Hydrol. Earth Syst. Sci., 15, 141–150, 2011
www.hydrol-earth-syst-sci.net/15/141/2011/
doi:10.5194/hess-15-141-2011

Hydrology and Earth System Sciences

Estimating strategies for multiparameter Multivariate Extreme Value copulas

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

[1]Dipartimento di Matematica, Università del Salento, Provinciale Lecce-Arnesano, P.O. Box 193, 73100 Lecce, Italy
[2]Sezione CIMI, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20132 Milano, Italy

Received: 19 July 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 4 October 2010
Revised: 21 December 2010 – Accepted: 22 December 2010 – Published: 17 January 2011

Abstract. Multivariate Extreme Value models are a fundamental tool in order to assess potentially dangerous events. Exploiting recent theoretical developments in the theory of Copulas, new multiparameter models can be easily constructed. In this paper we suggest several strategies in order to estimate the parameters of the selected copula, according to different criteria: these may use a single station approach, or a cluster strategy, or exploit all the pair-wise relationships between the available gauge stations. An application to flood data is also illustrated and discussed.

1 Introduction

Multivariate extremes occur in several hydrologic problems (like, e.g., space-time precipitation and floods (Singh, 1986; Pons, 1992; Wilks, 1998; Kim et al., 2003; Herr and Krzysztofowicz, 2005; Keef et al., 2009), or hydraulic conductivity in porous media – Journel and Alabert, 1988; Russo, 2009), as well as in many environmental problems (like, e.g., water quality and pollution (Grenney and Heyse, 1985), or sea levels – Butler et al., 2007).

The investigation of multivariate phenomena is best carried out via copulas. 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 described in a straightforward way in terms of suitable copulas. For a thorough bibliography see Nelsen, 2006; Salvadori et al., 2007.

The problem of measuring the amount of dependence between the variables involved is a central issue when modeling multivariate extremes. For instance, in literature the pair-wise dependence is generally measured via the canonical Pearson's correlation coefficient. However, it may not be the best measure of dependence when dealing with extremes (Joe, 1997), since it does not exist for heavy-tailed variables with infinite variance, and only involves a linear kind of dependence. Recently, other quantities were considered (Nelsen, 2006) to measure the association between pairs of random variables (hereafter, r.v.s): among others, Kendall's τ and Spearman's ρ rank correlation coefficients, or the Blomqvist's β medial correlation coefficient. These measures always exist (being based on the ranks), and model several types of association (for a practical discussion see, e.g., the case studies illustrated in Salvadori et al., 2007).

Instead, the notion of cluster-type dependence, when the size of the cluster is larger than two (i.e., beyond the simple pair-wise case), has only been partially explored. Generalizations of Kendall's τ (Nelsen, 1996), Spearman's ρ (Schmid and Schmidt, 2007a,b), and Blomqvist's β (Durante et al., 2007; Schmid and Schmidt, 2007c) to the d-variate case ($d > 2$) were only recently introduced – see below. These extensions may be of practical importance: on the one hand, they provide useful tools to quantify the dependence within clusters; on the other hand, they can be used to estimate the parameters of the multivariate model at play (see later). However, at present the application of these measures in actual case studies is still quite limited.

Another important issue is represented by the construction of Multivariate Extreme Value (hereafter, MEV) models involving a significant number of parameters. Using the results of Liebscher (2008), recent works (Durante and Salvadori, 2010; Salvadori and De Michele, 2010) have shown how multiparameter MEV models can be easily constructed via copulas and suitable techniques of extra-parameterization, leading either to the formulation of new models, or to the generalization of existing ones.

A further fundamental question is represented by the estimate of the parameters of the multivariate copulas considered

Correspondence to: G. Salvadori
(gianfausto.salvadori@unisalento.it)

(see Genest et al., 1995; Shih and Louis, 1995; Joe, 1997; Genest and Favre, 2007, and references therein). Maximum Likelihood (hereafter, ML) or Pseudo-likelihood procedures involving the ranks of the data are generally used to fit these parameters. Alternatively, the parameters may be sometimes estimated via the Method of Moments and some pair-wise measures of association (usually, the Kendall's τ, the Spearman's ρ, or the Blomqvist's β). Apparently, no application of the d-variate generalizations of these measures to the parameters' estimation is available in literature.

In this paper we focus the attention on the estimation of the parameters in copula-based MEV models, presenting some new fitting strategies. In particular, each procedure exploits a different source of information: (i) a suitable single station, (ii) an appropriate cluster of stations, (iii) all the pairs of the available stations. Below, in Sect. 2 we introduce the concept of multivariate Extreme Value copulas, describing some of the mathematical features of interest here. In Sect. 3 we show several strategies for estimating the relevant parameters. In Sect. 4 an application to maximum annual flood data is presented and discussed.

2 MEV copulas: an overview

In this Section we briefly outline the mathematics of copulas needed in the sequel; for a thorough theoretical and practical introduction see, respectively, 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 $\mathbf{I}^d$ the d-dimensional unit cube.

The main target pursued here is to provide a general multivariate framework for modeling non-independent extreme 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 r.v.s used in the sequel may represent, for instance, rainfall or flood measurements collected in a given basin, or pollution samples in a region, or wave measurements collected by marine buoys. Below, $\mathcal{S} = \{S_1, \ldots, S_d\}$ will denote a set of d gauge stations.

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

$$F(x_1, \ldots, x_d) = \mathbf{C}(F_1(x_1), \ldots, F_{\mathrm{d}}(x_d)). \tag{1}$$

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

A copula $\mathbf{C}$ is MEV if it is max-stable, i.e. if it satisfies the equation

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

for all $\mathbf{u} \in \mathbf{I}^d$ and all $t > 0$. As a simple example consider the following two copulas:

$$\Pi_d(\mathbf{u}) = u_1 \cdots u_d, \tag{3}$$

$$\mathbf{M}_d(\mathbf{u}) = \min\{u_1, \ldots, u_d\}. \tag{4}$$

The former one models independent variates, while the latter one models comonotone dependent ones, where each variable is a monotone increasing function of the others. Evidently, both Π_d and $\mathbf{M}_d$ are max-stable, and hence MEV. A distribution F is MEV if, and only if, all its margins F_i's are Generalized Extreme Value laws (hereafter, GEV), and the corresponding copula $\mathbf{C}$ is MEV. Note that not all copulas are MEV (i.e., satisfy the max-stability property (2)), and consequently should not be used to construct consistent MEV models. In addition, since the GEV law is continuous, the representation $F = \mathbf{C}(F_1, \ldots, F_d)$ of a MEV distribution F is unique. Most importantly, by exploiting the invariance property of copulas (Nelsen, 2006), we may restrict our attention to copulas only, and do not worry about the GEV margins, as we shall do hereinafter.

The construction of multivariate measures of association and/or dependence is an involved mathematical problem, and is still an open question in statistics. Several ideas were developed in the last few years, and various measures were introduced in order to describe concepts like, e.g., concordance for random vectors (Joe, 1990; Nelsen, 1996, 2002; Úbeda-Flores, 2005; Schmid and Schmidt, 2007a; Taylor, 2007).

For bivariate problems, several measures of association are available (Joe, 1997; Nelsen, 2006). Among others, Kendall's τ and Spearman's ρ are frequently used in applications. The former one is the difference between the probability of concordance and discordance of the variables, the latter one measures the average distance between Π_2 (i.e., independence) and the bivariate copula of interest. As is well known, these measures only depend upon the copula joining the variables under investigation, and not upon the margins (i.e., they are scale invariant). As already mentioned above, a further advantage is that, if the variables involved are characterized by heavy-tailed distributions, then the second order moment (and, in turn, Pearson's coefficient) may not exist, whereas these latter measures always exist, being based on the ranks.

Interesting extensions of Kendall's τ (Nelsen, 1996) and Spearman's ρ (Schmid and Schmidt, 2007a,b) to a general d-variate framework ($d > 2$) were recently proposed, and several new measures involving the generic d-copula $\mathbf{C}$ were introduced:

$$\tau_d = \frac{1}{2^{d-1} - 1}\left(2^d \int_{\mathbf{I}^d} \mathbf{C}(\mathbf{u}) d\mathbf{C}(\mathbf{u}) - 1\right), \tag{5}$$

$$\rho_{d,1} = h(d)\left(2^d \int_{\mathbf{I}^d} \mathbf{C}(\mathbf{u})\, d\mathbf{u} - 1\right), \tag{6}$$

$$\rho_{d,2} = h(d)\left(2^d \int_{\mathbf{I}^d} \Pi_d(\mathbf{u})\, d\mathbf{C}(\mathbf{u}) - 1\right), \tag{7}$$

$$\rho_{d,3} = h(2)\left(2^2 \sum_{i<j} \binom{d}{2}^{-1} \int_{\mathbf{I}^2} \mathbf{C}_{ij}(u,v)\, du\, dv - 1\right), \tag{8}$$

where $h(d) = (d+1)/(2^d-(d+1))$, and $\mathbf{C}_{ij}$ is the bivariate (i,j)-margin of $\mathbf{C}$. Note that $\rho_{d,3}$ is essentially the average Spearman's ρ for all the pairs in a set of d variables.

Another useful multivariate measure of association is the medial correlation coefficient β_d (see Durante et al., 2007; Schmid and Schmidt, 2007c, and references therein), which generalizes the well known Blomqvist's β coefficient (Nelsen, 2006):

$$\beta_d = \frac{2^{d-1}\left(\mathbf{C}(\mathbf{1/2}) + \overline{\mathbf{C}}(\mathbf{1/2})\right) - 1}{2^{d-1}-1}, \tag{9}$$

where $\overline{\mathbf{C}}$ is the survival function associated with $\mathbf{C}$, given by $\overline{\mathbf{C}}(\mathbf{u}) = \mathbf{P}\{\mathbf{C} > \mathbf{u}\}$, and $\mathbf{1/2} = (1/2, \ldots, 1/2)$. Clearly, also β_d is invariant with respect to the distributions of the margins. As pointed out in Schmid and Schmidt (2007c), β_d has some advantages over competing measures such as τ_d or $\rho_{d,i}$'s. In fact, it can explicitly be derived whenever the copula is of explicit form, which is often not possible for other measures, and its estimation requires a low computational complexity. Thus, β_d may represent a fast alternative for estimating the copula parameters (see below).

A further notion of interest is represented by Pickands' dependence function A (Pickands, 1981). Recall that 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

$$\mathbf{C}(u,v) = \exp\left[\ln(uv)\, A\left(\frac{\ln v}{\ln(uv)}\right)\right] \tag{10}$$

for all $(u,v) \in \mathbf{I}^2$. In particular, if $A(t) \equiv 1$ then $\mathbf{C} = \Pi_2$, and if $A(t) = \max\{t, 1-t\}$ then $\mathbf{C} = \mathbf{M}_2$. Conversely, given a bivariate MEV copula $\mathbf{C}$, the corresponding dependence function A is given by

$$A(t) = -\ln \mathbf{C}\left(e^{-(1-t)}, e^{-t}\right), \tag{11}$$

where $t \in \mathbf{I}$. It is worth noting that the value $\tau_{\mathbf{C}}$ of the Kendall's τ associated with $\mathbf{C}$, as well as the one of the Spearman's ρ, can be expressed in terms of A via (Nelsen, 2006; Salvadori et al., 2007)

$$\tau_{\mathbf{C}} = \int_0^1 \frac{t(1-t)}{A(t)}\, dA'(t) \tag{12}$$

and

$$\rho_{\mathbf{C}} = 12 \int_0^1 \frac{1}{(1+A(t))^2}\, dt - 3. \tag{13}$$

A generalization of Pickands' dependence function to the multivariate case is shown in Falk and Reiss (2005). 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.

Finally, below we shall also use the Kendall's measure function $K_{\mathbf{C}}$ (Genest and Rivest, 1993, 2001) given by

$$K_{\mathbf{C}}(t) = \mathbf{P}\{W \le t\} = \mathbf{P}\{\mathbf{C}(U_1, \ldots, U_d) \le t\}, \tag{14}$$

where $t \in \mathbf{I}$ is a probability level, $W = \mathbf{C}(U_1, \ldots, U_{\mathrm{d}})$ 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 $\mathbf{C}$. In the bivariate Extreme Value case, $K_{\mathbf{C}}$ is given by (Ghoudi et al., 1998)

$$K_{\mathbf{C}}(t) = t - (1-\tau_{\mathbf{C}})\, t \ln t, \tag{15}$$

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 Eq. (15) are known for the general multivariate case $d > 2$.

The Kendall's measure $K_{\mathbf{C}}$ is a fundamental tool for introducing a mathematically consistent (copula-based) definition of the return period for multivariate events (see also the discussion in Salvadori, 2004; Salvadori and De Michele, 2004; Salvadori et al., 2007; Durante and Salvadori, 2010; Salvadori and De Michele, 2010). In fact, Eq. (14) represents a multivariate quantile relationship, since it corresponds to a multidimensional Probability Integral Transform (Genest et al., 2006).

Let μ be the average interarrival time of the events in the sequence observed (e.g., $\mu = 1$ year for annual maxima), and let $p \in \mathbf{I}$ be an arbitrary critical probability level (usually, $p = 90, 95, 99\%$, or any other threshold of interest). The multivariate return period T_p associated with p (hereafter, Kendall's return period) is defined as

$$T_p = \frac{\mu}{1-p} = \frac{\mu}{1-K_{\mathbf{C}}(t)} = \frac{\mu}{1-\mathbf{P}\{\mathbf{u} \in \mathbf{I}^d : \mathbf{C}(\mathbf{u}) \le t\}}, \tag{16}$$

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

$$t = \inf\{s \in \mathbf{I} : K_{\mathbf{C}}(s) = p\} = K_{\mathbf{C}}^{[-1]}(p), \tag{17}$$

by analogy with the correct definition of quantile. Here $K_{\mathbf{C}}^{[-1]}$ indicates the generalized (or pseudo-) inverse (Nelsen (2006)) of the corresponding function. Since $K_{\mathbf{C}}$ is generally non-linear ($K_{\mathbf{C}}(t) = t$ only if $\mathbf{C} = \mathbf{M}_{\mathrm{d}}$), then $t \neq p$. More particularly, the relation $K_{\mathbf{C}}(t) \ge t$ holds (Capéraà et al. (1997)), and therefore

$$T_p = T_{K_{\mathbf{C}}(t)} = \frac{\mu}{1-K_{\mathbf{C}}(t)} \ge \frac{\mu}{1-t} = \frac{\mu}{1-\mathbf{C}(\mathbf{u})}, \tag{18}$$

where $\mathbf{u} \in \mathbf{I}^d$ is such that $\mathbf{C}(\mathbf{u}) = t$. The right-most term corresponds to the standard definition of multivariate return period (for a thorough review see Zhang (2005); Singh et al.

(2007), and references therein). Evidently, the traditional approach may yield an incorrect calculation of the return period, and, in turn, a wrong estimation of the risk. Since empirical estimators of the Kendall's measure function are available (Genest et al., 2009; Salvadori and De Michele, 2010), we shall see later how to use them to perform a return period analyses of practical utility.

3 Parameters' estimation

As is well known, the estimate of the parameters of multivariate distributions is an involved problem, and still an open question in mathematical statistics. Usually, procedures like Maximum Likelihood are used to simultaneously fit all the parameters of interest. However, if, e.g., the copulas under investigation have singular components, then ML may be difficult to implement and use.

Below, we outline several approaches for estimating the parameters of interest: each procedure will exploit different sources of information, and estimations achieved via different techniques will generally differ from one another. For instance, the estimate may rely only upon the information drawn from a suitable single station (Sect. 3.1), or an appropriate cluster of stations (Sect. 3.2), or the set of pairs of all the stations (Sect. 3.3). The methods are general, and can be applied to any MEV copula, including those with singular components. Clearly, other approaches are possible, depending upon the specific needs. Note that the estimates calculated via the methods mentioned above could be used as starting guesses for running other procedures (e.g., ML).

Generally, in the strategies presented below, the fitting criterion is represented by the best agreement, in the Least Squares sense (hereafter, LS), with the "local" dependence structures or association measures: clearly, this may yield estimates different from the ones achieved via other procedures (e.g., the global ML). However, the overall fitting ability will always be certified via global Goodness-of-Fit tests (see Sect. 4), in order to verify whether the resulting parametric model could be accepted or not.

3.1 The single station approach

The first approach we propose for the estimate of the parameters of interest consists in using the information drawn from a single station at a time. Practically, for each of the available gauge stations S_i's, a suitable "companion" station $S_j = S_{j(i)}$ is identified, possibly according to specific physio-geomorphological conditions and/or hydrological constraints. Then, we may estimate the parameters via a LS fit, involving the empirical estimates of the Pickands' dependence functions A_{ij}'s of the companion pairs. As an alternative, also the Kendall's measure function $K_{\mathbf{C}}$ could be used. However, while the former is specific for any copula, the latter is not, for it only depends upon the corresponding value of the Kendall's τ – see Eq. (15), and the comment following it. Therefore, we suggest to use Pickands' representation.

The procedure is as follows. Let n be the sample size, i.e. the number of available d-dimensional observations. For each station S_i, $i = 1,\dots,d$, a companion station $S_j = S_{j(i)}$ is identified, and an estimate $\widehat{A}_{ij}$ of the dependence function A of the model under investigation is calculated (Genest and Segers, 2009). In particular, in order to use all the information, since only n bivariate pairs are available, and given the constraints $A(0) = A(1) = 1$, the unit interval $\mathbf{I}$ is partitioned into n uniformly spaced intervals via the set of abscissas $x_k = k/n$, $k = 0,\dots,n$ (clearly, other choices are possible). Then, the $\widehat{A}_{ij}(x_{\mathrm{k}})$'s are estimated over the given grid, and the LS objective function

$$Z^{(1)} = \sum_{i=1}^{d} Z_i^{(1)} = \sum_{i=1}^{d} \sum_{\mathrm{k}=1}^{n-1} \left| A_{i,j(i)}(x_{\mathrm{k}}) - \widehat{A}_{i,j(i)}(x_{\mathrm{k}}) \right|^2 \qquad (19)$$

is minimized, yielding the LS estimates of the parameters of interest.

Note that, if $\{S_i, S_{j(i)}\}$ forms a pair-cluster (i.e., $S_{j(i)}$ is the station companion to S_i, and vice-versa), then there is no need to compute also the symmetric contribution $\{S_{i'=j(i)}, S_{j(i')=i}\}$: this may reduce the computational burden. Essentially, the single station approach (hereafter, 1-MEV) exploits the relationships of the S_i's with the corresponding companion stations, i.e. it only uses the local (station based) bivariate dependence structures.

A natural criterion for selecting the companion station would be the use of the Euclidean distance: then $S_{j(i)}$ would simply be the station closest to S_i. Note that, except for mathematically "pathological" cases of no interest here, usually S_j is unique: a counter-example is given by a (practically improbable) situation in which several stations are exactly positioned on a circle centered in S_i. Denoting by Δ_{ij} the distance between S_i and S_j, only two things may happen:

1. either $\{S_i, S_j\}$ forms a pair-cluster,
2. or, there exists another station S_{k} closer to S_j than S_i; clearly, S_{k} may belong to a pair-cluster.

From a geometrical point of view, at least a couple of stations must form a pair-cluster. In fact, the set of $N_d = d(d-1)/2$ pair distances Δ_{ij}'s is finite, and hence it has (at least) a minimum: this corresponds to a pair-cluster.

We stress that the use of the Euclidean distance as a criterion for choosing the source of information (i.e., adopting a *nearest neighbor* principle) may not always be the most advisable strategy. In fact, it has been shown (see, e.g., GREHYS, 1996; St-Hilaire et al., 2003; Merz and Bloeschl, 2004; Galéa and Canali, 2005; Wagener and Wheater, 2006; Ouarda et al., 2008; Shu and Ouarda, 2008) that the geometrical distance may not completely explain the dependence structure of the hydrological behavior of catchments: indeed,

several are the physio-geomorphological factors that may influence it. Therefore, the validity of the nearest neighbor approach should be tested out by carefully checking the practical case study under investigation.

It is worth pointing out that, if the model involves global parameters (i.e., common to all stations), and these can be estimated a priori via other techniques, then the local parameters (if any) can be calculated as follows. For each station S_i, the companion station S_j is identified, and the local parameters are estimated via a LS fit of the dependence function A_{ij}, using the values of the global parameters already estimated (i.e., only $Z_i^{(1)}$ is minimized). If $\{S_i, S_{j(i)}\}$ is a pair-cluster, then all the estimates of the local parameters associated with S_i and S_j are kept; otherwise, only those associated with S_i are stored. This latter strategy can easily deal with sets of stations of any size: in fact, only two stations at a time are considered for estimating the local parameters. In other words, a global estimate is necessary only if the global parameters cannot be estimated otherwise.

3.2 The cluster approach

The 1-MEV approach adopted in the previous Section only exploited the information drawn by a single station. This strategy can be generalized: in fact, a full cluster of companion gauge stations (instead of just one) may be chosen as a source of information. Clearly, the cluster can be fixed according to specific physio-geomorphological conditions and/or hydrological constraints (e.g., by identifying a homogeneous region, or a basin of influence).

Let S_i be the i-th station, and let $\mathcal{C}_{m_i}^{(\mathrm{i})}$ be a cluster of m_i stations "pertinent" to S_i, with $1 \le m_i < d$. Clearly, the choice of m_i, as well as the selection of the set of relevant companion stations belonging to $\mathcal{C}_{m_i}^{(\mathrm{i})}$, can be made dependent upon specific basin characteristics, and changed when considering different stations S_i's. Evidently, the case $m_i = 1$ for all i's corresponds to the 1-MEV approach. Here the idea is to estimate the parameters by exploiting suitable multivariate measures of association $\phi_{\mathbf{C}}^{(\mathrm{i})}$ calculated over the families of stations $\mathcal{F}_i = \{S_i \cup \mathcal{C}_{m_i}^{(\mathrm{i})}\}$, with $i = 1, \ldots, d$. For instance, any of the five measures outlined in Eqs. (5)–(9) could be used.

The procedure is as follows. First, for each station S_i, an estimate $\widehat{\phi}_i$ of $\phi_{\mathbf{C}}^{(i)}$ is calculated over the cluster $\mathcal{F}_i$. Then, the LS objective function

$$Z^{(\mathcal{F})} = \sum_{i=1}^{d} Z_i^{(\mathcal{F})} = \sum_{i=1}^{d} \left| \phi_{\mathbf{C}}^{(i)} - \widehat{\phi}_i \right|^2 \tag{20}$$

is minimized, yielding the LS estimates of the parameters of interest. Note that, if $\mathcal{F}_i$ is a closed cluster (i.e., if the station $S_j \in \mathcal{C}_{m_i}^{(i)}$, then $\mathcal{F}_j = \mathcal{F}_i$), then the contribution of the cluster can be calculated only once: this may reduce the computational burden. Essentially, the cluster approach (hereafter, c-MEV) exploits the relationships of the S_i's with suitable cluster of stations, i.e. it is based on the local m_i-variate association structures.

The c-MEV strategy is quite flexible, and potentially most promising. Unfortunately, its efficacy may be limited by the current lack of easily usable mathematical tools, which at present may turn it into a "weak" approach. In fact, the use of measures of association $\phi_{\mathbf{C}}^{(i)}$'s for ruling the fits (instead of full dependence structures, as in the 1-MEV approach) may discard some important details: roughly speaking, a few "moments" of a distribution may not provide the same information as of the distribution itself. As an obvious alternative, we might suggest to use in the fits some multivariate equivalents of the bivariate Pickands' dependence functions (Falk and Reiss, 2005), but the research in this area is still in its infancy, and it is not yet clear how to proceed.

Again, it is worth pointing out that, if the model involves global parameters, and these can be estimated a priori via other techniques, then the local parameters (if any) can be calculated as follows. For each station S_i, the family $\mathcal{F}_i$ is identified, and the local parameters are estimated via a LS fit of $\phi_{\mathbf{C}}^{(i)}$, using the values of the global parameters already estimated. If $\mathcal{F}_i$ is a closed cluster, then all the estimates of the local parameters associated with the stations in $\mathcal{F}_i$ are kept; otherwise, only those associated with S_i are stored. Thus, a global estimate is necessary only if the global parameters cannot be estimated otherwise.

3.3 The all-pairs approach

A further approach to the estimate of the parameters may rely upon the use of all the $d(d-1)/2$ bivariate margins, by simultaneously considering the dependence structures of all the pairs of stations. The strategy is to fix all the parameters in such a way that the Pickands' dependence functions A_{ij}'s best fit (in the LS sense) the corresponding empirical ones. From a practical point of view, this approach provides the closest "bivariate" approximation to a global fit: mathematically speaking, it is a "combinatorial" strategy. Clearly, as d gets larger and larger, the calculations required may become computationally demanding.

The LS objective function to be minimized is given by

$$Z^{(p)} = \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \sum_{\mathrm{k}=1}^{\mathrm{n}-1} \left| A_{i,j}(x_{\mathrm{k}}) - \widehat{A}_{i,j}(x_{\mathrm{k}}) \right|^2, \tag{21}$$

yielding the LS estimates of the parameters of interest. We call this method p-MEV approach.

By exploiting the same strategy, a faster alternative would be to calculate the parameters by simultaneously fitting all the bivariate Kendall's τ, or Spearman's ρ, or Blomqvist's β coefficients (or any other measure of association): essentially, this corresponds to a Method of Moments procedure. However, while the use of Pickands' function involves the full functional form of the dependence structure (which is

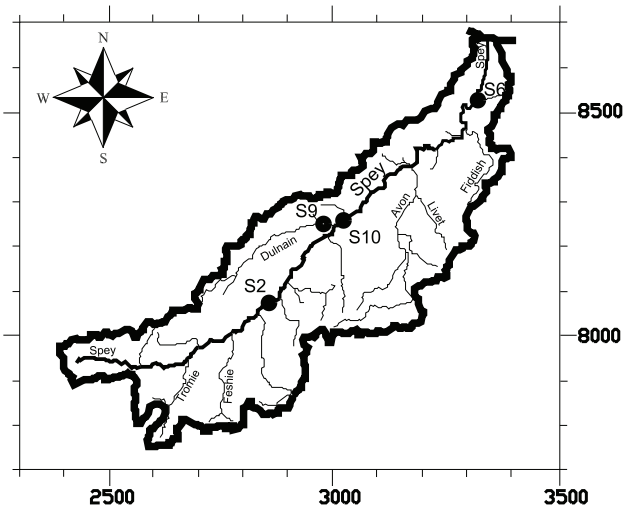

Fig. 1. Map of the Spey catchment. The black circles indicate the four gauge stations of interest – see text.

specific for every copula), the coefficients mentioned above may not distinguish between different copulas. For this reason, we neither suggest nor investigate this alternative.

4 Case study

For the sake of illustration, here we consider the same data and copulas as used in Salvadori and De Michele (2010), to which we make reference for further details: a short summary is reported below. Also, in the following, the use of the Euclidean distance as a criterion for choosing the stations of interest is motivated by illustrative purposes only.

The data are maximum annual flood measurements collected in the Spey catchment (northern Highlands of Scotland). The basin is equipped with a network of 17 flow gauge stations, and is managed by the Scottish Environment Protection Agency (2009). Further details can be found in Gilvear (2004) and Black and Fadipe (2009). In this study we consider four gauge stations located in the middle and lower part of the Spey catchment (see Fig. 1): three on the main stream (i.e., S_2, S_{10}, and S_6), and one on Dulnain tributary (i.e., S_9).

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 extreme data bases are considered. However, here the target is not to provide an ultimate extreme flood model, and no practical project of hydrological works is undertaken. Instead, our point 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.

As a dependence model, here we use the multiparameter 4-variate MEV copula **H** introduced in Salvadori and De Michele (2010):

Table 1. (*Upper triangular*) Inter-station distances (in km). (*Diagonal*) Labels of the nearest neighbor station. (*Lower triangular*) Empirical estimates of the Kendall's τ for all the pairs of the four stations, with the p-values in parentheses – see text.

Station	S_2	S_6	S_9	S_{10}
S_2	S_9	61.7	19.1	24.0
S_6	0.06	S_{10}	43.6	37.9
	(0.62)			
S_9	0.25	0.34	S_{10}	6.0
	(0.03)	(4e-3)		
S_{10}	0.43	0.29	0.54	S_9
	(2e-4)	(0.01)	(3e-6)	

$$\begin{aligned}\mathbf{H}(\mathbf{u}) &= \mathbf{G}_\xi(\mathbf{u}^{\mathbf{a}}) \times \mathbf{G}_\chi(\mathbf{u}^{\mathbf{1}-\mathbf{a}}) \\ &= \mathbf{G}_\xi(u_1^{a_1}, u_2^{a_2}, u_3^{a_3}, u_4^{a_4}) \\ &\quad \times \mathbf{G}_\chi(u_1^{1-a_1}, u_2^{1-a_2}, u_3^{1-a_3}, u_4^{1-a_4}),\end{aligned} \tag{22}$$

with Gumbel parameters $\xi, \chi \geq 1$, and "extra-parameters" $a_1, a_2, a_3, a_4 \in \mathbf{I}$, which represents a 4-variate generalization of the well known Gumbel copula $\mathbf{G}_\theta$ (Nelsen (2006); Salvadori et al. (2007))

$$\mathbf{G}_\theta(\mathbf{u}) = \exp\left\{-\left[\sum_{i=1}^{4}(-\ln u_i)^\theta\right]^{1/\theta}\right\}, \tag{23}$$

with parameter $\theta \geq 1$. Note that the Gumbel copula $\mathbf{G}_\theta$ represents a sort of "standard" MEV model in hydrology (see, e.g., Yue (2000a,b); Zhang and Singh (2007), and references therein). A straightforward interpretation of the parameters a_i's is as follows. Suppose that $\mathbf{a} = \mathbf{1}$: then, $\mathbf{H} = \mathbf{G}_\xi$. Conversely, should it be $\mathbf{a} = \mathbf{0}$, then $\mathbf{H} = \mathbf{G}_\chi$. For other values of $\mathbf{a}$, $\mathbf{H}$ is a sort of "mixture" between $\mathbf{G}_\xi$ and $\mathbf{G}_\chi$: in particular, the a_i's play the role of "local" mixing parameters.

The generic bivariate dependence function A_{ij} of $\mathbf{H}$ is

$$\begin{aligned}A_{ij}(t) = &\left\{[(1-a_i)(1-t)]^\chi + \left[(1-a_j)t\right]^\chi\right\}^{1/\chi} \\ &+ \left\{[a_i(1-t)]^\xi + \left[a_j t\right]^\xi\right\}^{1/\xi},\end{aligned} \tag{24}$$

i.e. a non-linear, possibly asymmetric, function, able to model non-exchangeable variables (an important issue in applications, not shared by $\mathbf{G}_\theta$ – see, e.g., the discussion in Grimaldi and Serinaldi, 2006). From a practical point of view, this latter feature may provide a consistent model of the asymmetric relationship between upstream and downstream river stations, viz. the upstream stations may "influence" the downstream ones, but the converse may be difficult to prove.

In Table 1 (upper triangular) we show the inter-station distances Δ's. It is then immediate to identify, for each site, the nearest neighbor station: namely, $S_2 \leftarrow S_9$, $S_6 \leftarrow S_{10}$,

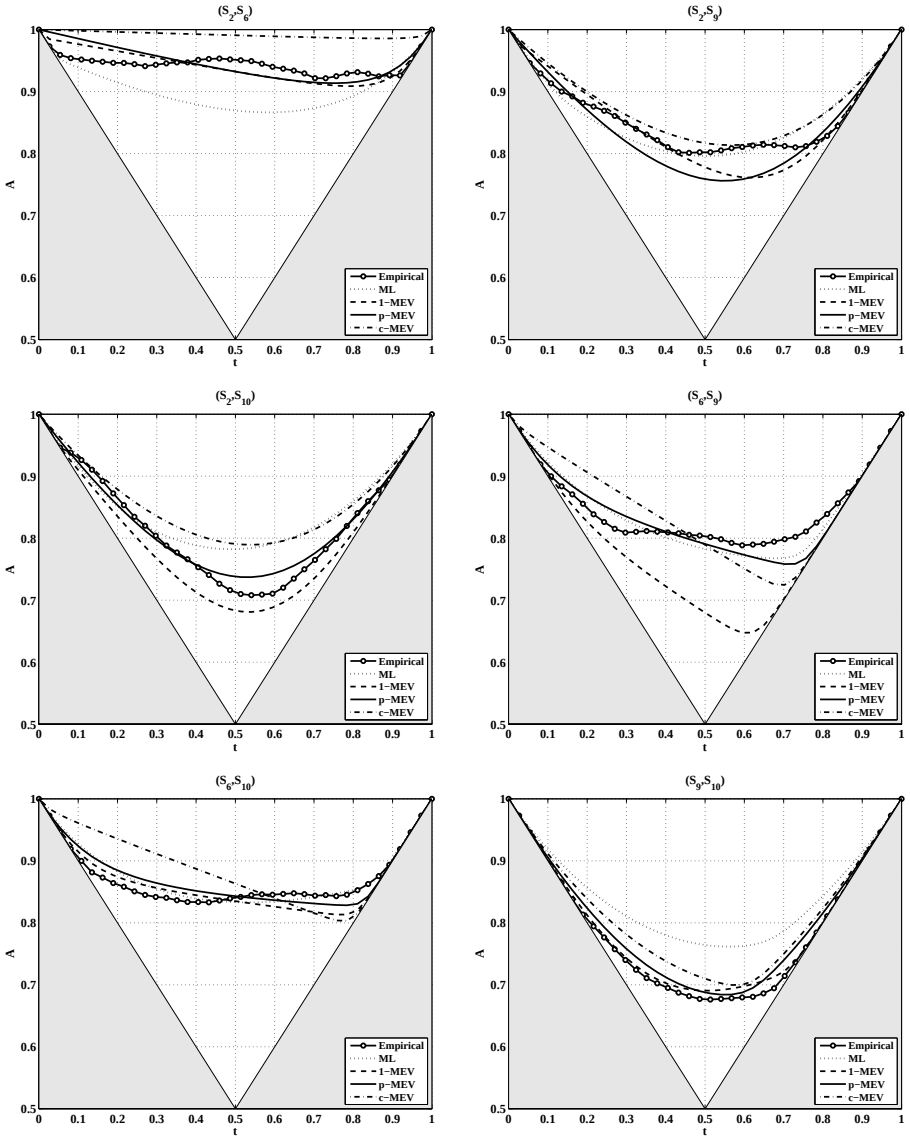

Fig. 2. Plots of empirical and fitted Pickands' dependence functions for all the pairs of stations and the models of interest – see text.

$S_9 \leftrightarrow S_{10}$, i.e. the latter two stations form a pair-cluster (see Table 1 – diagonal). In Table 1 (lower triangular) we show the empirical estimates of the bivariate Kendall's τ, for all the pairs of the four stations of interest here. It is interesting to note that the 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-values, though this does not imply that the stations are statistically independent (as, instead, is commonly misinterpreted). On the contrary, the analysis of the p-values shows that the estimates of the coefficients for all the other pairs are statistically significantly different from zero: this means that the corresponding stations are definitely dependent.

Below we shall statistically compare the performances of the copula model provided by Eq. (22), using sets of parameters fitted via different methods. The estimates are reported in Table 2. The six parameters of **H** have been estimated in Salvadori and De Michele (2010) via ML (see the first row of Table 2): this will give us the possibility to compare and discuss the results of fitting techniques different from the standard one. The estimates of the same parameters according to, respectively, the 1-MEV, the c-MEV, and the p-MEV strategies are also reported in Table 2. For illustrative purposes, the c-MEV approach is run using as multivariate measure of association the Spearman's $\rho_{d,3}$ – see Eq. (8), and considering the following four clusters of stations, having different sizes: $\mathcal{F}_2 = \{S_2, S_9, S_{10}\}$, $\mathcal{F}_6 = \{S_6, S_9, S_{10}\}$, and $\mathcal{F}_9 = \mathcal{F}_{10} = \{S_9, S_{10}\}$. As a variant (not shown), also the multivariate Blomqvist's β_d – see Eq. (9) – was used, but the results did not significantly change. It is interesting to note

Table 2. Estimates of the parameters of the 4-variate copula **H** using different fitting techniques – see text. Also shown are the p-values of the corresponding models.

Method	$\widehat{\xi}$	$\widehat{\chi}$	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	p-v.
ML	1.55	11.04	0.97	0.36	0.78	0.89	0.40
1-MEV	2.73	11.03	0.99	0.12	0.48	0.79	0.78
c-MEV	1.69	11.91	1.00	0.02	0.60	0.75	0.38
p-MEV	1.99	11.03	1.00	0.15	0.71	0.82	0.67

Table 3. Values of the Kendall's τ for all the pairs of the four stations – see text: (*upper triangular*) estimates using the 1-MEV strategy; (*lower triangular*) estimates using the p-MEV strategy.

Station	S_2	S_6	S_9	S_{10}
S_2	1	0.12	0.37	0.55
S_6	0.12	1	0.60	0.32
S_9	0.39	0.39	1	0.57
S_{10}	0.43	0.29	0.54	1

that, independently of the fitting procedure, $\mathbf{G}_\xi \approx \Pi_4$ – the copula of independence – see Eq. (3), – whereas $\mathbf{G}_\chi \approx \mathbf{M}_4$ – the copula of full dependence – see Eq. (4). Thus, as already mentioned, the extra-parametrized copula **H** is a sort of "mixture" between Π_4 and $\mathbf{M}_4$, ruled by the "local" mixing parameters a_i's.

In Fig. 2 we plot the empirical and fitted Pickands' functions A's for all the pairs of stations and the models of interest. The graphs allow for a preliminary visual analysis of the different performances: clearly, being just low-dimensional bivariate slices of the four-dimensional copula **H**, they cannot (and should not) be used to judge the overall fitting ability of the different models – see below. Furthermore, it must be stressed that the empirical estimates of the true (but unknown) dependence functions do not generally respect the convexity constraint, i.e. they are not intrinsic estimators – see also the discussion in Genest and Segers (2009), and references therein.

Apparently, none of the ML, 1-MEV, and c-MEV strategies seem to provide uniformly consistent fits, whereas the p-MEV method overall provides valuable approximations. However, the lacks of fit are more apparent than real: in fact, due to the small sample size, the confidence bands are expected to be quite large. Most interestingly, the copula **H** fitted via the "local" strategies is able to match the asymmetries shown by the empirical functions, and may adapt itself to the "in situ" behaviors of the data. Furthermore, the "degree of dependence", as measured via the Kendall's τ, ranges from ≈ 0.1 to ≈ 0.6 (see Table 3, reporting the estimates for the 1-MEV and p-MEV strategies), whereas the corresponding values fitted via ML only range from ≈ 0.2 to ≈ 0.4 (see Salvadori and De Michele, 2010).

However, when the problem is multivariate, what should always be analyzed is the full dependence structure, and its global ability to fit the actual data. For this purpose, we exploit some robust Goodness-of-Fit tests for multivariate copulas (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. In our case, the p-values are as reported in Table 2. In turn, all the models investigated here should be accepted, since the p-values are much larger than 5%. 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.

A further issue of interest concerns the investigation of the Kendall's return period: this is a fundamental point in applications, since it provides crucial information of practical utility. In principle, it might be possible to use the estimates of the multivariate return period function in order to choose between models fitted via different strategies. In Fig. 3 we show the empirical and the fitted Kendall's return periods for all the four stations and the models of interest: the plot shows the return periods associated with all critical probability levels $t \in \mathbf{I}$. Note that, due to the limited sample size, the estimates of the largest empirical return periods are spoiled (as is well evident in Fig. 3). Visually, both the ML and the p-MEV fits are valuable, whereas the 1-MEV and c-MEV ones apparently fail to provide a consistent approximation: this may not be surprising, since these latter strategies either use the least amount of information, or rely upon measures of association instead of full distributional functions.

As illustrated and discussed in Salvadori and De Michele (2010), these multivariate return periods are generally much larger than the ones calculated via the formulas usually found in literature – see Eq. (18), – and the ensuing discussion). 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 illustrated here, a correct risk analysis can be performed.

5 Conclusions

In order to properly assess the risk, MEV models are fundamental in all areas of geophysics. This paper is of methodological nature, and introduces new estimation techniques for dealing with extremes. In particular, we outline several strategies in order to estimate the parameters of MEV copulas according to different criteria: we use either a "companion" station/cluster approach, or exploit all the pair-wise

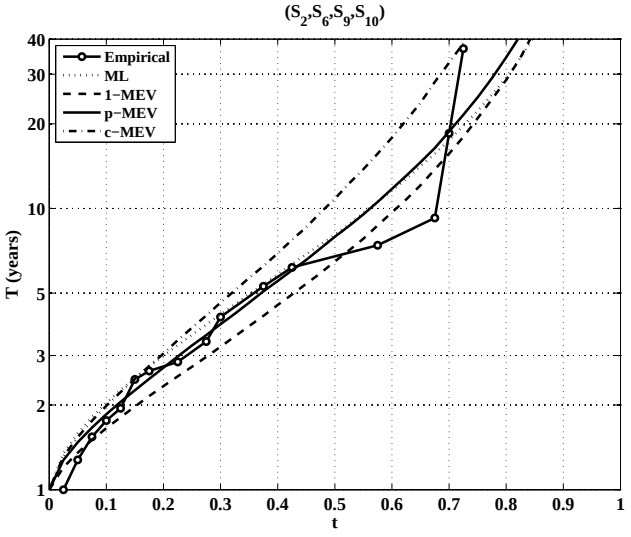

Fig. 3. Plot of empirical and fitted return periods for all the four stations and the models of interest – see text.

relationships between the available gauge stations. The techniques suggested may offer interesting alternatives to standard fitting methods (e.g., ML). An application to flood data is also presented, and a comparison between different estimating strategies is illustrated: this shows how the techniques outlined in the paper can be used in practice.

Acknowledgements. The Authors thank C. Sempi (Università del Salento, Lecce, Italy) and F. Durante (Free University of Bozen-Bolzano, Bolzano, Italy) 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.

Edited by: S. Grimaldi

References

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

Butler, A., Heffernan, J. E., Tawn, J. A., and Flather, R. A.: Trend estimation in extremes of synthetic North Sea surges, J. R. Stat. Soc. – Series C – Appl. Stat., 56, 395–414, 2007.

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

Durante, F. and Salvadori, G.: On the construction of Multivariate Extreme Value models via copulas, Environmetrics, 21, 143–161, 2010.

Durante, F., Quesada-Molina, J., and Úbeda-Flores, M.: On a family of multivariate copulas for aggregation processes, Informat. Sciences, 177, 5715–5724, 2007.

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

Galéa, G. and Canali, S.: Régionalisation des modules annuels et des régimes d'étiage du bassin hydrographique de la Moselle française: liens entre modèles régionaux, Rev. Sci Eau, 18, 331–352, 2005.

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

Genest, C. and Rivest, L.-P.: Statistical inference procedures for bivariate Archimedean copulas, J. Amer. Statist. Assoc., 88, 1034–1043, 1993.

Genest, C. and Rivest, L.-P.: On the multivariate probability integral transformation, Statist. Probab. Lett., 53, 391–399, 2001.

Genest, C. and Segers, J.: Rank-based inference for bivariate Extreme Value copulas, The Annals of Statistics, 37, 2990–3022, 2009.

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

Genest, C., Quessy, J.-F., and Rémillard, B.: Goodness-of-fit procedures for copula models based on the probability integral transformation, Scandinavian, J. Statist., 33, 337–366, 2006.

Genest, C., Remillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study. Insurance Mathematics and Economics, 44(2), 199–213, 2009.

Ghoudi, K., Khoudraji, A., and Rivest, L.: Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles, Canad. J. Statist., 26, 187–197, 1998.

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

GREHYS: Presentation and review of some methods for regional flood frequency analysis, J. Hydrol., 186, 63–84, 1996.

Grenney, W. and Heyse, E.: Suspended sediment – river flow analysis, J. Environ. Eng., 111, 790–803, 1985.

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

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

Joe, H.: Multivariate concordance, J. Multivariate Anal., 35, 12–30, 1990.

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

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

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

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

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

Merz, B. and Bloeschl, G.: Regionalization of catchment model parameters, J. Hydrol., 287, 95–123, 2004.

Nelsen, R.: Nonparametric measures of multivariate association,

in: Distributions with fixed marginals and related topics (Seattle, WA, 1993), vol. 28 of *IMS Lecture Notes Monogr. Ser.*, 223–232, Inst. Math. Statist., Hayward, CA, 1996.

Nelsen, R.: Concordance and copulas: a survey, in: Distributions with given marginals and statistical modelling, Kluwer Acad. Publ., Dordrecht, 169–177, 2002.

Nelsen, R.: An introduction to copulas, Springer-Verlag, New York, second edn., 2006.

Ouarda, T., Bâ, K., Diaz-Delgado, C., Cârsteanu, A., Chokmani, K., Gingras, H., Quentin, E., Trujillo, E., and Bobée, B.: Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study, J. Hydrol., 348, 40–58, 2008.

Pickands, J.: Multivariate Extreme Value Distributions, Bull. Int. Statist. Inst., 49, 859–878, 1981.

Pons, F.: Regional flood frequency analysis based on multivariate lognormal models, Ph.D. thesis, Colorado State University, Fort Collins, 1992.

Russo, D.: On probability distribution of hydraulic conductivity in variably saturated bimodal heterogeneous formations, Journal of Vadose Zone, 8, 611–622, 2009.

Salvadori, G.: Bivariate return periods via 2-copulas, Statist. Methodol., 1, 129–144, 2004.

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

Salvadori, G. and De Michele, C.: Multivariate multiparameter Extreme Value models and Return Periods: a copula approach, Water Resour. Res., 46, W10501, doi:10.1029/2009WR009040, 2010.

Salvadori, G., De Michele, C., Kottegoda, N., and Rosso, R.: Extremes in nature. An approach using copulas, vol. 56 of *Water Science and Technology Library*, Springer, 2007.

Schmid, F. and Schmidt, R.: Multivariate extensions of Spearman's rho and related statistics, Statist. Probab. Lett., 77, 407–416, 2007a.

Schmid, F. and Schmidt, R.: Multivariate conditional versions of Spearman's rho and related measures of tail dependence, J. Multivar. Anal., 98, 1123–1140, 2007b.

Schmid, F. and Schmidt, R.: Nonparametric inference on multivariate versions of Blomqvist's beta and related measures of tail dependence, Metrika, 66, 323–354, 2007c.

Scottish Environment Protection Agency: http://www.nwl.ac.uk, 2009.

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

Shu, C. and Ouarda, T.: Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system, J. Hydrol., 349, 31–43, 2008.

Singh, V., (Ed.): Hydrologic frequency modeling, Reidel Publishing Company, 1986.

Singh, V., Jain, S., and Tyagi, A.: Risk and Reliability Analysis, ASCE Press, Reston, Virginia, 2007.

Sklar, A.: Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris, 8, 229–231, 1959.

St-Hilaire, A., Ouarda, T., Lachance, M., Bobée, B., Barbet, M., and Bruneau, P.: La régionalisation des précipitations: une revue bibliographique des développements récents, Rev. Sci. Eau, 16, 27–54, 2003.

Taylor, M.: Multivariate measures of concordance, Ann. Inst. Statist. Math., 59, 789–806, 2007.

Úbeda-Flores, M.: Multivariate versions of Blomqvist's beta and Spearman's footrule, Ann. Inst. Statist. Math., 57, 781–788, 2005.

Wagener, T. and Wheater, H.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, J. Hydrol., 320, 132–154, 2006.

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

Yue, S.: The Gumbel mixed model applied to storm frequency analysis, Water Resour. Management, 14, 377–389, 2000a.

Yue, S.: The Gumbel logistic model for representing a multivariate storm event, Advan. in Water Resour., 24, 179–185, 2000b.

Zhang, L.: Multivariate hydrological frequency analysis and risk mapping, Ph.D. thesis, Louisiana State University, Baton Rouge, Louisiana (USA), 2005.

Zhang, L. and Singh, V.: Gumbel-Hougaard Copula for Trivariate Rainfall Frequency Analysis, ASCE – J. Hydrol. Eng., 12, 409–419, 2007.