Estimating strategies for multiparameter Multivariate Extreme Value copulas

G. Salvadori and C. De Michele

Reply to Referee 2

The Authors wish to express their gratitude to Referee 2: her/his illuminating comments gave us the possibility to properly fix several questionable issues, and improve the overall quality of the paper. Below please find our point-to-point reply to the Referee's objections.

1. Referee 2 writes: "However, as noticed by the authors, many of the maxima of different gauges are not corresponding to a common flood event. Therefore, the dependence between unrelated maxima does not provide useful information from the regional flood frequency perspectives. To my personal point of view, a qualified joint peak flow probabilistic model should be able to reflect the flood consequence from an upstream gauge to downstream ones. Therefore, the appropriate strategy of sample selection is also very important and needs to be realistic (some mathematical scarification may be needed). The current quadruple sample sets used in this study cannot preserve the correlation of the flood movement from gauge to gauge. It deserves for further investigation."

   The target of this paper is the introduction of fitting techniques for multivariate copulas. The choice of a particular flow probabilistic model, only used for illustrative purposes, is of minor relevance. In all cases, the problem raised by Referee 2 is intrinsic to the sampling process of multivariate extremes (maxima and/or minima). For instance, except for very particular cases, the same storm may generate floods in a basin which may not all be classified as maxima in all gauge stations, though their values may be quite close to the (say, annual) recorded maximum in each single station. In turn, rarely the analysis of flood observations may return a set of temporally coincident extremes (also considering the fact that, usually, basins are regulated). A possible way to bypass the problem, and preserve the correlation of the flood movement, would be to use suitable maximal order statistics $X_{(n-1)}, X_{(n-2)}, \ldots$, but this strategy may only represent an approximation (which needs a lot of statistical care), and requires the knowledge of the full data set (not available in this particular case).

   In addition, in Salvadori and De Michele (WRR, 2010), we showed that the estimates of the percentages p_{ij}'s of contemporary occurrences of annual maxima for the pair of sites (i, j) are empirically consistent with those of the Kendall's τ_{ij}'s: thus, the co-occurrence of annual maxima can be a physical explanation of the pair-wise statistical association of the stations.

2. Referee 2 writes: "The authors claimed their techniques to be physically-based, which seemed to be an over-statement."

   This is our fault, for we are not English mother-tongue: indeed, we misinterpreted the meaning of the word "physically-based". Then, throughout the revised version of the paper we discarded the adjective "physically-based" when making reference to our techniques.

3. Referee 2 writes: "If I understand it correctly, instead of estimating the local parameters using all streamflow gauge observation all together, the authors identified a suitable subset based on the inter-gauge distance. The approach could be justifiable for extreme rainfall analysis, but it may not be appropriate for streamflow."

   We agree with the objection of Referee 2. This gave us the possibility to generalize the techniques outlined in the previous version of the paper. In turn, the fitting strategies outlined in the revised version do not necessarily use the geometrical distance as a criterion to choose the source of information. More particularly:

   - The 1-MEV approach is now named "The single station approach", and the new strategy is described as follows: "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...".

     Also, the following final comment/warning is added: "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."
   - The new c-MEV approach is described as follows: "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)."

     The practical description of the new method introduces further generalizations: "Let S_i be the i-th station, and let $\mathcal{C}_{m_i}^{(i)}$ be a cluster of m_i stations "pertinent" to S_i, with $1 \leq m_i < d$. Clearly, the choice of m_i, as well as the selection of the set of relevant companion stations belonging to $\mathcal{C}_{m_i}^{(i)}$, can be made dependent upon specific basin characteristics, and changed when considering different stations."

   In the Case Study section the minimum Euclidean distance is adopted in the 1-MEV approach as a criterion for choosing the source of information: however, it is clearly pointed out that such a choice "is motivated by illustrative purposes only". Similarly, also the selection of the four clusters $\mathcal{F}_2 = \{S_2, S_9, S_{10}\}$, $\mathcal{F}_6 = \{S_6, S_9, S_{10}\}$, and $\mathcal{F}_9 = \mathcal{F}_{10} = \{S_9, S_{10}\}$ used in the c-MEV approach, and having different sizes, is motivated by illustrative purposes only.

4. Referee 2 writes: "The authors proposed three estimation approaches in this paper, namely the nearest neighbor (1-MEV), cluster (c-MEV), and p-MEV. However, only 1-MEV and p-MEV results are shown in the case study, so there is no way to compare the performance of c-MEV. Being a methodological discussion paper, it seems a requirement that the authors should show some results of c-MEV, even if it is in a very trivial form."

   The c-MEV approach is used and thoroughly discussed in the revised version of the paper.