Multivariate multiparameter extreme value models and return periods: A copula approach

G. Salvadori[1,2] and C. De Michele[2]

Received 23 December 2009; revised 18 June 2010; accepted 29 June 2010; published 6 October 2010.

[1] Multivariate extreme value models are a fundamental tool in order to assess potentially dangerous events. The target of this paper is two-fold. On the one hand we outline how, exploiting recent theoretical developments in the theory of copulas, new multivariate extreme value distributions can be easily constructed; in particular, we show how a suitable number of parameters can be introduced, a feature not shared by traditional extreme value models. On the other hand, we introduce a proper new definition of multivariate return period and show the differences with (and the advantages over) the definition presently used in literature. An illustration involving flood data is presented and discussed, and a generalization of the well-known multivariate logistic Gumbel model is also given.

Citation: Salvadori, G., and C. De Michele (2010), Multivariate multiparameter extreme value models and return periods: A copula approach, *Water Resour. Res.*, *46*, W10501, doi:10.1029/2009WR009040.

1. Introduction

[2] The multivariate extreme problems in hydrology and water resources engineering include, among others, (1) flood and drought occurrences at different sites [*Singh*, 1986; *Pons*, 1992; *Kim et al.*, 2003; *Keef et al.*, 2009], (2) the precipitation dynamics (rain and snow) [*Wilks*, 1998; *Herr and Krzysztofowicz*, 2005], (3) the link between water quality and quantity in a river section [*Grenney and Heyse*, 1985], and (4) the hydraulic conductivity in porous media [*Journel and Alabert*, 1988; *Russo*, 2009].

[3] As frequently stressed in the hydrologic literature, the statistical analysis of multivariate extremes is difficult, essentially due to (1) the complexity of the phenomena, (2) the reduced sample size of the actual multivariate data sets, and (3) a lack of availability of suitable multivariate probability distributions [*Katz et al.*, 2000].

[4] In the past, the joint probability distributions were often assumed to be multinormal (or its extensions, such as the multivariate lognormal's, Student's *t* and Fisher's *F* distributions [see, e.g., *Alexander*, 1954; *Stedinger*, 1983; *Hosking and Wallis*, 1988; *Pons*, 1992; *Kottegoda and Natale*, 1994]), or multivariate logistic Gumbel [see, e.g., *Raynal-Villasenor and Salas*, 1987; *Bacchi et al.*, 1994; *Yue*, 2000b].

[5] Unfortunately, quite a few multivariate distributions present in literature, which are direct extensions of well known univariate ones, suffer from several limitations and constraints: for instance, the marginals may belong to the same probability family. As a further drawback, consider, e.g., the multinormal distribution, where all the pair-wise dependencies are parametrized via Pearson's correlation coefficient. On the one hand, it may not be the best way to model the nonlinear associations exhibited by extremes; on the other hand, it may not exist when heavy-tailed variables are at play and thus may not be suitable to describe extremes. In addition, distributions like, e.g., the multivariate logistic Gumbel, may provide a limited description of the extreme dynamics, since they involve only one (or few) parameter(s) to model all the dependencies, independently of the dimension of the problem. Only recently, new multivariate extreme value models were introduced to overcome some of these limitations (see below and also *Tawn* [1990], *Coles and Tawn* [1991], *Joe* [1994], and *Rootzén and Tajvidi* [2006]).

[6] The use of copulas in hydrology, as well as in other geophysical and environmental sciences, is recent and rapidly growing. Incidentally, we observe that all the multivariate distributions present in literature can be written in a straightforward way in terms of suitable copulas. In this paper we show how, exploiting recent theoretical developments, multivariate extreme value models (hereafter, MEV) can be easily constructed via copulas, leading to new model formulations. In addition, we also demonstrate how the notion of multivariate return period is best investigated in terms of copulas. Our approach is quite general and physically based. Furthermore, the simulation of multivariate scenarios is simple and fast.

[7] Below, in section 2, we introduce the concept of multivariate extreme value copulas, describing some of the mathematical features of interest here. In section 3 we discuss a suitable new definition of multivariate return period, and show the differences with (and the advantages over) the definition presently used in literature. In section 4 we outline a simple procedure for introducing a suitable number of extra parameters in a given copula model, which may yield the construction of new MEV models. Finally, in section 5 an application to maximum annual flood data is presented and discussed. Incidentally, in the present paper we also show and

[1]Dipartimento di Matematica, Università del Salento, Lecce, Italy.
[2]DIIAR (Sezione CIMI), Politecnico di Milano, Milan, Italy.

0043-1397/10/2009WR009040

comment some of the common errors performed by unskilled practitioners while carrying out a copula analysis.

2. MEV Copulas: An Overview

[8] In this section we briefly outline the mathematics of copulas needed in the sequel; for a thorough theoretical and practical introduction see, e.g., *Joe* [1997], *Nelsen* [2006], and *Salvadori et al.* [2007]. Hereafter, for any integer $d > 1$, we use the vector notation in $\mathbf{R}^d$, i.e., $\mathbf{x} = (x_1, \ldots, x_d)$; operations and inequalities are to be intended componentwise. Also, $\mathbf{I} = [0, 1]$ will denote the unit interval, and $\mathbf{I}^d$ will denote the d dimensional unit cube.

[9] The main target pursued here is to provide a general multivariate framework for modeling extreme non-independent observations sampled via a network of gauge stations; the particular situation of independent ones will be included as a special case. As shown below, this can easily be achieved by using copulas. The random variables (hereafter, r.v.s) considered in this work may represent, for instance, rainfall or flood measurements collected in a given basin, or pollution samples in a region, or wave data collected by marine buoys. Below, $\mathbf{S} = \{S_1, \ldots, S_d\}$ will denote a set of d gauge stations.

[10] The problem of specifying a probability model for dependent multivariate observations can be simplified by expressing the corresponding d dimensional joint cumulative distribution F in terms of its marginals $F_1, \ldots, F_d$, and the associated copula $\mathbf{C}$, implicitly defined through the following functional identity stated by Sklar's Theorem [*Sklar*, 1959],

$$F(x_1, \ldots, x_d) = \mathbf{C}(F_1(x_1), \ldots, F_d(x_d)). \quad (1)$$

A multivariate copula $\mathbf{C}(u_1, \ldots, u_d)$ is simply a joint distribution over $\mathbf{I}^d$ with Uniform marginals. The link between d-copulas and multivariate distributions is provided by equation (1). If $F_1, \ldots, F_d$ are all continuous, then $\mathbf{C}$ is unique [*Sklar*, 1959].

[11] A copula $\mathbf{C}$ is MEV if it is max-stable [*Galambos*, 1987; *Nelsen*, 2006], i.e., if it satisfies the equation

$$\mathbf{C}(u_1^t, \ldots, u_d^t) = [\mathbf{C}(u_1, \ldots, u_d)]^t \quad (2)$$

for all $\mathbf{u} \in \mathbf{I}^d$ and all $t > 0$. A distribution F is MEV if, and only if, all its marginals F_i's are generalized extreme value laws (hereafter, GEV), and the corresponding copula $\mathbf{C}$ is MEV [*Joe*, 1997]. Furthermore, since the GEV law is continuous, the representation (1) of a MEV distribution F is unique. Most importantly, by exploiting the invariance property of copulas [*Nelsen*, 2006; *Salvadori et al.*, 2007], we may restrict our attention to copulas only, and do not worry about the GEV marginals, as we shall do hereinafter.

[12] It is fundamental to note that not all copulas are MEV (i.e., satisfy the max-stability property (2) (see, e.g., many of the examples by *Nelsen* [2006] and the illustrations given by *Salvadori et al.* [2007])), and consequently cannot be used to construct consistent MEV models. For instance, a distribution having GEV marginals and a Gaussian copula (as in the model proposed by *Renard and Lang* [2007] for the multivariate analysis of annual maxima) is not a MEV law, except for trivial values of the parameters corresponding to the independence MEV copula $\mathbf{\Pi}_d(\mathbf{u}) = u_1 \cdots u_d$: thus, for the sake of theoretical consistency, it should not be used for modeling extremes. However, in several practical problems, the fulfillment of the MEV assumptions is not the ultimate scope, and alternatives cannot be excluded, even though they are not theoretically consistent in terms of MEV hypotheses.

[13] An important notion of interest is represented by the Pickands' dependence function A [*Pickands*, 1981]. A bivariate copula $\mathbf{C}$ is MEV if there exists a convex function $A{:}\mathbf{I} \to [1/2, 1]$, satisfying the constraint $\max\{t, 1 - t\} \le A(t) \le 1$ for all $t \in \mathbf{I}$, such that

$$\mathbf{C}(u, v) = \exp\left[\ln(uv)\, A\left(\frac{\ln v}{\ln(uv)}\right)\right]$$

for all $(u, v) \in \mathbf{I}^2$. In particular [see, e.g., *Nelsen*, 2006], if $A(t) \equiv 1$ then $\mathbf{C} = \mathbf{\Pi}_2 = uv$ (the independence bivariate copula), and if $A(t) = \max\{t, 1 - t\}$ then $\mathbf{C} = \mathbf{M}_2 = \min\{u, v\}$ (the comonotone bivariate copula), modeling full dependence. For a generalization and extension of the above dependence structure to the multivariate case see [*Falk and Reiss*, 2005]. The interesting point is that the one-dimensional function A is a substitute for the bivariate MEV copula $\mathbf{C}$, and provides the overall structure of dependence for the pair of variables of interest. Since A can be estimated via empirical data [*Genest and Segers*, 2009], then it may be used to check the statistical adequacy of different models. We shall see later how to use Pickands' dependence function.

[14] Finally, below we shall also use the Kendall's distribution (or measure) function $K_{\mathbf{C}}{:}\mathbf{I} \to \mathbf{I}$ [*Genest and Rivest*, 1993, 2001] given by

$$K_{\mathbf{C}}(t) = \mathbf{P}\{W \le t\} = \mathbf{P}\{\mathbf{C}(U_1, \ldots, U_d) \le t\}, \quad (3)$$

where $t \in \mathbf{I}$ is a probability level, $W = \mathbf{C}(U_1, \ldots, U_d)$ is a univariate r.v. taking value on $\mathbf{I}$, and the U_i's are Uniform r.v.s on $\mathbf{I}$ with copula $\mathbf{C}$. Note that, in the bivariate extreme value case, $K_{\mathbf{C}}$ is given by [*Ghoudi et al.*, 1998],

$$K_{\mathbf{C}}(t) = t - (1 - \tau_{\mathbf{C}})\, t \ln t, \quad (4)$$

where $\tau_{\mathbf{C}}$ is the value of the Kendall's τ associated with the copula $\mathbf{C}$. Clearly, bivariate MEV copulas with the same value of τ share the same function $K_{\mathbf{C}}$. Unfortunately, at present no useful expressions similar to equation (4) are known for the general multivariate case $d > 2$, and it is necessary to resort to simulations. A possible algorithm is as follows.

Algorithm 1: Calculation of Kendall's $K_{\mathbf{C}}$.

1. Generate a sample $\mathbf{u}_1, \ldots, \mathbf{u}_m$ from copula $\mathbf{C}$.
2. For $i = 1, \ldots, m$ calculate $v_i = \mathbf{C}(\mathbf{u}_i)$.
3. For $t \in \mathbf{I}$ estimate $\widehat{K}_{\mathbf{C}}(t) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}(v_i \le t)$.

[15] Elementary statistical considerations show that the procedure outlined above yields a maximum likelihood estimator of $K_{\mathbf{C}}$: in fact, it is only a matter of calculating the probability of "success" in the Bernoulli process ruled by the indicator function in Step (3). Moreover, by the Glivenko-Cantelli Theorem [*Billingsley*, 1995], the empirical distribution function $\widehat{K}_{\mathbf{C}}$ is a consistent estimator of $K_{\mathbf{C}}$.

[16] The function $K_{\mathbf{C}}$ plays a fundamental role in the new definition of a consistent multivariate return period (see section 3): since empirical estimators of the Kendall's distribution function exist [*Genest et al.*, 2009], then interesting

analyses of practical utility can be performed on the available data (see section 5).

3. Multivariate Return Periods

[17] The return period of a prescribed event is adopted in hydrology as a criterion for design purposes, and provides a simple means for risk analysis: its definition is "the average time elapsing between two successive realizations of the given event."

[18] While the theory of return periods is clear in the univariate case, it requires some care in the multivariate one. First of all, note that equation (3) represents a multivariate quantile relationship, since it corresponds to a multidimensional probability integral transform [*Genest et al.*, 2006]. Thus, the Kendall's measure $K_{\mathbf{C}}$ is a fundamental tool for introducing a consistent (copula-based) definition of a return period for multivariate events (see also the discussions by *Salvadori* [2004], *Salvadori and De Michele* [2004], *Salvadori et al.* [2007], and *Durante and Salvadori* [2010]).

[19] In the univariate setting, where only a single r.v. X with distribution F_X is involved, the return period is generally defined as

$$T_p = \frac{\mu}{1-p} = \frac{\mu}{1-F_X(x_p)}, \tag{5}$$

where μ is the average interarrival time of the events in the sequence observed (e.g., $\mu = 1$ year for annual maxima), $p \in \mathbf{I}$ is a critical probability level (usually, p = 90, 95, 99%, or any other threshold of interest), and $x_p = F_X^{[-1]}(p)$ is the critical design quantile of X of order p. Here $F_X^{[-1]}$ indicates the generalized (or pseudo-) inverse [*Nelsen*, 2006] of the corresponding function. Thus, given an arbitrary return period T_p, both the corresponding critical probability level p and the critical design quantile x_p can be calculated by inverting equation (5).

[20] Evidently, the definition of the return period involves a dichotomic dynamics: either an event is (1) subcritical, i.e., $X < x_p$ (or, equivalently, $F_X(x) < p$), with return period smaller than T_p, or it is (2) supercritical, i.e., $X > x_p$ (or, equivalently, $F_X(x) > p$), with return period larger than T_p. Clearly, subcritical events are "safe" realizations, whereas the supercritical ones are potentially dangerous and destructive. It is essential to understand that $p = F_X(x_p)$ is the probability measure of the subcritical region $\{x \in \mathbf{R}: x < x_p\}$ as induced by the law of F_X of X, i.e., the probability that an event be subcritical at any realization of the process. A similar comment holds for the supercritical region $\{x \in \mathbf{R}: x > x_p\}$, having probability $1 - F_X(x_p)$. As a conclusion, the return period is essentially a measure theoretic notion.

[21] The standard definition of multivariate return period mimics the univariate one given above (for a thorough review see *Zhang* [2005 and references therein] and *Singh et al.* [2007 and references therein]). In particular, thanks to Sklar's Theorem, the multivariate law $F(\mathbf{x})$ can be expressed in terms of a suitable copula $\mathbf{C}(\mathbf{u})$ (see equation (1)), and then $\mathbf{C}$ is used in place of the univariate law F_X in equation (5),

$$T_p = \frac{\mu}{1-p} = \frac{\mu}{1-F(\mathbf{x})} = \frac{\mu}{1-\mathbf{C}(\mathbf{u})}, \tag{6}$$

where $\mathbf{u} = (u_1, \ldots, u_d) = (F_1(x_1), \ldots, F_d(x_d))$ is any point that lies on the critical iso-hypersurface given by $\mathcal{S}_p = \{\mathbf{u} \in \mathbf{I}^d: \mathbf{C}(\mathbf{u}) = p\}$. Therefore, since $p = \mathbf{C}(\mathbf{u}) = \mathbf{P}\{U_1 \le u_1, \ldots, U_d \le u_d\}$, each point $\mathbf{u} \in \mathcal{S}_p$ identifies a (hyper-rectangular) subcritical region $R_p(\mathbf{u})$ given by

$$R_p(\mathbf{u}) = \{\mathbf{v} \in \mathbf{I}^d: \mathbf{v} < \mathbf{u}\},$$

having probability $p = \mathbf{C}(\mathbf{u})$. Clearly, since $R_p(\mathbf{u})$ depends upon $\mathbf{u}$, different points $\mathbf{u}$ belonging to $\mathcal{S}_p$ generate different subcritical regions, all carrying a probability p. As shown below, the definition (6) is incorrect from a measure theoretic point of view, and provides an incoherent tool for dealing with multivariate return periods.

[22] Indeed, equation (6) does not introduce a dichotomic dynamics, as in the univariate case: all the infinite events on $\mathcal{S}_p$ have the same return period T_p, but different points belonging to $\mathcal{S}_p$ yield different (and overlapping) subcritical regions. In fact, let $\mathbf{u}_1 \neq \mathbf{u}_2$ lie on $\mathcal{S}_p$; then, the corresponding subcritical regions $R_p(\mathbf{u}_1)$ and $R_p(\mathbf{u}_2)$ are different, and partially overlap. In other words, although $\mathbf{C}$ is the multivariate distributional equivalent of F_X, this definition does not identify a unique subcritical region of $\mathbf{I}^d$ that exactly supports a probability p as induced by $\mathbf{C}$. Practically, it is then impossible to decide which are the subcritical events (as well as the dangerous supercritical ones).

[23] The way to bypass the inconsistency problem mentioned above is to make a proper use of the Kendall's measure $K_{\mathbf{C}}$, which only depends upon $\mathbf{C}$, and not on any event belonging to $\mathcal{S}_p$. Namely, we define the multivariate return period as

$$T_p = \frac{\mu}{1-p} = \frac{\mu}{1-K_{\mathbf{C}}(t)} = \frac{\mu}{1-\mathbf{P}\{\mathbf{u} \in \mathbf{I}^d: \mathbf{C}(\mathbf{u}) \le t\}}, \tag{7}$$

where the critical threshold $t \in \mathbf{I}$ is given by

$$t = \inf\{s \in \mathbf{I}: K_{\mathbf{C}}(s) = p\} = K_{\mathbf{C}}^{[-1]}(p),$$

by analogy with the correct definition of quantile: in fact, recall that $K_{\mathbf{C}}$ is a distribution function, and hence t represents the corresponding proper quantile of order p. Since $K_{\mathbf{C}}$ is generally nonlinear ($K_{\mathbf{C}}(t) = t$ only if $\mathbf{C} = \mathbf{M}_d = \min\{u_1, \ldots, u_d\}$, the comonotone copula modeling full dependence), then $t \neq p$, and therefore $\mathcal{S}_t \neq \mathcal{S}_p$, the latter one being the critical iso-hypersurface involved by equation (6). In particular, for MEV copulas the relation

$$K_{\mathbf{C}}(t) \ge t$$

holds [*Capéraà et al.*, 1997], and therefore

$$T_p = T_{K_{\mathbf{C}}(t)} = \frac{\mu}{1-K_{\mathbf{C}}(t)} \ge \frac{\mu}{1-t} = T_t, \tag{8}$$

where the right-most term corresponds to the standard return period prescribed by equation (6). Evidently, the above result shows that the use of the standard definition may yield underestimates of the correct value: practically, it is assumed that the Kendall's distribution of the generic copula $\mathbf{C}$ always equals the one of $\mathbf{M}_d$, which can make little sense.

[24] As a conclusion, a unique subcritical region, only depending upon $\mathbf{C}$ and p, and exactly carrying the critical design probability p, is simply given by the set

$$\mathcal{B}_{\mathbf{C}}(p) = \{\mathbf{u} \in \mathbf{I}^d: \mathbf{C}(\mathbf{u}) \le t = K_{\mathbf{C}}^{[-1]}(p)\}.$$

Then, $\mathbf{I}^d$ is partitioned into two disjoint regions by the iso-hypersurface

$$\mathcal{S}_t = \{\mathbf{u} \in \mathbf{I}^d: \mathbf{C}(\mathbf{u}) = t = K_{\mathbf{C}}^{[-1]}(p)\}. \tag{9}$$

At any realization of the process, only two mutually exclusive things may happen: either the event is subcritical and belongs to $\mathcal{B}_{\mathbf{C}}$ (with probability p, and return period not larger than T_p), or it is supercritical and belongs to the complementary set $\overline{\mathcal{B}}_{\mathbf{C}}$ (with probability $1 - p$, and return period not smaller than T_p).

[25] Note that also the iso-hypersurface $\mathcal{S}_p$ partitions the unit hypercube $\mathbf{I}^d$ into a subcritical and a supercritical region: say, $R_1 = \{\mathbf{u} \in \mathbf{I}^d{:}\mathbf{C}(\mathbf{u}) < p\}$, and $R_2 = \{\mathbf{u} \in \mathbf{I}^d{:}\mathbf{C}(\mathbf{u}) > p\}$. However, generally these two regions do not carry the correct required probabilities, i.e., p and $1 - p$. The error performed by unskilled practitioners is to confuse the value taken by the copula on $\mathcal{S}_p$, i.e., $\mathbf{C}(\mathbf{u}) = p$, as given by equation (6), with the probability measure induced by $\mathbf{C}$. In fact, R_1 has probability $p_1 = K_{\mathbf{C}}(p) \ge p$, and R_2 has probability $p_2 = 1 - p_1 \le 1 - p$.

[26] Practically, at any realization of the process, either the event falls in the subcritical region R_1 (with probability $p_1 \ge p$), or in the supercritical one R_2 (with probability $p_2 \le 1 - p$). Thus, the correct return period of the supercritical (potentially destructive) events in R_2 is $T = \mu/p_2$, different from (and larger than) the standard return period $T_p = \mu/(1 - p)$, as given by equation (6). Clearly, this may result in a wrong estimation of the risk.

4. An Extraparametrization Technique

[27] In this section we sketch out a simple procedure for introducing a suitable number of extra parameters in a given copula model. In general, this may improve the modeling features of the multivariate dependence structure considered: in fact, it may provide a better way to capture the interdependencies between the variables at play. Note that the approach presented here may be applied to any copula, and that the new parameters may describe and rule how each variable affects the others, and/or how each variable is affected by the others. The basic theoretical result is given by the following theorem, which represents a special case of a more general result stated by *Liebscher* [2008].

Theorem 1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be d-copulas, and let* $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbf{I}^d$ *be a set of d parameters. Then the function* $\mathbf{C}_{a_1,\ldots,a_d}$ *given by*

$$\begin{aligned}\mathbf{C}_{\mathbf{a}}(\mathbf{u}) &= \mathbf{A}(\mathbf{u}^{\mathbf{a}}) \cdot \mathbf{B}(\mathbf{u}^{\mathbf{1}-\mathbf{a}}) \\ &= \mathbf{A}(u_1^{a_1}, \ldots, u_d^{a_d}) \cdot \mathbf{B}(u_1^{1-a_1}, \ldots, u_d^{1-a_d})\end{aligned} \tag{10}$$

also defines a family of d-copulas. In particular, if $\mathbf{A}$ *and* $\mathbf{B}$ *are MEV, then so is* $\mathbf{C}_{\mathbf{a}}$.

[28] These copulas have been referred to as the Khoudraji family by *Durante and Salvadori* [2010] and generalize the ones introduced by *Khoudraji* [1995] and *Genest et al.* [1998]. The construction method stated by Theorem 1 provides a straightforward way to give a clear interpretation of the new parameters a_i's. In fact, suppose that $\mathbf{a} = \mathbf{1}$: then, $\mathbf{C}_1 = \mathbf{A}$. Conversely, should it be $\mathbf{a} = \mathbf{0}$, then $\mathbf{C}_0 = \mathbf{B}$. For other values of $\mathbf{a}$, the resulting copula $\mathbf{C}_{\mathbf{a}}$ would be a sort of "mixture" between $\mathbf{A}$ and $\mathbf{B}$: in particular, the a_i's play the role of "local" mixing parameters.

[29] Samples extracted from copulas of type (10) can be easily simulated via the following algorithm.

Algorithm 2: Simulation of Khoudraji's copulas.

1. Generate the variates $s_1,\ldots,s_d$ from copula $\mathbf{A}$.
2. Generate the variates $t_1,\ldots,t_d$ from copula $\mathbf{B}$.
3. For $i = 1,\ldots,d$ return $u_i = \max\{s_i^{1/a_i}, t_i^{1/(1-a_i)}\}$.

[30] In particular, the copulas used in the case study analyzed in section 5 can be easily investigated and simulated via a package written for "R" [*Yan and Kojadinovic*, 2009; *Kojadinovic and Yan*, 2010b], or by the procedure described by *Salvadori et al.* [2007, Appendix A.2].

[31] Most importantly, should further parameters be needed (say, another set $\mathbf{b} = \{b_1,\ldots,b_d\}$), the above procedure can be easily iterated as follows. First, simply consider the copula $\mathbf{C}_{\mathbf{a}}$ in equation (10), and a new copula $\mathbf{D}$. Then, construct

$$\mathbf{C}_{\mathbf{a};\mathbf{b}}(\mathbf{u}) = \mathbf{C}_{\mathbf{a}}(\mathbf{u}^{\mathbf{b}})\, \mathbf{D}(\mathbf{u}^{\mathbf{1}-\mathbf{b}}).$$

Again, if the copulas $\mathbf{C}_{\mathbf{a}}$ and $\mathbf{D}$ are MEV, then so is $\mathbf{C}_{\mathbf{a};\mathbf{b}}$. Practically, this building technique may provide a sort of hierarchical constructing method. Clearly, the actual number of new parameters introduced may depend upon the specific analytical structure of the copulas at play.

[32] Evidently, the MEV model given by equation (10) is quite general and rich: an infinite number of variants is possible via a suitable selection of the copulas of interest, and new MEV dependence structures can be easily introduced. In particular, as we shall see in section 5, a known MEV copula can be used as a starting model, and then extraparametrized and enriched via Theorem 1. As an illustration, later we shall use the following MEV copulas.

[33] 1. The Gumbel model, which is the four-variate Gumbel copula [*Nelsen*, 2006; *Salvadori et al.*, 2007],

$$\mathbf{G}_\theta(\mathbf{u}) = \exp\left\{-\left[\sum_{i=1}^{4}(-\ln u_i)^\theta\right]^{1/\theta}\right\}, \tag{11}$$

with parameter $\theta \ge 1$.

[34] 2. The X-Gumbel model, which is a four-variate version of the Gumbel copula $\mathbf{G}$ given above, extraparametrized via Theorem 1 [see also *Liebscher*, 2008, section 2.4]:

$$\begin{aligned}\mathbf{H}(\mathbf{u}) &= \mathbf{G}_\xi(\mathbf{u}^{\mathbf{a}}) \cdot \mathbf{G}_\chi(\mathbf{u}^{\mathbf{1}-\mathbf{a}}) \\ &= \mathbf{G}_\xi(u_1^{a_1}, u_2^{a_2}, u_3^{a_3}, u_4^{a_4}) \\ &\quad \cdot \mathbf{G}_\chi(u_1^{1-a_1}, u_2^{1-a_2}, u_3^{1-a_3}, u_4^{1-a_4}),\end{aligned} \tag{12}$$

with Gumbel parameters ξ, $\chi \ge 1$, and extra parameters a_1, a_2, a_3, $a_4 \in \mathbf{I}$.

[35] The reason for choosing these two models is coherent with the methodological nature of this paper. In fact, the Gumbel copula $\mathbf{G}_\theta$ represents a sort of "standard" MEV model in hydrology [see, e.g., *Yue*, 2000a, 2000b; *Zhang and Singh*, 2007 and references therein]; in particular, it is the only Archimedean copula that is also MEV [*Genest and Rivest*, 1989]. However, it may be inadequate for modeling multivariate phenomena: for any given $\tilde{d} < d$, its lower $\tilde{d}$-dimensional marginals are all identical, and the use of a single global dependence parameter θ to deal with

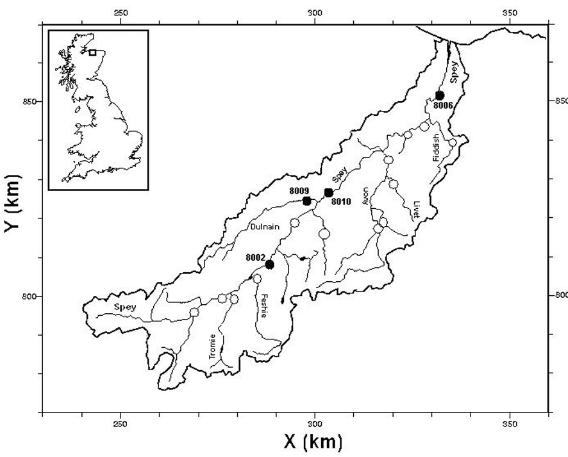

Figure 1. Map of the Spey catchment (Universal Transverse Mercator coordinates). The black circles indicate the four gauge stations of interest (see section 5).

all the stations may represent a too elementary approximation, unable to rule possibly different marginal behaviors. In order to improve the modeling features, further local dependence parameters a_i's are then introduced via the model (12), by exploiting Theorem 1.

[36] Note that our target is not to provide a well-defined MEV model. Instead, our goal is only to introduce and discuss a set of new techniques for constructing MEV models with "richer" dependence structures: the one given by equation (12) will be used as a simple example. As we shall see, the extraparametrized copula may also show asymmetries, i.e., it could also model nonexchangeable variables, whereas the standard (Archimedean) Gumbel copula is necessarily symmetric (i.e., constant for all permutations of the arguments): this may represent an important feature in many practical applications (see, e.g., the discussion by *Grimaldi and Serinaldi* [2006].

[37] Finally, it is important to stress that, potentially, multiparameter models constructed via Theorem 1 might suffer from an identifiability issue. This means that, due to the analytical structure and possible symmetries of the copulas involved, different sets of parameters (say, $\mathbf{a}' \neq \mathbf{a}''$) might yield the same model. However, as claimed by *Casella and Berger* [1990], problems with identifiability can usually be solved by redefining the model, or by adding suitable constraints on the parameters: this latter procedure may be particularly easy in hydrological practice.

[38] In the present case, it is clear that the same model (12) can be constructed via the two different ordered sets of parameters $\{\xi, \chi, \mathbf{a}\}$ and $\{\chi, \xi, \mathbf{1} - \mathbf{a}\}$. However, the identifiability problem is only apparent: in fact, the use of one or the other set simply corresponds to writing the copula $\mathbf{H}$ either as $\mathbf{H}(\mathbf{u}) = \mathbf{A}(\mathbf{u}^{\mathbf{a}}) \cdot \mathbf{B}(\mathbf{u}^{1-\mathbf{a}})$ or as $\mathbf{H}(\mathbf{u}) = \mathbf{B}(\mathbf{u}^{1-\mathbf{a}}) \cdot \mathbf{A}(\mathbf{u}^{\mathbf{a}})$, according to construction shown in equation (10). Evidently, the swap of the copulas **A** and **B** does not give raise to any practical problem, it is purely a mere mathematical argument.

Table 2. Interstation Distances, Nearest Neighbor Stations, and Estimated Percentages of Contemporary Temporal Occurrences of Annual Maxima[a]

Station	S_2	S_6	S_9	S_{10}
S_2	$\mathbf{S_9}$	*61.7*	*19.1*	*24.0*
S_6	28%	$\mathbf{S_{10}}$	*43.6*	*37.9*
S_9	41%	36%	$\mathbf{S_{10}}$	*6.0*
S_{10}	58%	49%	63%	$\mathbf{S_9}$

[a]Italic numbers are interstation distances (in kilometers), bold numbers are labels of the nearest neighbor station, and roman numbers are estimated percentages of contemporary temporal occurrences of annual maxima (see section 5).

5. Case Study

[39] In order to illustrate the use of the construction methods outlined above, we analyze maximum annual flood data (i.e., the largest instantaneous flows observed in each calendar year) collected in the Spey catchment, the seventh largest river in Britain, located in the Northern Highlands of Scotland (see Figure 1). The river raises from the Monadhliath Mountains and flows 157 km northeastward into the North Sea; the catchment covers an area of $\approx$3367 km^2.

[40] The Spey catchment is equipped with a network of 17 flow gauge stations, many of which have records dating back to 1951. The network is managed by the Scottish Environment Protection Agency (2009, http://www.nwl.ac.uk), which provides measurements of peak levels and flows: the certified data used here are extracted from the Flood Estimation Handbook data sets [*Bayliss*, 1999]. Further details can be found in the works by *Gilvear* [2004] and *Black and Fadipe* [2009].

[41] In this study we consider four gauge stations (i.e., a four-dimensional problem) (see Figure 1), three on the mainstream (S_2: Kinrara [8002], S_{10}: Grantown [8010], and S_6: Boat o Brig [8006]), and one on Dulnain tributary (S_9: Balnaan Bridge [8009]). On the one hand, these stations are of hydrological interest; on the other hand, they provide one of the largest multivariate database when considering all the Spey stations: the available observations amount to 37 quadruples of maximum annual floods. Evidently, from a statistical point of view, the sample size is very small for investigating a multivariate problem: unfortunately, this is a typical situation when databases of extremes are considered. However, here the target is not to provide an ultimate flood model, and no design of actual structures is involved. Instead, our goal is only to show, in a relatively simple case, how the techniques outlined above can be used in practice: in other words, this is a methodological paper.

Table 1. Basic Features of the Available Data

Station	Area (km^2)	Mean (m^3/s)	Standard Deviation (m^3/s)	Minimum (m^3/s)	Maximum (m^3/s)
S_2	1011.7	145.8	48.3	80.6	257.3
S_6	2861.2	556.9	275.0	298.2	1597.8
S_9	272.2	105.9	41.9	49.5	245.5
S_{10}	1748.8	244.7	82.3	129.6	508.8

Table 3. Values of the Kendall's τ for All Pairs of the Four Stations S_2, S_6, S_9, and S_{10}[a]

Station	S_2	S_6	S_9	S_{10}
S_2	1	*0.21*	*0.34*	*0.36*
S_6	**0.06** (0.62)	1	*0.39*	*0.29*
S_9	**0.25** (0.03)	**0.34** ($4e^{-3}$)	1	*0.39*
S_{10}	**0.43** ($2e^{-4}$)	**0.29** (0.01)	**0.54** ($3e^{-6}$)	1

[a]Italic numbers are estimates using the extraparametrized Gumbel copula given by equation (12); bold numbers are empirical estimates (with the p values in parentheses).

[42] In Table 1 we report some basic features of the available data: the drainage area, the mean, the standard deviation, and the minimum and maximum observations. In Table 2 we show the interstation distances, and, for each site, the nearest neighbor station. In addition, we also report the estimated percentages $\widehat{p}_{ij}$'s of contemporary temporal occurrences of annual maxima for all the pairs of the four stations (S_2, S_6, S_9, S_{10}): these are simply the empirical probabilities that, for a given pair of stations (i, j), the annual maximum floods occur in the same day (±1 day). Clearly, this provides some interesting information about the physical origin of the stochastic dependence between the observed data. In fact, the analysis of the results yields the following conclusions:

[43] 1. The value $\widehat{p}_{ij}$ is minimum for the pair (S_2, S_6), the two farthest stations.

[44] 2. The value $\widehat{p}_{ij}$ is maximum for the pair (S_9, S_{10}), the two closest stations.

[45] 3. The values $\widehat{p}_{ij}$'s are empirically consistent with the estimates of the Kendall's τ reported in Table 3.

[46] In Figure 2 we show the nonparametric bivariate rank plots [*Genest and Favre*, 2007] of all the six pairs of interest here. These plots provide an empirical estimate of the pairwise dependence structures (i.e., the 2-copulas) ruling the statistical behavior of the multivariate phenomenon. It is worth noting that the data do not fill uniformly the unit square, but rather they tend to arrange along the main diagonal: this suggests that the stations should not be independent, but rather positively dependent, as is required for extreme values. Similar conclusions can be drawn by analyzing the trivariate rank plots of the available data (not shown). It is important to point out that all the previous plots are nonparametric, and are not equivalent to the corresponding bivariate or trivariate plots of the original observations, as is frequently found in many papers: in fact, in the latter case, what is shown is *not* the actual dependence structure (i.e., the copula, the object of our study).

[47] In Table 3 we show the empirical estimates of the bivariate Kendall's τ (and the corresponding p values) for all the pairs of interest here. It is interesting to note that this coefficient is very small for the two farthest stations $\{S_2, S_6\}$: this means that the association between the two is negligible, as confirmed by the corresponding p value, though this does not imply that the stations are statistically independent (as, instead, is commonly misinterpreted). A doubtful result concerns the pair $\{S_2, S_9\}$, involving the most upstream station and the one on the Dulnain tributary: the value $\tau \approx 0.25$ cannot be unquestionably classified as significantly positive, since the p value is about 3%. On the contrary, the analysis of the p values shows that the estimates of the Kendall's coefficients for all the other pairs are statistically significantly different from zero: this means that the corresponding stations are definitely dependent. It is worth pointing out that the analysis of the p values is an important step: it provides an objective way to judge whether or not the estimate of a parameter is statistically significant, and, in turn, to guess which model might be suitable for describing the phenomenon investigated.

[48] A four-variate MEV copula can be used for modeling the dependence between the four stations. From an inferential point of view, the first step in fitting extreme value copulas would be to test whether the underlying dependence structure is indeed of this form. One such test is described by *Ghoudi et al.* [1998] and further studied by *Ben Ghorbal et al.* [2009] [see also *Kojadinovic and Yan*, 2010a; I. Kojadinovic et al., Large-sample tests of extreme-value dependence for multivariate copulas, submitted to *Canadian Journal of Statistics*, 2010]. However, in the present case, since we are dealing with observations of maxima, we use a priori MEV copulas. For the sake of illustration, the two competing MEV models (11) and (12) are fitted on the available data. Again, our target is only to show how to construct (and properly contrast) different copula models, not to promote any of them.

[49] The parameters are calculated via a pseudo-likelihood technique involving the ranks of the data [see *Genest et al.*, 1995; *Shih and Louis*, 1995; *Joe*, 1997; *Genest and Favre*, 2007 and references therein]: this fitting procedure is feasible, since the Gumbel copula is absolutely continuous. An alternative way to fit the parameters would be to use a $(d - 1)$-variate version of Pickands's dependence function, as outlined in some recent works [see *Zhang et al.*, 2008; *Genest and Segers*, 2009]. However, we shall not explore this approach here.

[50] The estimate of the Gumbel parameter θ is $\widehat{\theta} \approx 1.37$, while the estimates for the extraparametrized model are: $\widehat{\xi}$ (global) ≈ 1.55, $\widehat{\chi}$ (global) ≈ 11.04, $\widehat{a}_1$ (local: S_2) ≈ 0.97, $\widehat{a}_2$ (local: S_6) ≈ 0.36, $\widehat{a}_3$ (local: S_9) ≈ 0.78, and $\widehat{a}_4$ (local: S_{10}) ≈ 0.89. It is interesting to note that, given the actual estimates of the Gumbel parameters ξ and χ, it turns out that $\mathbf{G}_\xi \approx \mathbf{\Pi}_4$ (the copula of independence), whereas $\mathbf{G}_\chi \approx \mathbf{M}_4$ (the copula of full dependence). Thus, as already discussed, the extraparametrized copula $\mathbf{H}$ is a sort of "mixture" between $\mathbf{\Pi}_4$ and $\mathbf{M}_4$, ruled by the "local" mixing parameters a_i's.

[51] In Figure 3 we plot the empirical and fitted Pickands' functions A for all the pairs of stations and the two models of interest. It must be stressed that the empirical estimates of the true (but unknown) dependence functions do not generally respect the convexity constraint (the estimator is not intrinsic, i.e., not convex by design), and the estimates do not improve by using different algorithms to calculate the empirical A's (see, e.g., the discussions by *Genest and Segers* [2009 and references therein]). However, the lack of fit, particularly evident for the pairs $\{S_2, S_6\}$, $\{S_2, S_9\}$, and $\{S_6, S_{10}\}$, are more apparent than real: in fact, due to the small sample size, the confidence bands are expected to be quite large.

[52] Note that the Gumbel dependence functions are the same in all plots, whereas the extraparametrized Gumbel ones (hereafter, X-Gumbel) adapt themselves to the "local" behaviors of the data. Visually, the latter model provides

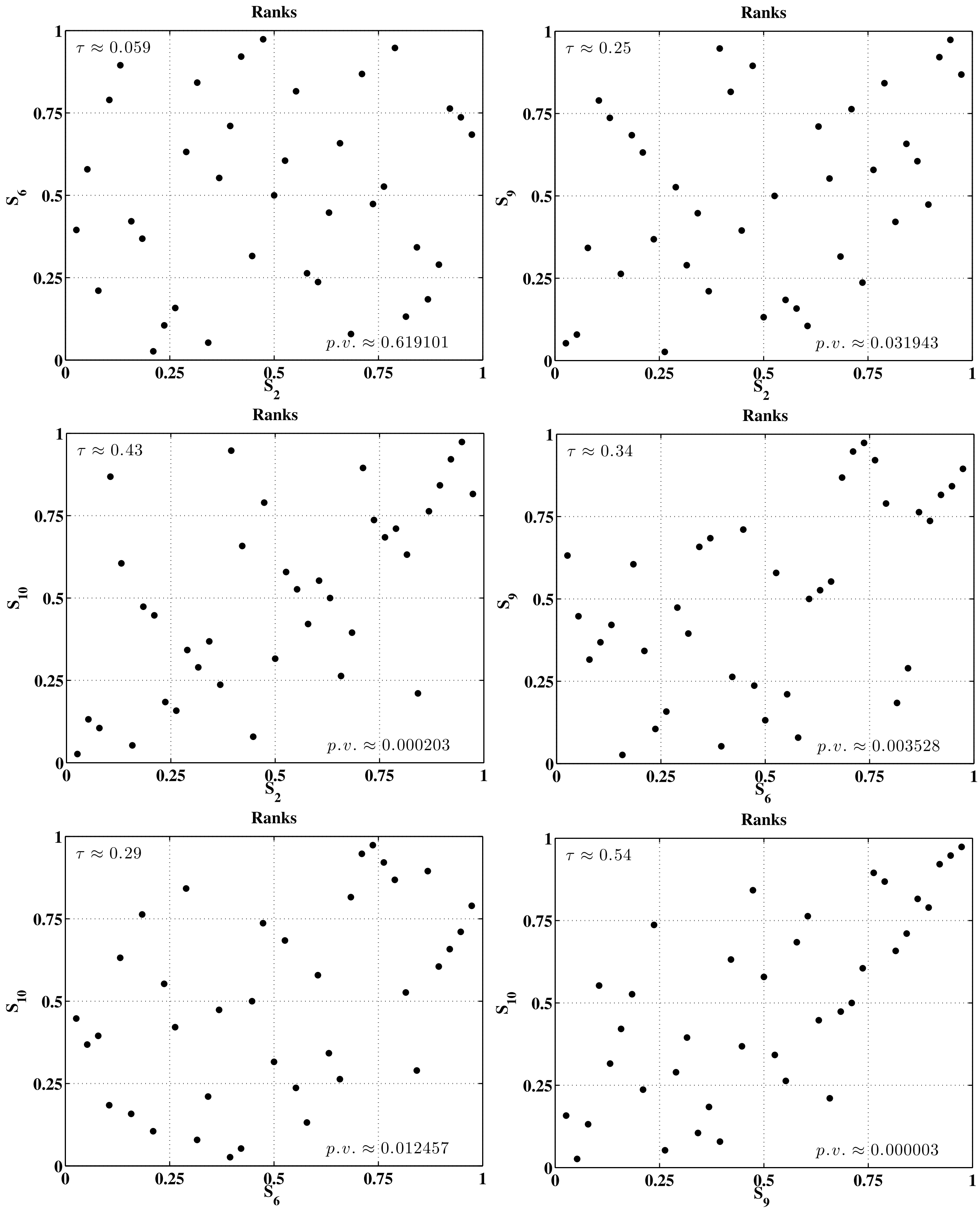

Figure 2. Nonparametric bivariate rank plots of the available data for all the pairs of interest (see section 5). Also shown are the empirical estimates of the Kendall's τ coefficients and the corresponding p values.

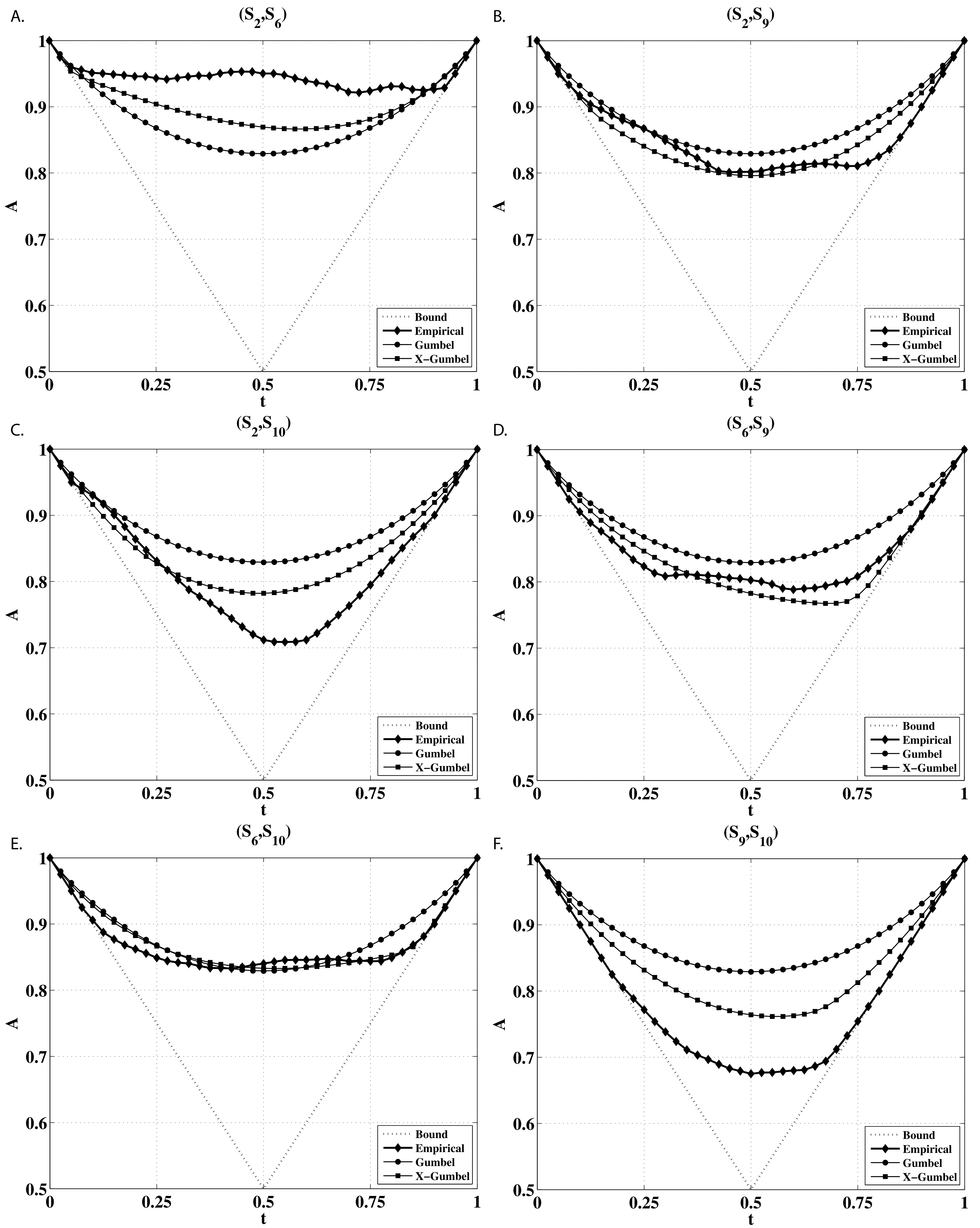

Figure 3. Plots of empirical and fitted Pickands' dependence functions for all the pairs of stations and the two models of interest (see section 5).

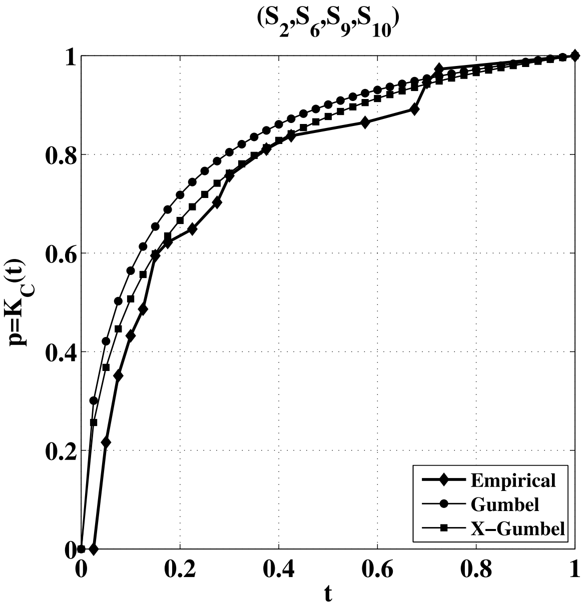

Figure 4. Plot of empirical and fitted Kendall's distribution functions K_C for all of the four stations and the two models of interest (see section 5).

better fits than the Gumbel one, being also able to match the asymmetries shown by the data: this is particularly evident for the pairs $\{S_6, S_9\}$ and $\{S_6, S_{10}\}$. In addition, whereas the Gumbel copula could only model a single "degree of dependence" (i.e., the one associated with the value $\widehat{\tau} = (\widehat{\theta} - 1)/\widehat{\theta} \approx 0.27$ [*Nelsen*, 2006]), the X-Gumbel copula may fit a wider range of dependencies, since the corresponding Kendall's τ's span from ≈0.2 to ≈0.4 (see Table 3).

[53] A common error performed by practitioners is to stop the analysis at this point and, given the investigation of the marginal behaviors and the resulting indications, draw superficial conclusions, and provide a premature guess about which model looks the most suitable. However, rarely the sole analysis of the marginals may help, and yield valuable decisions. In fact, when the problem is multivariate, what should always be analyzed is the full dependence structure and its global ability to fit the empirical data.

[54] For this purpose, we exploit some robust goodness-of-fit tests for multivariate copulas, as recently investigated in the work by *Genest et al.* [2009]. These tests use Cramér-von-Mises statistics, and acceptance or rejection of a model is based on the p values calculated via bootstrap techniques: small ones suggest to discard the corresponding copula, whereas large ones support its suitability. It is important to stress that these tests require some regularity of the copulas investigated, and should be used with care; however, MEV copulas are generally quite regular.

[55] As a result, it turns out that the Gumbel model should be rejected: the p values are always less than 1%! However, this is not too surprising: although some of the bivariate fits may be acceptable, overall the Gumbel copula is only monoparametric, and too rigid to cope with a variegated four-dimensional problem such as the one considered here. On the contrary, the extraparametrized Gumbel model show p values larger than 40%, and is therefore definitely accepted.

[56] It is worth mentioning that the p values should only be used to reject a copula, according to some standard criterion (like, e.g., a value smaller than 1%). It is a common error to consider as "better" those models yielding the highest p values: mathematically speaking, this is generally false.

[57] A further point of interest concerns the investigation of the multivariate Kendall's distribution function $K_{\mathbf{C}}$, used to define the return period for all the four stations, i.e., for the full four-variate problem. As already mentioned in section 2, the calculation of $K_{\mathbf{C}}$ is carried out via simulations: here samples of size $m = 10^6$ are generated from the copulas $\mathbf{G}$ and $\mathbf{H}$ via Algorithm 2, and then Algorithm 1 is used to estimate $K_{\mathbf{G}}$ and $K_{\mathbf{H}}$, respectively.

[58] The results are shown in Figure 4. Apparently, the X-Gumbel copula provides a reasonable fit of the empirical $K_{\mathbf{C}}$, whereas the Gumbel one yields a systematic failure. However, the visual analysis may be deceiving. On the one hand, the calculation of the empirical $K_{\mathbf{C}}$ is affected by the small sample size. On the other hand, the comparison of the empirical and the theoretical Kendall's distribution functions is a sort of Kolmogorov-Smirnov test, and this test is not powerful in a multivariate setting. Indeed, since the Kendall distribution is a univariate representation of some stochastic behavior in a high-dimensional space (here, dimension four), the pictures may be somewhat misleading, and neither the confidence bands nor the corresponding goodness-of-fit test can be expected to be very discriminating [see also *Genest et al.*, 2009]. Overall, the Gumbel model should be discarded according to the Cramér-von-Mises goodness-of-fit tests for multivariate copulas mentioned above, and the investigation of the Kendall's distribution function seems to support such a conclusion.

[59] In terms of the multivariate return periods introduced in section 3, further significant comments can be expressed. To the best of our knowledge, this is the first time that a similar analysis is presented. In Figure 5 we show the empirical and the fitted return periods for all the four stations and the two models of interest: the plot shows the

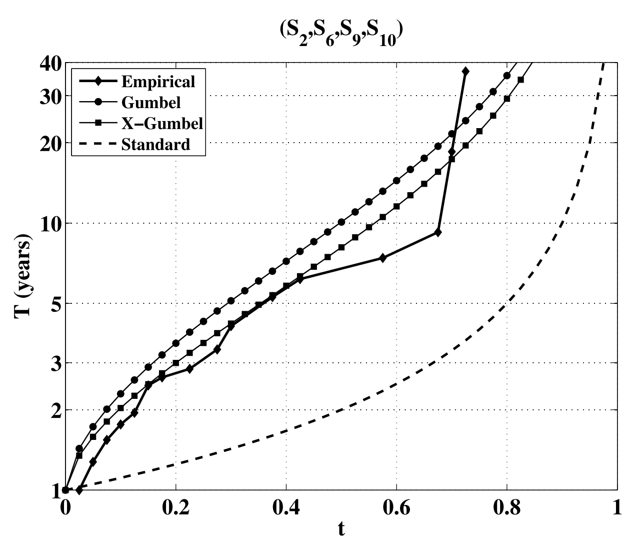

Figure 5. Plot of empirical and fitted return periods for all of the four stations and the two models of interest (see section 5). Also shown are the standard return periods prescribed by equation (6).

return periods associated with all critical probability levels $t \in \mathbf{I}$. Visually, the fit provided by the X-Gumbel model is valuable and consistent, also considering the limited sample size used: in fact, this spoils the estimates of the largest empirical return periods (as is well evident in Figure 5). On the contrary, the standard Gumbel model (the one to be discarded), again shows a systematic lack of fit.

[60] For any given t, the unit cube $\mathbf{I}^4$ is partitioned by the iso-hypersurface $\mathcal{S}_t$ into a subcritical and a supercritical region, the latter one having probability $1 - p = 1 - K_{\mathbf{C}}(t)$, and containing all the events having return period larger than T_p. For the sake of comparison, we also show the corresponding standard return periods T_t's as prescribed by equation (6). Evidently, the difference between the (correct) new definition and the (wrong) standard one is striking. For instance, the supercritical events corresponding to the iso-hypersurface $\mathcal{S}_{0.8}$ are assigned a standard return period equal to 5 years. However, the supercritical region has probability $1 - K_{\mathbf{H}}(0.8) \approx 3.44\%$ (according to the results shown in Figure 4), and therefore a supercritical event should be expected (on average) only once every, say, 29 realizations (i.e., 29 years, since annual maxima are investigated), as correctly estimated by the X-Gumbel model. Even more impressive is the situation for larger thresholds t's (not shown). For instance, the standard return period associated with $t = 0.999$ is 1000 years, whereas the correct one is about 6660 years, a much larger value.

[61] Clearly, the underestimates provided by the standard approach, i.e., a return period (much) smaller than the correct one, may have sizable consequences. Instead, following the Kendall's measure approach, the subcritical and supercritical regions can easily be identified, and a correct risk analysis can be performed.

6. Conclusions

[62] MEV models are fundamental in all areas of geophysics in order to properly assess risk and hazard. This paper is of methodological nature and introduces new techniques for dealing with extremes.

[63] 1. We outline how, exploiting recent theoretical developments in the theory of copulas, new MEV models can be easily constructed. In particular, a straightforward generalization of the standard multivariate logistic Gumbel model is presented.

[64] 2. We show how a suitable number of parameters can be easily introduced in existing models, thus generalizing many traditional MEV models.

[65] 3. We introduce a suitable new definition of multivariate return period and show why the traditional definition presently used in literature fails to provide meaningful results.

[66] An application to flood data is also illustrated. Our target is not to provide an ultimate extreme flood model, and no design of actual structures is involved here. Instead, our goal is only to show and discuss how the techniques outlined in the paper can be used in practice. Throughout the paper we also show and comment some of the common errors performed by unskilled practitioners while carrying out a copula analysis: this should serve as a guideline.

[67] **Acknowledgments.** The authors thank C. Sempi (Università del Salento, Lecce, Italy), F. Durante (Free University of Bozen-Bolzano, Bolzano, Italy), and I. Kojadinovic (University of Auckland, Auckland, New Zealand) for invaluable helpful discussions and suggestions. The research was partially supported by the Italian M.I.U.R. via the project "Metodi stocastici in finanza matematica." The support of Centro Mediterraneo Cambiamenti Climatici (CMCC, Lecce, Italy) is also acknowledged.

References

Alexander, G. (1954), Some aspects of time series in hydrology, *J. Inst. Eng. Aust.*, *26*(9), 194.

Bacchi, B., G. Becciu, and N. Kottegoda (1994), Bivariate exponential model applied to intensities and durations of extreme rainfall, *J. Hydrol.*, *155*, 225–236.

Bayliss, A. (1999), Validation and update of flood peak data, in *Flood Estimation Handbook*, vol. 3, edited by A. Robson and D. Reed, pp. 261–272, Inst. of Hydrol., Wallingford, U. K.

Ben Ghorbal, N., C. Genest, and J. Neslehová (2009), On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence, *Can. J. Stat.*, *37*, 1–19.

Billingsley, P. (1995), *Probability and Measure*, 3rd ed., John Wiley, New York.

Black, A. R., and D. Fadipe (2009), Use of historic water level records for re-assessing flood frequency: Case study of the Spey catchment, *Water Environ. J.*, *23*(1), 23–31.

Capéraà, P., A.-L. Fougères, and C. Genest (1997), A stochastic ordering based on a decomposition of Kendall's tau, in *Distributions with Given Marginals and Moment Problems*, edited by V. Beneš and J. Štěpán, pp. 81–86, Kluwer Acad., Dordrecht, Netherlands.

Casella, G., and R. L. Berger (1990), *Statistical Inference*, Duxbury, Belmont, Calif.

Coles, S., and J. Tawn (1991), Modeling extreme multivariate events, *J. Royal Stat. Soc., Ser. B*, *53*(2), 377–392.

Durante, F., and G. Salvadori (2010), On the construction of multivariate extreme value models via copulas, *Environmetrics*, *21*, 143–161.

Falk, M., and R. D. Reiss (2005), On Pickands coordinates in arbitrary dimensions, *J. Multivariate Anal.*, *92*, 426–453.

Galambos, J. (1987), *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed., Kreiger, Malabar, Fla.

Genest, C., and A. Favre (2007), Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, *12*(4), 347–368.

Genest, C., and L.-P. Rivest (1989), A characterization of Gumbel's family of extreme value distributions, *Stat. Probab. Lett.*, *8*(3), 207–211.

Genest, C., and L.-P. Rivest (1993), Statistical inference procedures for bivariate Archimedean copulas, *J. Am. Stat. Assoc.*, *88*(423), 1034–1043.

Genest, C., and L.-P. Rivest (2001), On the multivariate probability integral transformation, *Stat. Probab. Lett.*, *53*(4), 391–399.

Genest, C., and J. Segers (2009), Rank-based inference for bivariate Extreme Value copulas, *Ann. Stat.*, *37*(5B), 2990–3022.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995), A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, *82*, 543–552.

Genest, C., K. Ghoudi, and L.-P. Rivest (1998), "Understanding relationships using copulas," by Edward Frees and Emiliano Valdez, January 1998, *North Am. Actuarial J.*, *2*(3), 143–149.

Genest, C., J.-F. Quessy, and B. Rémillard (2006), Goodness-of-fit procedures for copula models based on the probability integral transformation, *Scand. J. Stat.*, *33*, 337–366.

Genest, C., B. Rémillard, and D. Beaudoin (2009), Goodness-of-fit tests for copulas: A review and a power study, *Insur. Math. Econ.*, *44*, 199–213.

Ghoudi, K., A. Khoudraji, and L. Rivest (1998), Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles, *Can. J. Stat.*, *26*, 187–197.

Gilvear, D. J. (2004), Patterns of channel adjustment to impoundment of the upper River Spey, Scotland (1942–2000), *River Res. Appl.*, *20*(2), 151–165.

Grenney, W., and E. Heyse (1985), Suspended sediment—River flow analysis, *J. Environ. Eng.*, *111*(6), 790–803.

Grimaldi, S., and F. Serinaldi (2006), Asymmetric copula in multivariate flood frequency analysis, *Adv. Water Resour.*, *29*(8), 1155–1167.

Herr, H., and R. Krzysztofowicz (2005), Generic probability distribution of rainfall in space: The bivariate model, *J. Hydrol.*, *306*, 234–263.

Hosking, J., and J. Wallis (1988), The effect of intersite dependence on regional flood frequency analysis, *Water Resour. Res.*, *24*, 588–600.

Joe, H. (1994), Multivariate extreme-value distributions with applications to environmental data, *Can. J. Stat.*, *22*, 47–64.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

Journel, A., and F. Alabert (1988), Non-gaussian data expansion in the earth sciences, *Terra Nova*, *1*, 123–134.

Katz, R., M. Parlange, and P. Naveau (2000), Statistics of extremes in hydrology, *Adv. Water Resour.*, *25*, 1287–1304.

Keef, C., C. Svensson, and J. Tawn (2009), Spatial dependence in extreme river flows and precipitation for Great Britain, *J. Hydrol.*, *378*, 240–252.

Khoudraji, A. (1995), Contributions à l'étude des copules et à la modélisation des valeurs extrêmes bivariées, Ph.D. thesis, Université de Laval, Québec City, Que., Canada.

Kim, T.-W., J. Valdes, and C. Yoo (2003), Nonparametric approach for estimating return periods of droughts in arid regions, *J. Hydrol. Eng.*, *8*, 237–246.

Kojadinovic, I., and J. Yan (2010a), Nonparametric rank-based tests of bivariate extreme-value dependence, *J. Multivariate Anal.*, *101*(9), 2234–2249.

Kojadinovic, I., and J. Yan (2010b), Modeling multivariate distributions with continuous margins using the copula R package, *J. Stat. Software*, *34*(9), 20.

Kottegoda, N., and L. Natale (1994), Two-component log-normal distribution of irrigation-affected low-flows, *J. Hydrol.*, *158*, 187–199.

Liebscher, E. (2008), Construction of asymmetric multivariate copulas, *J. Multivariate Anal.*, *99*, 2234–2250.

Nelsen, R. (2006), *An Introduction to Copulas*, 2nd ed., Springer, New York.

Pickands, J. (1981), Multivariate extreme value distributions, *Bull. Int. Stat. Inst.*, *49*, 859–878.

Pons, F. (1992), Regional flood frequency analysis based on multivariate lognormal models, Ph.D. thesis, Colo. State Univ., Fort Collins.

Raynal-Villasenor, J., and J. Salas (1987), Multivariate extreme value distributions in hydrological analyses, in *Water for the Future: Hydrology in Perspective*, *IAHS Publ.*, vol. 164, pp. 111–119, IAHS Press, Wallingford, U. K.

Renard, B., and M. Lang (2007), Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology, *Adv. Water Resour.*, *30*(4), 897–912.

Rootzén, H., and N. Tajvidi (2006), Multivariate generalized Pareto distributions, *Bernoulli*, *12*(5), 917–930.

Russo, D. (2009), On probability distribution of hydraulic conductivity in variably saturated bimodal heterogeneous formations, *Vadose Zone J.*, *8*, 611–622.

Salvadori, G. (2004), Bivariate return periods via 2-copulas, *Stat. Methodol.*, *1*, 129–144.

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

Salvadori, G., C. De Michele, N. Kottegoda, and R. Rosso (2007), *Extremes in Nature: An Approach Using Copulas*, *Water Sci. Technol. Libr.*, vol. 56, Springer, Dordrecht, Netherlands.

Shih, J., and T. Louis (1995), Inferences on the association parameter in copula models for bivariate survival data, *Biometrics*, *51*(4), 1384–1399.

Singh, V. (Ed.) (1986), *Hydrologic Frequency Modeling*, D. Reidel, Norwell, Mass.

Singh, V., S. Jain, and A. Tyagi (2007), *Risk and Reliability Analysis*, ASCE Press, Reston, Va.

Sklar, A. (1959), Fonctions de répartition à *n* dimensions et leurs marges, *Publ. Inst. Stat. Univ. Paris*, *8*, 229–231.

Stedinger, J. (1983), Estimating a regional flood frequency distribution, *Water Resour. Res.*, *19*, 503–510.

Tawn, J. (1990), Modelling multivariate extreme value distributions, *Biometrika*, *77*(2), 245–253.

Wilks, D. (1998), Multisite generalization of a daily stochastic precipitation generation model, *J. Hydrol.*, *210*, 178–191.

Yan, J., and I. Kojadinovic (2009), *Copula: Multivariate dependence with copulas, R package*, version 0.8–12, http://cran.r-project.org/web/packages/copula/index.html, Compr. R Arch. Network, Inst. for Stat. and Math., Vienna Univ. of Econ. and Bus., Vienna.

Yue, S. (2000a), The Gumbel mixed model applied to storm frequency analysis, *Water Resour. Manage.*, *14*, 377–389.

Yue, S. (2000b), The Gumbel logistic model for representing a multivariate storm event, *Adv. Water Resour.*, *24*, 179–185.

Zhang, D., M. Wells, and L. Peng (2008), Nonparametric estimation of the dependence function for a multivariate extreme value distribution, *J. Multivariate Anal.*, *99*(4), 577–588.

Zhang, L. (2005), Multivariate hydrological frequency analysis and risk mapping, Ph.D. thesis, La. State Univ., Baton Rouge.

Zhang, L., and V. Singh (2007), Gumbel-Hougaard copula for trivariate rainfall frequency analysis, *J. Hydrol. Eng.*, *12*, 409–419.

C. De Michele, DIIAR (Sezione CIMI), Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20132 Milan, Italy. (carlo.demichele@polimi.it)

G. Salvadori, Dipartimento di Matematica, Università del Salento, P. O. Box 193, I-73100 Lecce, Italy. (gianfausto.salvadori@unisalento.it)