequency of the pseudo-observations which do not exceed the (x,y) value and tends to the underlying copula as well as, in the univariate case, the sample frequency distribution tends to the cumulative probability function. Being C_n a consistent estimator (Deheuvels, 1979), it represents the most objective tool for assessing the dependence structure and can be suitably utilized for facing the inference problem.

Herein, the copulas taken into consideration were those belonging to the one-parameter families of the Archimedean class, which includes absolutely continuous functions (a wide discussion of their properties is provided by Nelsen, 2006, pp. 106–156). The choice is justified by the advantages of their closed-form expressions and of their versatility, by which they are able to represent a variety of dependence structures both in terms of form and strength. For these reasons, the Archimedean copulas have already found application in many hydrologic problems.

In this situation, the inference problem is simplified in the estimation of the dependence parameter, denoted by θ, of the unknown copula C under the null hypothesis

$$H_0 : C \in \Gamma_0; \tag{8}$$

by which the function is a member of a certain parametric family Γ_0

$$\Gamma_0 = \{C_\theta : \theta \in \Sigma\} \quad \text{with} \quad \Sigma \subseteq R; \tag{9}$$

being Σ a subset of the real number R.

The fitting methods can be distinguished in parametric and semiparametric, in consideration of whether the hypotheses concerning the marginal distributions are involved or not. The full likelihood criterion and the inference function for margins (Joe, 1997) are parametric methods that employ the Sklar theorem for developing a maximum likelihood estimator where both the marginal and the copula parameters are included.

HESSD
8, 429–481, 2011

In the former, all the parameters are simultaneously estimated by maximizing the log-likelihood function. In the latter, the procedure is divided in two steps: initially only the marginal parameters are estimated, by using traditional consistent methods. The fitted margins are successively introduced into the log-likelihood function which is maximized to obtain those of the copula. The inference function for margins requires smaller computational burden than the full likelihood criterion, but usually determines an efficiency loss of the estimator. In addition, neither of them shields the dependence parameter estimation against making erroneous marginal assumptions and the uncertainty of their calibration. In such a condition, the methods have been proved to be affected by a severe bias (Kim et al., 2007).

One of the most popular semiparametric methods is based on the pseudo-likelihood estimator, that may be expressed as:

$$L(\theta) = \sum_{i=1}^{n} \log\left[c_\theta(\hat{\mathbf{u}}_i)\right], \tag{10}$$

where c_θ denotes the copula density (Genest et al., 1995; Shih and Louis, 1995). The $L(\theta)$ function accounts exclusively for the pseudo-observation samples, which substitute the marginal cdfs in the argument of the copula density.

Thus, $L(\theta)$ can be interpreted like a further development of the Joe inference function of margins, in which the marginal probabilities are non-parametrically estimated. In general this estimator is less computationally intensive than the previous ones but is not efficient, except for some particular cases (Genest and Werker, 2002).

Nevertheless, when the dependence parameter is a scalar quantity, the relationships between the copula function and the concordance measure can be exploited for fitting the copula to the data by the method of moments. In fact, both the theoretical versions of Kendall τ_k and Spearman ρ_s can be expressed in terms of the θ parameter. The fitting procedure therefore becomes very simple and easy and consists in inverting such relationships and substituting the estimates of the concordance measures gathered from the pseudo-observations.

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

In our view, despite their limits, the semiparametric methods represent more appropriate tools than the parametric ones, because they are consistent with the copula approach aim, which consists in maintaining the dependence structure assessment independent of those regarding the marginal distributions. Either of such methods can be applied in our case, thanks to the absolute continuity properties of the Archimedean models, that always ensures the existence of the copula density, and to the need of estimating only one dependence parameter.

Hence, considering the volume-duration pair, the most popular Archimedean copula were fitted to the various random vectors $\{\hat{x}_i, \hat{y}_i\}$ obtained by varying the IETD and the IA thresholds and visually compared to the corresponding empirical copulas. In the bivariate framework, this may be carried out by drawing the level curves of the two copulas providing a rough, but effective, evaluation of the global goodness-of-fit.

Among the examined families, the Frank copula (Frank, 1979) and, in particular, the Gumbel-Hougaard copula (Gumbel, 1960; Hougaard, 1986) gave the best fits. The bivariate member of this last family was therefore selected for modelling the $x-y$ distribution by defining the C_{XY} expression as:

$$C_{XY}(x,y) = \exp\left\{-\left[(-\ln x)^{\theta} + (-\ln y)^{\theta}\right]^{\frac{1}{\theta}}\right\} \quad \text{with} \quad x,y \in I. \tag{11}$$

In this copula the parameter θ must be greater than or equal to one and is univocally determined by the Kendall coefficient τ_k by means of the relationship:

$$\tau_k = \frac{\theta - 1}{\theta} \quad \text{with} \quad \theta \geq 1. \tag{12}$$

The Gumbel-Hougaard family is comprehensive of the independence copula, that is obtained when τ_K is zero and θ is equal to one, but is able to suit only positively associated data: the stronger the concordance, the higher the θ value is.

The contour plots reveal that the superiority of the Gumbel-Hougaard model with respect to the Frank one is due to its better performance in the upper-right corner of the unitary square I^2, where the larger and more severe occurrences are located.

HESSD
8, 429–481, 2011

Indeed, its upper tail behaviour is characterized by an asymptotic dependence lacking in the other and can be measured by the coefficient:

$$\lambda_u = 2 - 2^{1/\theta}. \tag{13}$$

The copula density concentrates both in the upper-right corner and in the lower-left one when the association degree increases, but only in the former the events appreciably align with the diagonal. This means that the higher extreme events show a deeper association than the commons, which emphasizes the tendency to generate joint values.

When the two calibration criteria were applied, the differences between the estimations of the dependence parameter amounted to few percentage points and no significant detriment or enhancement of the global fit arose by using the method of moment rather than the maximum likelihood criterion. Considering that the goodness-of-fit tests involve a very large number of estimations, the method of moments was therefore preferred to the maximum likelihood estimator in favour of the computational parsimony.

So, the estimations of the dependence parameter $\hat{\theta}$ were performed by inverting the Eq. (12) and substituting the sample version $\hat{\tau}_\mathrm{k}$ of the Kendall coefficient Eq. (5).

$$\hat{\theta} = \frac{1}{1 - \hat{\tau}_\mathrm{k}} \tag{14}$$

Its behaviour with respect to the discretization thresholds obviously agrees with the one previously discussed for the Kendall coefficient, except for the variability interval. If the same ranges are assumed for the IETD and IA values, it is bounded between 1.60 and 2.30; some estimated values are listed in Table 2 for the three series.

As a demonstration of the satisfactory goodness-of-fit achievable by using the Gumbel-Hougaard family, the level curves of the theoretical copulas are plotted in Fig. 5 against those derived from the empirical ones. The samples were derived from the continuous time series by assuming a IETD of 12 h and an IA of 2 mm.

Finally, the independence bivariate copula Π_2 was adopted to deal with the pairs of the interevent period. The bivariate functions C_{YZ} (Eq. 15) and C_{XZ} (Eq. 16) can be

HESSD
8, 429–481, 2011

defined when the random variable z is joined to the wet weather duration y and to the rainfall event volume x, respectively. This is a fundamental copula simply given by the product of the random variables and does not require any calibration.

$$C_{YZ}(y,z) = \Pi_2(y,z) = y\,z \quad \text{with} \quad y,z \in I \tag{15}$$

$$C_{XZ}(x,z) = \Pi_2(x,z) = x\,z \quad \text{with} \quad x,z \in I \tag{16}$$

The suitability of such copulas can be verified by observing the contour plots in Figs. 6 and 7, drawn for the samples extracted with the same threshold values, where the functions Eqs. (15) and (16) are compared to their empirical counterparts.

4.2 Copula goodness-of-fit tests

Although the contour plots showed a very satisfactory agreement between the theoretical and the empirical distributions, a quantitative estimation of the goodness-of-fit by test statistic was found to be needed to ensure the model reliability. The test is formally stated with regard to the null hypothesis H_0 (Eq. 8), under which the copulas were fitted. The objective is to verify whether the underlying unknown copula belongs to the chosen parametric family or such an assumption has to be rejected.

Several procedures have already been developed to meet this task, as observed by Genest et al. (2009) who provides a brief review of the most popular ones. The various methods are classified by the authors in three main categories (i) those that can be utilized only for a specific copula family (Shih, 1998; Malevergne and Sornette, 2003), (ii) those that have a general applicability but involve important subjective choices for their implementation, like a parameter (Wang and Wells, 2000), a smoothing factor (Scaillet, 2007) or a data categorization (Junker and May, 2005), (iii) those that do not have any of the previous constraints and for this reason are called blanket tests. The convenience of adopting the last kind of test is clear and was highlighted by the authors.

HESSD
8, 429–481, 2011

Furthermore, in the same paper, they analyzed the performances of some blanket procedures by way of large scale Montecarlo simulations, obtaining a general confirmation of their validity. One of the most powerful tests is based on the empirical copula process χ_n defined as:

$$\chi_n = \sqrt{n}\left(C_n - C_{\hat{\theta}}\right), \tag{17}$$

which evaluates the distance between the empirical copula C_n and the estimate $C_{\hat{\theta}}$ of the underlying copula C under the null hypothesis H_0.

A suitable test statistic S_n can be constructed by using a rank-based version of the Cramer-Von Mises criterion, as shown in the integral:

$$S_n = \int_{[0,1]^p} \chi_n(\mathbf{u})^2 \,\mathrm{d}\chi_n(\mathbf{u}) \quad \text{with} \quad \mathbf{u} \in I^p, \tag{18}$$

whose integration variable $\mathbf{u}$ is a generic uniform random vector having p dimension. When the S_n values are low, the fitted model and the pseudo-observation distribution are close, on the contrary, they disagree considerably. In the first condition the null hypothesis H_0 tends to be not-rejected while in the other to be rejected.

Dealing with the pseudo-observations $\hat{\mathbf{u}}_i$, the statistic S_n may be approximated by the sum:

$$S_n = \sum_{i=1}^{n} \left[C_n(\hat{\mathbf{u}}_i) - C_{\hat{\theta}}(\hat{\mathbf{u}}_i)\right]^2. \tag{19}$$

As previously argued by Fermanian (2005) and successively demonstrated by Genest and Rémilland (2008), the statistic S_n is actually able to yield an approximate p-value if it is implemented inside an appropriate parametric bootstrap procedure. According to the achievements previously discussed, in our case such a procedure was implemented as follows for each of the three pairs of random variables:

HESSD
8, 429–481, 2011

M. Balistrocchi and B. Bacchi

- For a given combination of the parameter IETD and IA, n bivariate vectors of pseudo-observations are derived by using the discretization procedure described in Sect. 2, being n the total number of independent storms.
- The sample Kendall coefficient $\hat{\tau}_k$ and the empirical copula C_n are computed according to the observed data by their expressions Eqs. (5) and (7).
- The dependence parameter $\hat{\theta}$ is estimated by the method of moments, assessing the underlying copula $C_{\hat{\theta}}$.
- The Cramer-Von Mises statistic S_n is directly calculated by Eq. (19), in which the analytical formulation of the Archimedean models is exploited.
- For a large integer N, the next sub-steps are repeated for every $m=1, \ldots, N$:
  - A sample of pseudo-observations of n dimension is generated by simulating the estimation of the underlying copula $C_{\hat{\theta}}$.
  - The sample Kendall coefficient $\hat{\tau}_{k,m}$ and the empirical copula $C_{n,m}$ of the simulated data are computed according by their expressions Eqs. (5) and (7).
  - The dependence parameter $\hat{\theta}_m$ is estimated by the method of moments, assessing the theoretical copula $C_{\hat{\theta}_m}$.
  - The Cramer-Von Mises statistic $S_{n,m}$ of the simulated sample is calculated by the same expression (Eq. 19), in which $C_{n,m}$ and $C_{\hat{\theta}_m}$ are substituted to C_n and $C_{\hat{\theta}}$, respectively.
- An approximate p-value is finally provided by the sum:

$$\frac{1}{N}\sum_{m=1}^{N}\mathbf{1}\left(S_{n,m} > S_n\right). \tag{20}$$

HESSD
8, 429–481, 2011

The tests were conducted varying the discretization parameters within the same IETD and IA ranges previously used for the association measure analysis. Firstly, the real existence of a dependence degree was verified by assuming that all the three bivariate distributions correspond to the independent 2-copula (Eq. 21). When this assumption is tested the procedure simplifies, because the steps concerning the estimation of the dependence parameter do not apply.

$$H_0 : C = \Pi_2 \tag{21}$$

The p-values listed in Tables 3–5 were obtained for the three couples by assuming N greater than 2500. On the whole, the independence assumption must be rejected for the event volume and the wet weather duration pair, because null or close to zero p-values are assessed, while for the other two variable joints it cannot be rejected with significance levels considerably greater than the usual levels of the 5–10%. In addition, no particular trends with regard to the IETD and IA settings or differences among the climates are shown.

Hence, only for the first couple the null hypothesis (Eq. 22), by which the dependence structure can be modeled by the Gumbel-Hougaard 2-copula, was examined. The further goodness-of-fit tests led to the estimations that are reported in Table 6.

$$H_0 : C_{XY} \in \left\{ \exp\left\{ -\left[(-\ln x)^{\theta} + (-\ln y)^{\theta} \right]^{\frac{1}{\theta}} \right\} : \theta \in [1, \infty[\right\} \tag{22}$$

The p-values mainly range between 40% and 100%, demonstrating that the selected family is able to suit the empirical data ensuring very high significance levels for all the IETD and IA choices and the precipitation regimes. Nevertheless, the goodness-of-fit shows a moderate tendency to improve according to the increments of both the discretization parameters. The Frank model was tested as well, but in nearly every case poorer adaptations were evidenced by p-values lower than the Gumbel-Hougaard ones (50–60%). The only exceptions to this situation were noticed in the Brescia rainfall series for the minor thresholds, for which the p-values of the two models are comparable as well.

HESSD
8, 429–481, 2011

The p-values exactly equal to the unity could certainty seem anomalous. The occurrence may be explained by the limited number N of repeated simulations that were performed, which was not set substantially larger than the sample dimension n as recommended by Genest et al. (2009) for limiting the estimation variance. This choice was due to the need to balance the assessment accuracy and the heavy computational demand of the procedure, in view of the sample sizes which can greatly exceed two thousand when the lower threshold values are utilized. An iteration number of 2500 has been considered by the same authors as an acceptable compromise, that still allows to meet the main objective of the test, even if a certain loss of the estimator efficiency has to be accounted for. In our application, the assessed p-values do not demonstrate any relevant variation when the procedure is carried out by repeating over 2000 simulations.

Thus, it can be concluded that the goodness-of-fit tests based on the Cramer-Von Mises statistic S_n definitely confirms the broad framework that has been delineated by the association measure analysis and by the graphical fits, since the assumed bivariate models cannot be rejected for the standard significance levels commonly adopted in the statistical tests.

4.3 Mixing method

Among the quite large set of strategies for constructing multivariate copulas, the conditional approach that was investigated by Chakak and Kueler (1995) seemed to be extremely interesting for our problem. In fact, it is possible to demonstrate that the trivariate copula C_{XYZ} must satisfy the equality

$$C_{XYZ}(x,y,z) = C_{XY}\left(\frac{C_{XZ}(x,z)}{z}, \frac{C_{YZ}(y,z)}{z}\right) z \tag{23}$$

which mixes together the lower order copulas.

An important simplification is determined in such an equation by the independence of the z random variable with respect to the others. If the copulas Eqs. (15) and (16) are substituted, the equation of the trivariate copula is easily reduced to the product of the bivariate copula C_{XY} and the uniform variable z.

$$C_{XYZ}(x,y,z) = C_{XY}\left(\frac{\Pi_2(x,z)}{z}, \frac{\Pi_2(y,z)}{z}\right) z = C_{XY}\left(\frac{xz}{z}, \frac{yz}{z}\right) z = C_{XY}(x,y)\, z \tag{24}$$

So, the joint distribution suiting the natural variability of the individual rainfall event variables may be expressed by the trivariate copula (Eq. 25), whose unique parameter θ is exclusively a function of the positive association between the event volumes and the wet weather durations.

$$C_{XYZ}(x,y,z) = \exp\left\{-\left[(-\ln x)^{\theta} + (-\ln y)^{\theta}\right]^{\frac{1}{\theta}}\right\} z \tag{25}$$

5 Marginal distribution assessment

Several probability models have been suggested for describing the natural variability of the rainfall event characteristics, including the Gamma (Di Toro and Small, 1979; Wood and Hebson, 1986), the Pareto (Salvadori and De Michele, 2006) and the Poisson functions (Wanielista and Yousef, 1992). Nevertheless, the most popular one is indeed the exponential model, that has been extensively employed in a huge number of problems (a detailed list is provided in Salvadori and De Michele, 2007). The reason for such a success mainly lies in its very simple formula, that often allows the possibility to analytically integrate the derived probability functions and so makes it particularly appealing in the analytical-probabilistic perspective.

Unfortunately, in the Italian precipitation regimes the exponential model does not suit the observed distributions of the rainfall event variables. An appropriate alternative has been detected in the Weibull model (Bacchi et al., 2008), that can be viewed like a more versatile generalization of the exponential one. Hence, the theoretical marginals P_V, P_T and P_A defined in Eq. (1) have been represented by means of the three cdfs Eqs. (26),

HESSD
8, 429–481, 2011

(27) and (28), respectively.

$$P_V(v) = \begin{cases} 1 - \exp\left[-\left(\frac{v - \mathrm{IA}}{\zeta}\right)^\beta\right] & \text{for } v \geq \mathrm{IA} \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

$$P_T(t) = \begin{cases} 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^\gamma\right] & \text{for } t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

$$P_A(a) = \begin{cases} 1 - \exp\left[-\left(\frac{a - \mathrm{IETD}}{\psi}\right)^\delta\right] & \text{for } a \geq \mathrm{IETD} \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

The exponents β, γ and δ are dimensionless parameters that specify the shape of the distributions, while the denominators ζ (mm), λ (h) and ψ (h) define the scale of the random process. The IA and the IETD parameters play the role of lower limits in Eqs. (26) and (28) and are set according to the minimum interevent time and the volume threshold utilized in the rainfall separation procedure.

In this kind of model the correct assessment of the exponent is a key aspect, because very dissimilar shapes of the probability density function (pdf) are possible according to its value: (i) when an exponent minor than one occurs, the pdf monotonically decreases from the lower limit in which a vertical asymptote is present, (ii) when it is exactly equal to one, the pdf owns a finite mode in the lower limit but does not lose the monotonic decreasing behaviour, (iii) when it exceeds the unity the distribution mode is greater than the lower limit and the pdf shows a right tail that is progressively less marked if it is further incremented. Hence, the greater the exponent is, the more shifted to the right the most probable values of the random variable are and the more symmetric the distribution is.

HESSD
8, 429–481, 2011

5.1 Marginal function fitting

A sensitivity analysis has been carried out referring to the previously investigated IETD and IA ranges. The marginal distributions were fitted to the sample data extracted from the continuous rainfall series by the maximum likelihood criterion yielding the graphs drawn in Fig. 8 for the shape parameters and in Fig. 9 for the scale ones: on the whole, it can be argued that the precipitation regime does not significantly affect the estimation of the distribution parameters, since almost identical behaviours are delineated.

The shape parameters β and δ are always fewer than the unity: the first one lies within the interval 0.60–0.98, while the second one is bounded by 0.55–0.90; on the contrary, the exponent γ of the storm duration cdf may be greater than one, being estimated between 0.80–1.35. As above remembered, the unity represents a fundamental boundary for the shape parameter. Therefore, unlike the other two distributions, the duration pdf may be subject to important changes in shape with reference to the discretization parameter settings.

For all the three exponents, greater values are generally estimated when the volume threshold is increased, while different trends are evidenced with respect to the minimum interevent time. In fact, the δ exponent increases according to the IETD, but the opposed behaviour occurs for the γ exponent and, finally, no clear tendency can be detected for the β exponent. The scale parameters ζ, λ and ψ are instead characterized by a wider variability, by which they increase according to both the discretization thresholds in a quite proportional manner.

The reasons for such outcomes reside in the effects of the increase in the discretization thresholds, illustrated in the association analysis context. The scale parameters mainly depend on the mean, whose increments can be immediately justified by the isolation of more extended events separated by longer dry weather periods. The shape parameters are instead principally related to the variance; given the wide scale variability, the dispersion around the mean must be discussed by using the coefficient of variation, which is exclusively a function of the exponent.

HESSD

8, 429–481, 2011

The properties of the Weibull probability function are such that the exponent increase is linked to a reduction of the coefficient of variation. Indeed, if the volume threshold is incremented the frequencies of the smallest observations decrease for all three variables, diminishing the overall distribution dispersion. An analogous explanation may be advanced for understanding the effects of the IETD extension on the dry whether period distributions, because the minimum interevent time directly acts on the variable like a threshold.

This filtering action is not systematically exercised on the wet weather durations, whose minor values do not always disappear from the sample. In fact, when the rainfalls are isolated by long interevent periods, as in the dry seasons, they are not aggregated in more extended events. So, the mean increase is associated with a larger enhancement of the standard deviation, which results in the reduction of the coefficient of variation.

Finally, the lack of a recognizable trend of the rainfall volume dispersion according to the IETD appears to be reasonable because this discretization parameter operates regardless of this quantity: even storms having completely different depths may be joined into a unique event and so the dispersion change is not univocally foreseeable.

The scale parameters deserve a last consideration because, although they do not seem to be sensitive to the precipitation regime, they evidence however to be influenced by another climatic aspect like the annual precipitation amount. Obviously, in the wetter climates, greater ζ and λ values have been estimated, while the dryer ones were found to be characterized by larger ψ. A demonstration of this assertion can be found in Fig. 9 by observing the curves belonging to the Parma series, whose annual mean depth is the lowest among the presented examples.

5.2 Marginal goodness-of-fit tests

The suitability of the Weibull probabilistic model for representing the natural variability of the rainfall event volumes has been already proved by using the confidence boundary test for the Brescia time series. In this work, the same technique has been adopted for assessing the goodness-of-fits in all the illustrated Italian climates regarding the complete set of random variables. The tests are intended to verify whether the Weibull distributions from Eqs. (26) to (28), fitted by way of the maximum likelihood method, can be rejected or not according to an a priori fixed significance level and are based on the standard variables υ, τ and α defined

$$\upsilon = \frac{v - \mathrm{IA}}{\hat{\xi}}; \quad \tau = \frac{t}{\hat{\lambda}}; \quad \alpha = \frac{a - \mathrm{IETD}}{\hat{\psi}}, \tag{29}$$

for the event volume, the wet weather duration and the interevent periods respectively: the corresponding sample versions are easily obtained by inverting the cdfs and substituting the frequencies F (Eq. 2) to the non exceedance probability P:

$$\hat{\upsilon} = \left[\ln\left(\frac{1}{1-F_V}\right)\right]^{\frac{1}{\hat{\beta}}}; \quad \hat{\tau} = \left[\ln\left(\frac{1}{1-F_T}\right)\right]^{\frac{1}{\hat{\gamma}}}; \quad \hat{\alpha} = \left[\ln\left(\frac{1}{1-F_A}\right)\right]^{\frac{1}{\hat{\delta}}}. \tag{30}$$

The confidence limits were centered with respect to the theoretical value and quantified for a significance level of 5% according to the interval half-width Δ (Eq. 31), computed as a function of the non exceedance probability P and the probability density p of the standard variable.

$$\Delta = \pm \frac{1.96}{\sqrt{n}} \frac{\sqrt{P\,(1-P)}}{p} \tag{31}$$

The goodness-of-fits achievable by means of the Weibull probabilistic model are illustrated in the plots of Figs. 10–12, in which the confidence boundaries are drawn between the probability interval 0.20–0.80 and the sample data are derived from the continuous record by using the thresholds IETD=12 h and IA=4 mm. The sample points are fairly aligned with the theoretical line and substantially included in the confidence region. Analogous results were gathered from all the series by using different sets of IETD and IA. So, on the whole, the hypothesis concerning the assumed probabilistic function cannot be rejected.

HESSD
8, 429–481, 2011

6 Conclusions

In this work the possibility of analysing the stochastic structure of the rainfall point process by a quite simple trivariate joint distribution, based on three random variables able to represent the storm main features, the rainfall volume, the wet weather duration and the interevent period, has been clearly demonstrated. The sample construction needs a criterion for isolating independent events from the time series, which may be operatively defined with regard to the derived hydrological process involved in the application.

The copula functions played a fundamental role in the development of the trivariate distribution function relating to these variables. The essential advantages consisted in the capability of constructing unconventional functions by exploiting mixing techniques applied to bivariate copulas and to test the model components separately. The first aspect is related to the a considerable enhancement of the model suiting capacity, while the second one plays a significant role in the control of the computational burden.

In general, it must be pointed out that the copula approach provides a rigorous and objective procedure to face the inference problem in the multivariate framework and to avoid the heuristic approaches and the arbitrary assumptions that affected the joint distribution assessment in the past. Hence, the extreme convenience of this approach must be strongly remarked and its systematic utilization may be definitely advocated.

More specifically, regarding the illustrated results, it is opportune to remark that the dependence structure relating to the rainfall event variables is finely modelled by means of a joint function constructed by mixing the Gumbel-Hougaard and the independence bivariate copulas. The first component, whose unique parameter can be conveniently estimated by using the method of moments, expresses the non negligible positive association between the rainfall volume and the wet weather duration.

The properties of the Gumbel-Hougaard copula evidence that their probability density is not characterized by elliptic level curves, but concentrates both in the lower-left and upper-right corners. However, only in the latter the observed data are sensibly

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

HESSD

8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

associated, while in the earlier area they vary in a more independent manner. The multivariate models that were adopted in the past, such as the normal or the exponential ones, are therefore unable to properly model this kind of dependence structure. This evidence provides a reason for the difficulties encountered in the stochastic representation of the dependent hydrological variables.

The ability of the three parameter Weibull function to fit the marginal distributions of all the analysed random variables was also statistically tested. In the proposed formulation, the distribution lower limit is a priori fixed according to the independent event discretization procedure, while the other two may be estimated by the maximum likelihood criterion. The main appeal of using this alternative model lies in its analytical form, that neither compromises the model versatility required by the sample data nor complicates the parameter assessment.

The analysis considered three time series, each of them correspond to an Italian rainfall regime. Despite the variability of their meteorological characteristics, the analysis revealed that, when the thresholds employed for isolating the independent storms are changed, both the dependence structure and the margins exhibit identical behaviours. Moreover, very similar fitting values were estimated for their parameters, except for those expressing the univariate distribution scale which show to be moderately influenced by the total annual rainfall depth. No detriment of the goodness-of-fit is evidenced in the three climates and a very satisfactory agreement between theoretical functions and observed data is always outlined by the performed statistical tests, for every couple of IETD and IA values that were considered. Thus, a general reliability of the proposed probabilistic model may be deduced for the Italian climates.

The natural research development could be addressed to the flood frequency analysis, by exploiting deterministic or stochastic routing models, or to the watershed hydrologic balance, considered both in its global terms or in its components, infiltration and evapotranspiration. Although the proposed model has quite a simple expression, its implementation in the derivation procedure should reasonably lead to pdfs that cannot be analytically integrated. So, the model effectiveness advance will have to be carefully

HESSD
8, 429–481, 2011

appraised in consideration of such an occurrence, since the capability of the analytical-probabilistic methodology to express the derived probability functions by an analytical form represents one of the most important advantages that justify its use.

Finally, another research perspective is offered by the concerns involved in the estimation of the return period. In fact, in the multivariate case its evaluation still remains an open question, because a universally accepted definition has not been proposed yet. The main reason lies in the ambiguity of having very dissimilar events associated with the same value that arises from the most immediate generalizations of this concept. The analytical-probabilistic methodology offers a way to face this problem from the univariate point of view, because the forcing meteorological variables are directly related to the dependent one.

Acknowledgement. Giuseppe De Martino of the University Federico II and his research group are kindly acknowledged for having provided the Naples precipitation data.

References

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

Bacchi, B., Becciu, G., and Kottegoda, N. T.: Bivariate exponential model applied to intensities and durations of extreme rainfall, J. Hydrol., 155(1–2), 225–236, doi:10.1016/0022-1694(94)90166-X, 1994.

Bacchi, B., Balistrocchi, M., and Grossi, G.: Proposal of a semi-probabilistic approach for storage facility design, Urban Water J., 5(3), 195–208, doi:10.1080/15730620801980723, 2008.

Balistrocchi, M., Grossi, G., and Bacchi, B.: An analytical probabilistic model of the quality efficiency of a sewer tank, Water Resour. Res., 45, W12420, doi:10.1029/2009WR007822, 2009.

Bandini, A.: Tipi Pluviometrici Dominanti Sulle Regioni Italiane, Servizio Idrografico Italiano, Roma, 1931.

Beven, K. J.: Rainfall-Runoff Modelling: the Primer, John Wiley and Sons, Bath, UK, 265–270, 2001.

HESSD

8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

Bonta, J. V. and Rao, R.: Factors affecting the identification of independent storm events, J. Hydrol., 98(3–4), 275–293, doi:10.1016/0022-1694(88)90018-2, 1988.
Chakak, A. and Koehler, K. J.: A strategy for constructing multivariate distributions, Commun. Stat. Simulat., 24(3), 537–550, doi:10.1080/03610919508813257, 1995.
Chan, S.-O. and Bras, R. L.: Urban storm water management: distribution of flood volumes, Water Resour. Res., 15(2), 371–382, doi:10.1029/WR015i002p00371, 1979.
Chow, V. T., Maidment, D. R., and Mays, L. W.: Applied Hydrology, McGraw-Hill International Edition, New York, NY, 1988.
Córdova, J. R. and Rodríguez-Iturbe, I.: On the probabilistic structure of storm surface runoff, Water Resour. Res., 21(5), 755–763, doi:10.1029/WR021i005p00755, 1985.
Cox, D. R. and Isham, V.: Point Processes, Chapman and Hall Ltd., London, UK, 75–81, 1980.
Deheuvels, P.: Empirical dependence function and properties: nonparametric test of independence, B. Cl. Sci. Ac. Roy. Belg., 65(6), 274–292, 1979.
Di Toro, D. M. and Small, M. J.: Stormwater interception and storage, J. Environ. Eng., 105(EE1), 43–54, 1979.
Díaz-Granados, M. A., Valdes, J. B., and Bras, R. L.: A physically based flood frequency distribution, Water Resour. Res., 20(7), 995–1002, doi:10.1029/WR020i007p00995, 1984.
Eagleson, S. P.: Dynamic Hydrology, McGraw-Hill, New York, p. 186, 1970.
Eagleson, S. P.: Dynamics of flood frequency, Water Resour. Res., 8(4), 878–898, doi:10.1029/WR008i004p00878, 1972.
Eagleson, S. P.: Climate, soil, and vegetation. 2. The distribution of annual precipitation derived from observed storm sequences, Water Resour. Res., 14(5), 713–721, doi:10.1029/WR014i005p00713, 1978a.
Eagleson, S. P.: Climate, soil, and vegetation. 5. A derived distribution of storm surface runoff, Water Resour. Res., 14(5), 741–748, doi:10.1029/WR014i005p00741, 1978b.
Eagleson, S. P.: Climate, soil and vegetation. 7. A derived distribution of annual water yield, Water Resour. Res., 14(5), 765–776, doi:10.1029/WR014i005p00765, 1978c.
Fermanian, J.-D.: Goodness-of-fit tests for copulas, J. Multivariate Anal., 95(1), 119–152, doi:10.1016/j.jmva.2004.07.004, 2005.
Foufoula-Georgiou, E. and Guttorp, P.: Assessment of a class of Neyman-Scott models for temporal rainfall, J. Geophys. Res., 92(D8), 9679–9682, doi:10.1029/JD092iD08p09679, 1987.
Frank, M. J.: On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$, Aequationes Math., 19(1), 194–226, doi:10.1007/BF02189866, 1979.

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

Genest, C. and Rémilland, B.: Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models, Ann. I. H. Poincare-Pr., 44(6), 1096–1127, 2008.
Genest, C. and Werker, B. J. M.: Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models, in: Distribution with Given Marginals and Statistical Modelling, Kluwer, Dordrecht, The Nederlands, 103–112, 2002.
Genest, C., Ghoudi, K., and Rivest, L. P.: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika, 82(3), 543–552, doi:10.1093/biomet/82.3.543, 1995.
Genest, C., Rémilland, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: a review and a power study, Insur. Math. Econ., 44(2), 199–213, doi:10.1016/j.insmatheco.2007.10.005, 2009.
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(1–2), 56–67, doi:10.1016/S0022-1694(00)00145-1, 2000.
Grimaldi, S. and Serinaldi, F.: Asymmetric copula in multivariate flood frequency analysis, Adv. Water Resour., 29(8), 1115–1167, doi:10.1016/j.advwatres.2005.09.005, 2006.
Gumbel, E. J.: Distributions des valeurs extrêmes en plusiers dimensions, Publ. Inst. Statist. Univ. Paris, 8, 171–173, 1960.
Guo, Y. and Adams, B. J.: Hydrologic analysis of urban catchments with event-based probabilistic models. 2. Peak discharge rate, Water Resour. Res., 34(12), 3433–3443, doi:10.1029/98WR02448, 1998.
Guo, Y. and Adams, B. J.: An analytical probabilistic approach to sizing flood control detention facilities, Water Resour. Res., 35(8), 2457–2468, doi:10.1029/1999WR900125, 1999.
Hougaard, P.: A class of multivariate failure time distributions, Biometrika, 73(3), 671–678, doi:10.1093/biomet/73.3.671, 1986.
Joe, H.: Multivariate Models and Dependence Concepts, Chapman and Hall, London, 1997.
Junker, M. and May, A.: Measurement of aggregate risk with copulas, Economet. J., 8(3), 428–454, doi:10.1111/j.1368-423X.2005.00173.x, 2005.
Kendall, M. G.: A new measure of the rank correlation, Biometrika, 30(1–2), 81–93, doi:10.1093/biomet/30.1-2.81, 1938.
Kim, G., Silvapulle, M. J., and Silvapulle, P.: Comparison of semiparametric and parametric methods for estimating copulas, Comput. Stat. Data. An., 51(6), 2836–2850, doi:10.1016/j.csda.2006.10.009, 2007.

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

Kruskal, W. H.: Ordinal measures of association, J. Am. Stat. Ass., 53(284), 814–861, doi:10.2307/2281954, 1958.
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(9), 2103–2107, doi:10.1029/97WR00812, 1997.
Li, J. Y. and Adams, B. J.: Probabilistic models for analysis of urban runoff control systems, J. Env. Eng., 126(3), 217–224, doi:10.1061/(ASCE)0733-9372(2000)126:3(217), 2000.
Malevergne, Y. and Sornette, D.: Testing the Gaussian copula hypothesis for financial assets dependences, Quant. Financ., 3(4), 231–250, doi:10.1088/1469-7688/3/4/301, 2003.
Moisello, U.: Grandezze e Fenomeni Idrologici, La Goliardica Pavese, Pavia, IT, 1985.
Nelsen, R. B.: An Introduction to Copulas, 2nd edn., Springer, New York, 2006.
Restrepo-Posada, P. J. and Eagleson, P. S.: Identification of independent rainstorms, J. Hydrol., 55(1–4), 303–319, doi:10.1016/0022-1694(82)90136-6, 1982.
Ruymgaart, F. H.: Asymptotic theory for rank tests for independence, MC Tract 43, Mathematisch Instituut, Amsterdam, 1973.
Salvadori, G. and De Michele, C.: Statistical characterization of temporal structure of storms, Adv. Water Resour., 29(6), 827–842, doi:10.1016/j.advwatres.2005.07.013, 2006.
Salvadori, G. and De Michele, C.: On the use of copulas in hydrology: theory and practice, J. Hydrol. Eng., 12(4), 369–380, doi:10.1061/(ASCE)1084-0699(2007)12:4(369), 2007.
Salvadori, G., De Michele, C., Kottegoda, N. T., and Rosso, R.: Extremes in nature: an approach using copulas, Springer, Dordrecht, The Nederlands, 2007.
Seto, M. Y. K.: Comparison of alternative derived probability distribution models for urban stormwater management, M. A. Sc. Thesis, Department of Civil Engineering, University of Toronto, Toronto, Ontario, 1984.
Scaillet, O.: Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters, J. Multivariate Anal., 98(3), 533–543, doi:10.1016/j.jmva.2006.05.006, 2007.
Shih, J. H.: A goodness-of-fit test for association in a bivariate survival model, Biometrika, 85(1), 189–200, doi:10.1093/biomet/85.1.189, 1998.
Shih, J. H. and Louis, T. A.: Inferences on the association parameter in copula models for bivariate survival data, Biometrics, 51(4), 1384–1399, 1995.
Sklar, A.: Fonctions de répartition à *n* dimensions et leures marges, Publ. Inst. Statist. Univ. Paris, 8, 229–231, 1959.
Singh, K. and Singh, V. P.: Derivation of bivariate probability density functions with exponential

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

marginals, Stoch. Hydrol. Hydraul., 5(1), 55–68, doi:10.1007/BF01544178, 1991.
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(1), W01512, doi:10.1029/2009WR007857, 2010.
Wang, W. J. and Wells, M. T.: Model selection and semiparametric inference for bivariate failure-time data, J. Am. Stat. Assoc., 95(449), 62–72, 2000.
Waymire, E., Gupta, V. K., and Rodriguez-Iturbe, I.: A spectral theory of rainfall intensity at the meso-β scale, Water Resour. Res., 20(10), 1453–1465, doi:10.1029/WR020i010p01453, 1984.
Wanielista, M. P. and Yousef, Y. A.: Storm Water Management, John Wiley and Sons, New York, 1992.
Wood, E. F. and Hebson, C. S.: On hydrologic similarity. 1. Derivation of the dimensionless flood frequency curve, Water Resour. Res., 22(11), 1549–1554, 10.1029/WR022i011p01549, 1986.

Table 1. Analyzed rainfall time series.

Location	Rain gauge	Observation period (years)	Sampling time (min)	Sampling resolution (mm)	Annual mean rainfall (mm)
Brescia	ITAS Pastori	45 (1949–1993)	30	0.2	920
Naples	San Mauro	11 (1998–2009)	10	0.1	999
Parma	University	11 (1987–1997)	15	0.2	811

HESSD
8, 429–481, 2011

M. Balistrocchi and B. Bacchi

Table 2. Estimations of the dependence parameter $\hat{\theta}$ (/) for the precipitation series of Brescia, Parma [ ], and Naples ( ).

IETD (h)	IA (mm) 1			4			7			10		
3	1.61	[1.94]	(1.97)	1.42	[1.57]	(1.64)	1.42	[1.46]	(1.57)	1.33	[1.44]	(1.48)
6	1.75	[1.90]	(2.02)	1.50	[1.62]	(1.73)	1.45	[1.43]	(1.63)	1.39	[1.37]	(1.54)
12	1.78	[1.88]	(2.14)	1.55	[1.63]	(1.86)	1.49	[1.47]	(1.68)	1.44	[1.41]	(1.60)
24	1.87	[1.88]	(2.13)	1.69	[1.72]	(1.96)	1.62	[1.56]	(1.83)	1.54	[1.47]	(1.77)
48	1.92	[2.02]	(2.20)	1.77	[1.79]	(1.90)	1.68	[1.60]	(1.78)	1.63	[1.66]	(1.74)
96	2.19	[1.94]	(2.29)	2.02	[1.80]	(1.93)	1.95	[1.66]	(1.89)	1.97	[1.68]	(1.90)

Table 3. *p*-values (%) obtained when testing the independence between the event volume and the wet weather duration for the rainfall series of Brescia, Parma [ ], and Naples ( ).

IETD (h)	IA (mm) 1			4			7			10		
3	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.4]	(0.0)
6	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[1.2]	(0.0)
12	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.2]	(0.0)	0.0	[0.6]	(0.0)
24	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.3]	(0.0)
48	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)
96	0.0	[0.0]	(0.0)	0.0	[0.0]	(0.0)	0.0	[0.1]	(0.0)	0.0	[0.3]	(0.0)

HESSD
8, 429–481, 2011

Table 4. p-values (%) obtained when testing the independence between the wet weather duration and the interevent period for the rainfall series of Brescia, Parma [ ], and Naples ( ).

IETD (h)	IA (mm) 1			4			7			10		
3	81.1	[78.6]	(84.0)	42.2	[97.8]	(82.8)	47.6	[100.0]	(98.5)	65.9	[100.0]	(96.0)
6	85.0	[96.9]	(91.6)	79.0	[97.2]	(78.2)	87.7	[100.0]	(66.5)	73.8	[100.0]	(59.0)
12	66.2	[100.0]	(100.0)	79.7	[100.0]	(89.8)	87.6	[100.0]	(86.7)	85.8	[100.0]	(54.1)
24	99.9	[100.0]	(94.3)	100.0	[99.0]	(92.0)	100.0	[94.0]	(99.7)	100.0	[99.8]	(97.8)
48	99.2	[96.8]	(99.8)	99.9	[98.7]	(94.2)	96.6	[99.8]	(96.4)	77.6	[100.0]	(100.0)
96	100.0	[100.0]	(100.0)	99.9	[100.0]	(100.0)	99.8	[100.0]	(100.0)	100.0	[100.0]	(98.0)

HESSD
8, 429–481, 2011

M. Balistrocchi and B. Bacchi

Table 5. p-values (%) obtained when testing the independence between the event volume and the interevent dry period for the rainfall series of Brescia, Parma [ ], and Naples ( ).

IETD (h)	IA (mm) 1			4			7			10		
3	91.0	[100.0]	(100.0)	100.0	[92.8]	(94.2)	100.0	[90.4]	(84.3)	99.8	[100.0]	(100.0)
6	98.0	[100.0]	(100.0)	99.8	[88.8]	(100.0)	100.0	[89.7]	(99.5)	100.0	[99.8]	(99.8)
12	75.2	[100.0]	(99.9)	100.0	[100.0]	(99.3)	100.0	[99.8]	(99.7)	98.4	[100.0]	(99.4)
24	94.8	[99.6]	(100.0)	99.9	[99.6]	(100.0)	99.5	[99.8]	(100.0)	91.0	[100.0]	(99.7)
48	100.0	[100.0]	(100.0)	100.0	[100.0]	(100.0)	100.0	[100.0]	(100.0)	100.0	[100.0]	(100.0)
96	100.0	[100.0]	(99.4)	100.0	[99.4]	(97.4)	100.0	[98.1]	(99.8)	100.0	[99.5]	(74.7)

HESSD
8, 429–481, 2011

M. Balistrocchi and B. Bacchi

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀| |▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Table 6. *p*-values (%) obtained when testing the goodness-of-fit of the Gumbel-Hougaard model to the event volume and the wet weather duration pair for the rainfall series of Brescia, Parma [ ], and Naples ( ).

IETD	IA (mm)											
(h)	1			4			7			10		
3	39.1	[100.0]	(98.7)	77.2	[100.0]	(99.8)	90.9	[99.9]	(100.0)	100.0	[100.0]	(99.8)
6	77.0	[99.7]	(99.2)	97.6	[100.0]	(100.0)	100.0	[100.0]	(99.6)	99.9	[100.0]	(98.8)
12	91.9	[100.0]	(100.0)	96.8	[100.0]	(100.0)	100.0	[100.0]	(99.5)	99.9	[100.0]	(99.2)
24	99.7	[100.0]	(100.0)	99.9	[100.0]	(100.0)	100.0	[100.0]	(100.0)	100.0	[100.0]	(100.0)
48	98.4	[100.0]	(100.0)	100.0	[100.0]	(100.0)	100.0	[100.0]	(100.0)	100.0	[100.0]	(100.0)
96	100.0	[100.0]	(100.0)	99.0	[100.0]	(100.0)	91.7	[100.0]	(100.0)	98.3	[100.0]	(100.0)

HESSD
8, 429–481, 2011

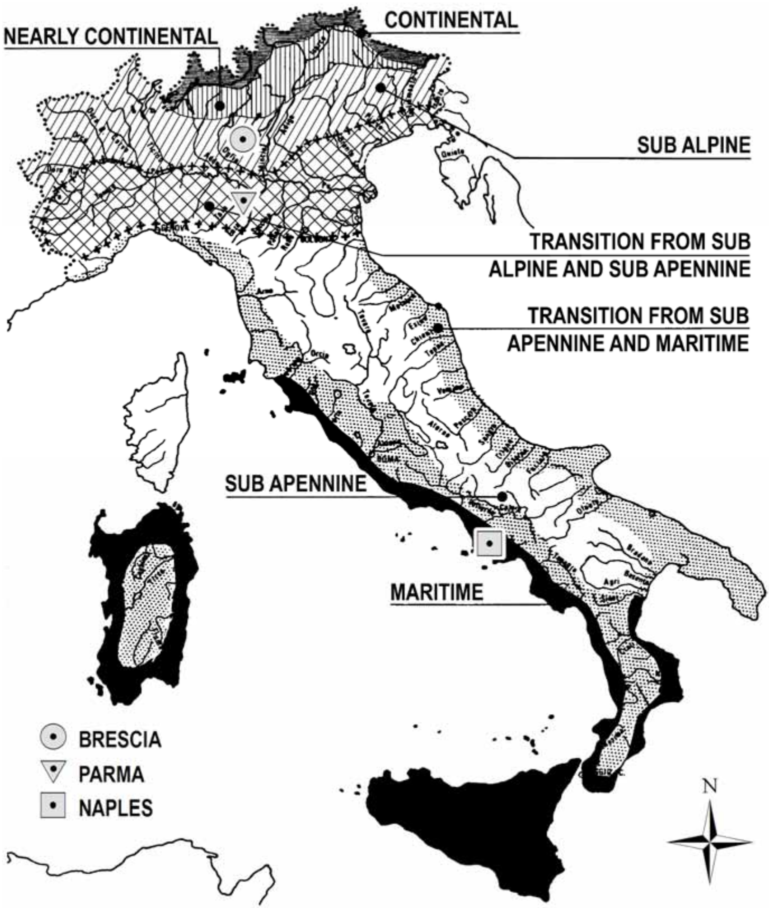

Fig. 1. Italian precipitation regimes and raingauge locations (redrawn from Moisello, 1985).

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

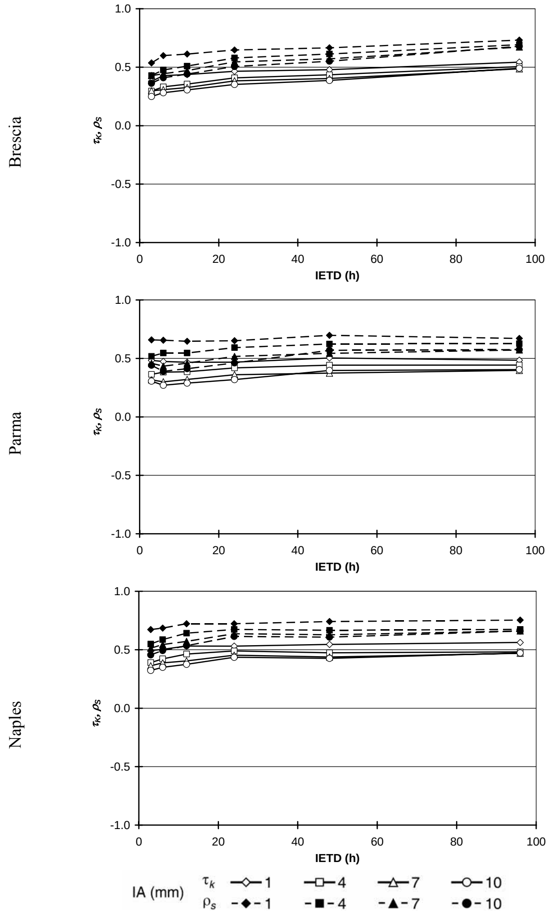

Fig. 2. Trends of the volume-duration rank correlation coefficients with regard to the discretization thresholds.

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

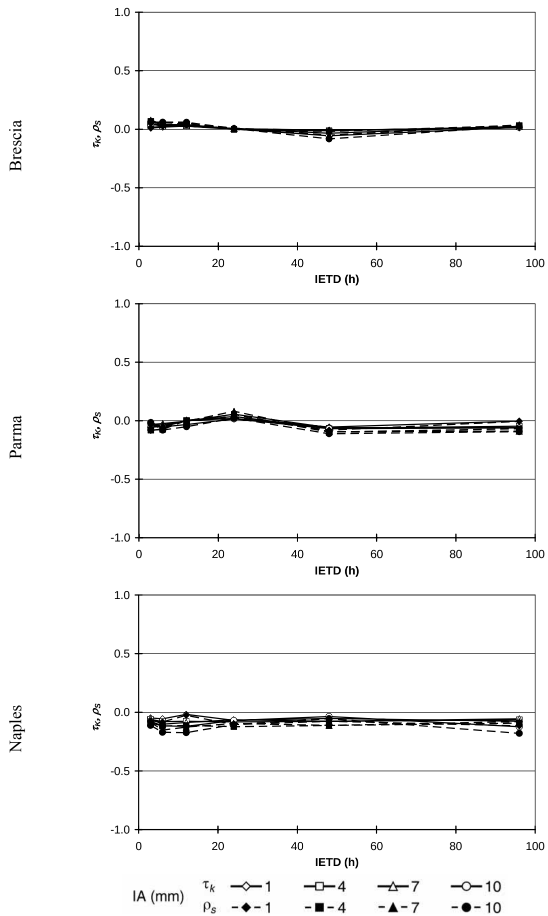

Fig. 3. Trends of the duration-interevent rank correlation coefficients with regard to the discretization thresholds.

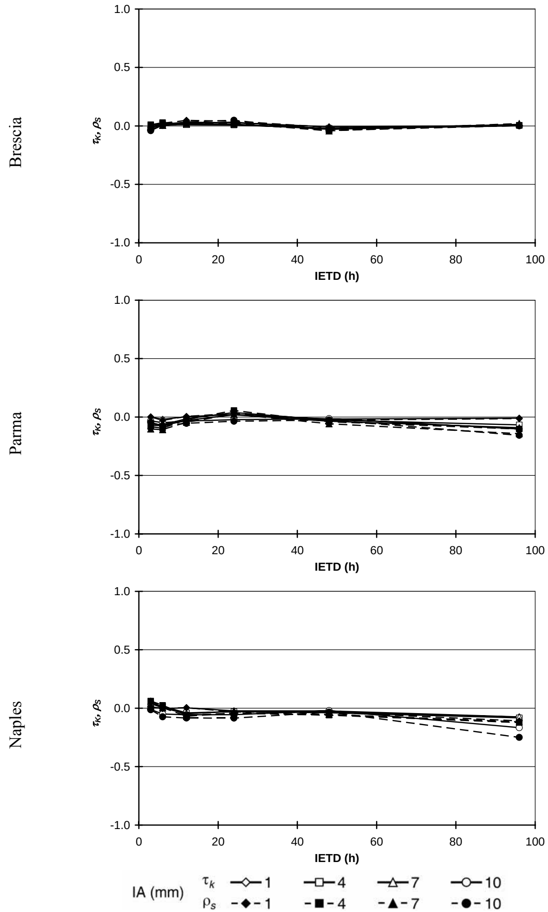

Fig. 4. Trends of the volume-interevent rank correlation coefficients with regard to the discretization thresholds.

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

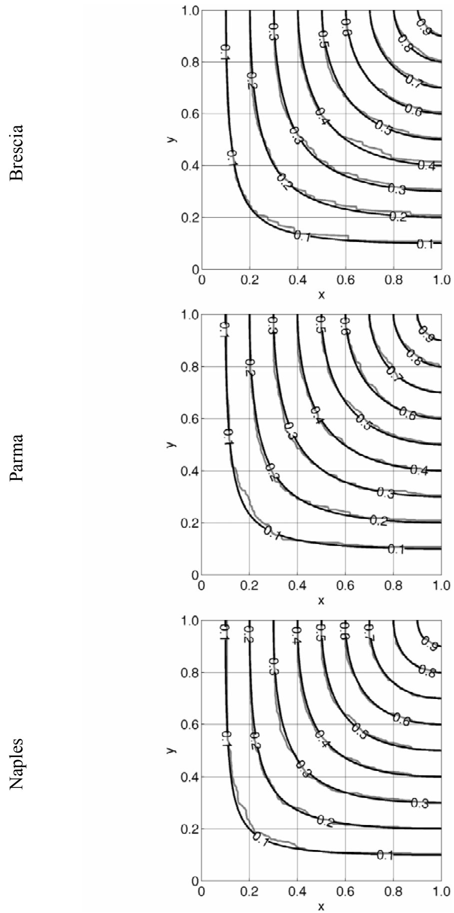

Fig. 5. Contours of the G.-H. (black lines) and the empirical copulas (grey lines) for the volume-duration pair (IETD=12 h; IA=4 mm).

HESSD
8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

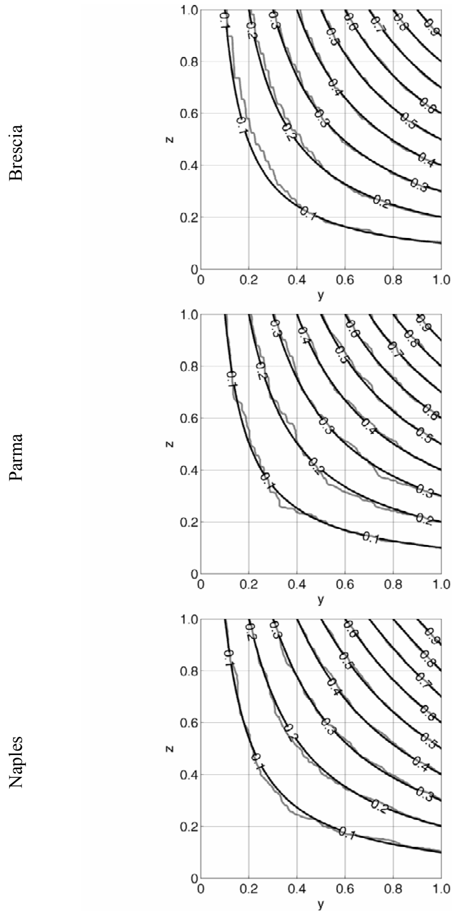

Fig. 6. Contours of the independence (black lines) and the empirical copulas (grey lines) for the duration-interevent pair (IETD=12 h; IA=4 mm).

HESSD

8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

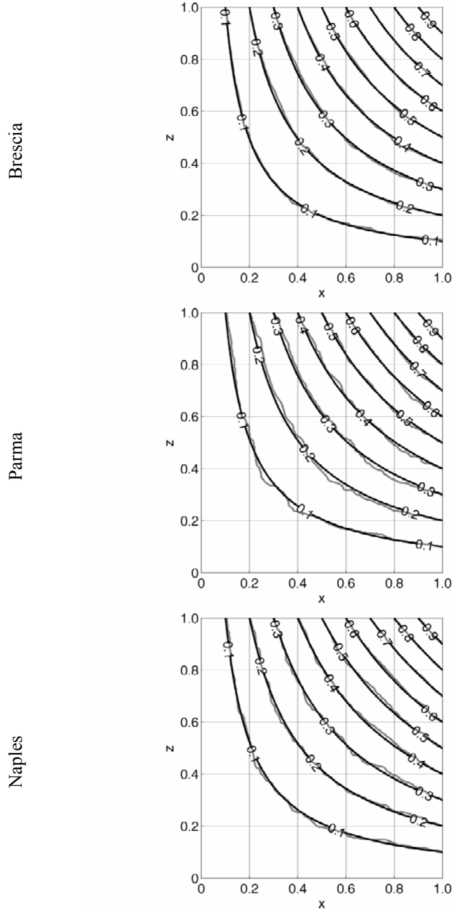

Fig. 7. Contours of the independence (black lines) and the empirical copulas (grey lines) for the volume-interevent pair (IETD=12 h; IA=4 mm).

HESSD
8, 429–481, 2011

M. Balistrocchi and B. Bacchi

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

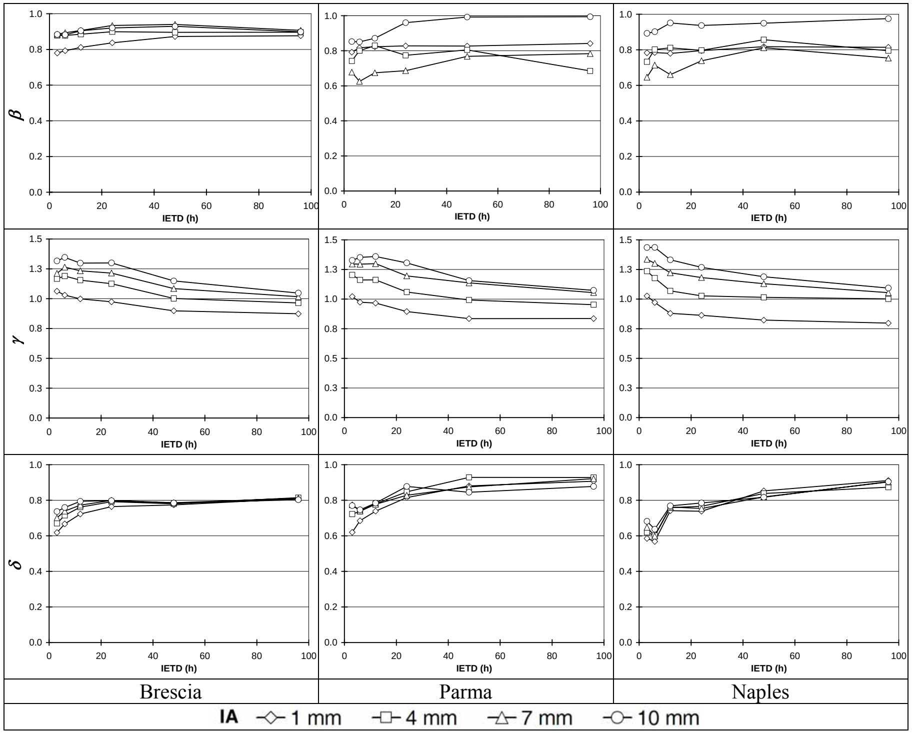

Fig. 8. Trends of the margin shape parameters with regard to the discretization thresholds.

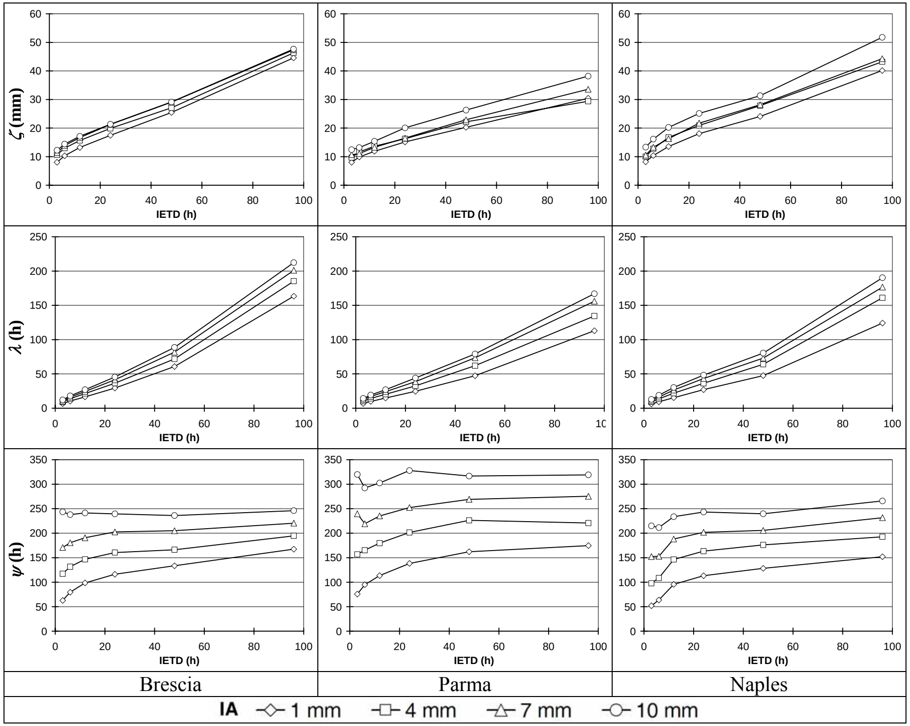

Fig. 9. Trends of the margin scale parameters with regard to the discretization thresholds.

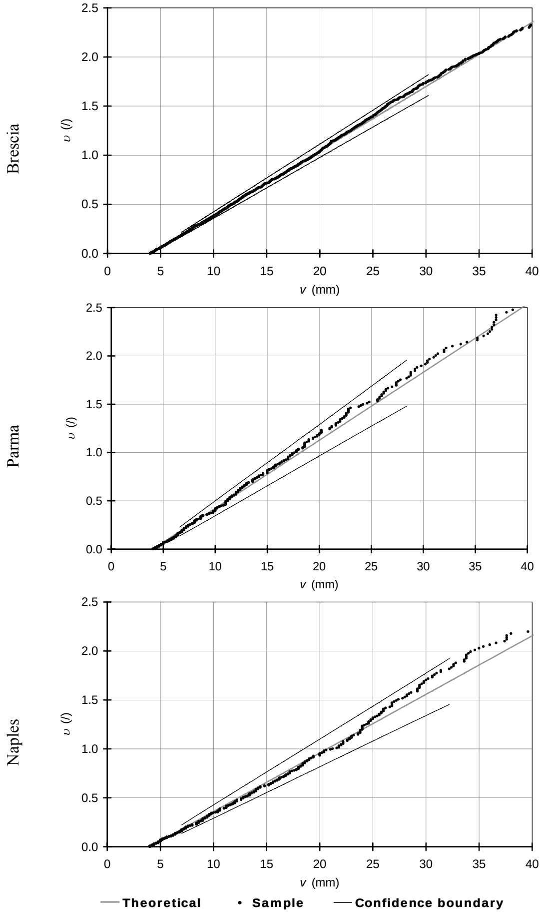

Fig. 10. Confidence boundary tests for the event volume distribution fits (IETD=12 h; IA=4 mm).

HESSD
8, 429–481, 2011

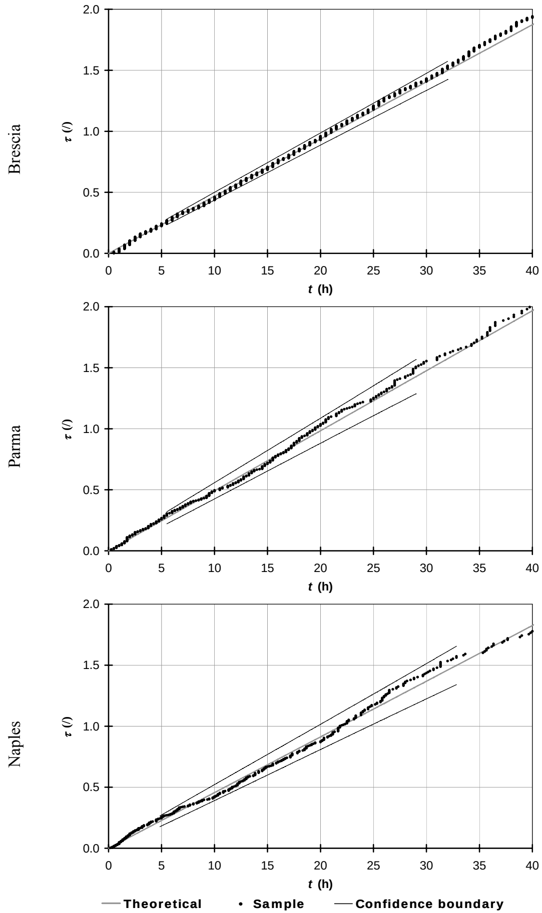

Fig. 11. Confidence boundary tests for the wet weather duration distribution fits (IETD=12 h; IA=4 mm).

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

HESSD

8, 429–481, 2011

Modelling the statistical dependence

M. Balistrocchi and B. Bacchi

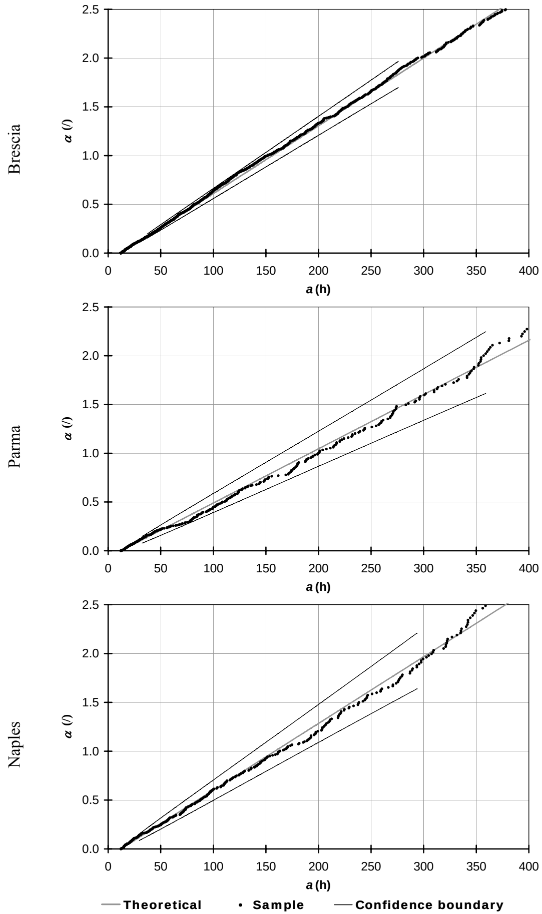

Fig. 12. Confidence boundary tests for the interevent time distribution fits (IETD=12 h; IA=4 mm).

HESSD

8, 429–481, 2011