Serinaldi F. *An uncertain journey around the tails of multivariate hydrological distributions*. *Water Resources Research* 2013, 49(10), 6527-6547.

Copyright: ©2013 The Authors. Water Resources Research published by Wiley on behalf of American Geophysical Union.

DOI link to article: [http://dx.doi.org/10.1002/wrcr.20531](http://dx.doi.org/10.1002/wrcr.20531)

Date deposited: 16\textsuperscript{th} January 2014

This work is licensed under a [Creative Commons Attribution 3.0 Unported License](http://creativecommons.org/licenses/by/3.0/).
An uncertain journey around the tails of multivariate hydrological distributions

Francesco Serinaldi¹ ²

Received 1 March 2013; revised 5 September 2013; accepted 12 September 2013; published 9 October 2013.

Moving from univariate to multivariate frequency analysis, this study extends the Klemes’ critique of the widespread belief that the increasingly refined mathematical structures of probability functions increase the accuracy and credibility of the extrapolated upper tails of the fitted distribution models. In particular, we discuss key aspects of multivariate frequency analysis applied to hydrological data such as the selection of multivariate design events (i.e., appropriate subsets or scenarios of multiplets that exhibit the same joint probability to be used in design applications) and the assessment of the corresponding uncertainty. Since these problems are often overlooked or treated separately, and sometimes confused, we attempt to clarify properties, advantages, shortcomings, and reliability of results of frequency analysis. We suggest a selection method of multivariate design events with prescribed joint probability based on simple Monte Carlo simulations that accounts for the uncertainty affecting the inference results and the multivariate extreme quantiles. It is also shown that the exploration of the p-level probability regions of a joint distribution returns a set of events that is a subset of the p-level scenarios resulting from an appropriate assessment of the sampling uncertainty, thus tending to overlook more extreme and potentially dangerous events with the same (uncertain) joint probability. Moreover, a quantitative assessment of the uncertainty of multivariate quantiles is provided by introducing the concept of joint confidence intervals. From an operational point of view, the simulated event sets describing the distribution of the multivariate p-level quantiles can be used to perform multivariate risk analysis under sampling uncertainty. As an example of the practical implications of this study, we analyze two case studies already presented in the literature.


1. Introduction and General Overview

Many hydrological phenomena are characterized by properties that are (or seem to be) stochastically correlated. These features are often treated as correlated random variables and modeled by multivariate joint distribution functions. In particular, the application of multivariate frequency analysis to hydrological variables has grown quickly since the introduction of copulas in hydrology and geosciences by De Michele and Salvadori [2003]. The possibility of splitting the statistical inference of marginal distributions and dependence structure (or copula) provided by Sklar’s theorem [Sklar, 1959] along with the increasing availability of open source statistical software, allowed for simplifying the analysis and constructing new flexible joint distribution functions with heterogeneous dependence structures and marginal distributions. In the last 10 years, copula-based joint distributions have been defined for a number of hydrological phenomena that are deemed to be characterized by correlated features, such as the peak, volume and duration of flood events, the peak, mean intensity, duration and volume of hyetographs or peak, severity, duration and areal extension of droughts. A large part of the literature on multivariate frequency analysis of hydrological data has focused on the inference procedures, providing a copula-based joint distribution and their lower dimensional marginals and conditional distributions as a final result [e.g., Yue et al., 1999; Yue, 2000, 2001a, 2001b; Grimaldi and Serinaldi, 2006; Zhang and Singh, 2007; Karmakar and Simonovic, 2009; Ghizzoni et al., 2010]. The resultant joint and conditional probabilities are sometimes transformed in return periods by relationships such as \( T = \mu/(1 - p) \), where \( \mu \) is the average interarrival time between two events, \( p \) is a generic probability of nonexceedance (univariate, conditional, or multivariate), and \( T \) is the return period corresponding to \( p \).
interpretation of results and their operational use in practical design problems. In this respect, only a few studies have gone beyond the inference stage and have tackled the problems of the application to real-world hydrologic engineering. A parallel with univariate frequency analysis can help to introduce the key aspects that likely prevented a wider (and appropriate) application and diffusion of multivariate frequency analysis. In a univariate framework, a big effort has been ongoing for several decades to find the distribution that shows the most accurate fit of the data at hand. With this objective, a variety of multiparameter distributions which are more or less theoretically based and flexible were suggested [Rao and Hamed, 2000], along with a number of inference procedures such as moments, L-moments, maximum likelihood, maximum entropy, and Bayesian techniques. However, this effort to achieve high accuracy and a perfect fit is often frustrated by the limited sample size of hydrological data, which implies large uncertainty on the extreme quantiles. This uncertainty must be clearly defined and assessed and taken into account in design applications [Montanari, 2007, 2011; Xu et al., 2010]. Even in a univariate framework, the uncertainty is not always easy to be incorporated in the design practice and communicated [Di Baldassarre et al., 2010]. Therefore, practical problems are often solved by using simple point estimates. However, in univariate frequency analysis, rather simple analytical and Monte Carlo methods allow us to explore the uncertainty of extreme quantiles, such as extreme flood peaks, and to assess the impact on the final output of the risk analysis. In addition, the lack of information resulting from a small sample size is well known and several approaches, such as regionalization procedures [Hosking and Wallis, 1997; Reed, 2002] and information merging [Reed, 2002; Merz and Blöschl, 2008a, 2008b; Viglione et al., 2013], have been suggested to reduce it.

In multivariate frequency analysis, the situation is similar. Most of the literature focuses on finding the best joint distribution that fits the multivariate empirical distribution of the data without accounting for the uncertainty and then the reliability of the estimates and their applicability to design problems, whereas regionalization procedures are in their early stage of development [Chebana and Ouarda, 2009]. As the small size of hydrological samples has a dramatic impact on the reliability of predictions of extreme quantiles [Klemes, 2000a, 2000b; Reed, 2002; Serinaldi, 2009; Xu et al., 2010], a strong impact is also expected in a multivariate framework, in which the uncertainty of the marginal distributions combines with the uncertainty of the dependence structure, whereas the sample size is kept generally small (e.g., annual maxima). Moreover, in a multivariate context, an additional problem is the existence of infinite combinations of the studied variables (e.g., flood peak and duration) that share the same joint probability even though their impact on the design can be very different [e.g., Salvadori et al., 2011]. This makes the selection of appropriate design events more difficult than in the univariate analysis.

1.1. Overview on Multivariate Design Events

To make the presentation clearer and with no loss of generality, it is worth visualizing the concepts mentioned above by referring to the bivariate case. Figure 1 shows the bivariate cumulative distribution function (CDF) \( F_{XY}(x, y) \) and probability density function (PDF) \( f_{XY}(x, y) \) of two generic random variables \( X \) and \( Y \) that follow a Gumbel logistic model [Gumbel and Mustafi, 1967] resulting from the combination of standard Gumbel marginal distributions \( F_X(x) = \exp(-\exp(-x)) \) and \( F_Y(y) = \exp(-\exp(-y)) \) and a Gumbel-Hougaard copula with association parameter \( \theta = 2 \) (Kendall correlation coefficient \( \tau = 0.5 \)). The figure also shows 5000 realizations simulated by this joint distribution and the level curves obtained by cutting the CDF and PDF surfaces through horizontal planes. Each level curve is the locus of points \( \{ L_p^F = (x, y) \in \mathbb{R}^2 : F_{XY}(x, y) = p \} \) and \( \{ L_p^f = (x, y) \in \mathbb{R}^2 : f_{XY}(x, y) = r \} \). As specified by Salvadori et al. [2011], these curves become p-level surfaces and hyper-surfaces when the distribution is trivariate and multivariate. Keeping the discussion in two dimensions, Figure 1 illustrates that a virtually infinite number of combinations of \( X \) and \( Y \) share the same joint probability of nonexceedance \( \Pr[X \leq x \cap Y \leq y] = F_{XY}(x, y) \). However, the simulations show that some points lying on a p-level curve are simulated more likely than others and the cloud of point obviously reflects the shape of the joint density. Therefore, even though all points on a p-level region have the same joint probability, under the hypothesis that the process is bivariate logistic, the different likelihood of each point must be taken into account to select appropriate design scenarios.

The problem of selecting design events along a p-level curve of a bivariate CDF has been well known since at least the mid 1980s when it was tackled for instance by Sackl and Bergman [1987] in flood frequency analyses involving flood peak \( Q \) and flood volume \( V \) [see also Bergman and Sackl, 1989; Bergmann et al., 2001]. Namely, Sackl and Bergman [1987] used the peak runoff time \( t_m = V/Q \) to define an appropriate subset of a p-level curve through a physical boundary. After being overlooked for some years, the problem has been newly considered by Chebana and Ouarda [2011] and Salvadori et al. [2011] by exploiting the ease of mathematical manipulation offered by copulas and the advances in multivariate frequency analysis. In particular, Chebana and Ouarda [2011] used geometric properties of the joint CDF p-level curves to distinguish a so-called “proper part” of the curves (which corresponds to the central truly curved part), from a “naïve part” (the straight horizontal and vertical segments (Figure 1, bottom-left)). This approach is free from physical considerations and can be applied to every pair of random variables, thus generalizing to some extent the early idea of Sackl and Bergman [1987]. Almost simultaneously, Salvadori et al. [2011] suggested another general approach to restrict the range of design events with probability \( p \) by using the copula-based formalism, and denoting a p-level region with the unique and general term of critical layer, they introduced the idea to use a suitable function that weights the realizations lying on the critical layer of interest. Following this approach, one can then freely choose the weight function that best fits the practical needs. One of these weight functions could be the joint PDF, which actually describes the likelihood associated with each point lying on a CDF p-level curve (or critical layer). Therefore, one of the solutions proposed by Salvadori et al. [2011] is...
to choose the point with the highest likelihood along the p-level curve. It should be noted that focusing on the density along a p-level curve corresponds to looking for the conditional distribution resulting from the integration of the joint density along the curve. Following this reasoning, Volpi and Fiori [2012] derived such a distribution and defined the proper part of the p-level curve as the subset of points within the two pairs of points \( (x_{\alpha/2}, y_{\alpha/2}) \) and \( (x_{\alpha/2}, y_{\alpha/2}) \) that define the \((1 - \alpha)\%\) area of the conditional distribution mentioned above. This method introduces a probabilistic selection of the proper part of a p-level curve, thus generalizing the approach of Chebana and Ouarda [2011].

Two examples of these conditional distributions are shown in Figure 2. The figure illustrates the level curves \( F_{XY}(x, y) = 0.9 \) and \( F_{XY}(x, y) = F_X(x) + F_Y(y) - F_{XY}(x, y) = \Pr[X \leq x \cup Y \leq y] = 0.9 \) corresponding to the logistic model described previously. The 5000-size simulated sample is also reported to better visualize how the conditional distributions reflect the spread of the points that follow the logistic model along the p-level curves.

1.2. Overview on Uncertainty of Joint Quantiles

The problem of the uncertainty that affects the joint distribution estimated on a small sample was tackled in a very few studies. For instance, Huard et al. [2006] and Silva and Lopes [2008] considered a Bayesian approach to improve the accuracy of the model selection. Even though this method can be well devised in principle to explore the uncertainty of the joint quantiles, the authors have not investigated this aspect. Durante and Salvadori [2010] and Salvadori and De Michele [2011] recognized the need of accounting for uncertainty, suggesting caution about the reliability of the inference results and their practical application. In spite of its importance, the uncertainty assessment seems to have been widely overlooked in multivariate frequency analysis even though it is a prominent aspect of the univariate frequency analysis and often might make the applicability of results questionable.

Following Montanari [2011], the uncertainty can be classified as uncertainty related to data, inherent (or structural) uncertainty, and epistemic uncertainty. In this study, inherent uncertainty refers to the intrinsic behavior of hydrological processes, is irreducible, and is described by probabilistic models (univariate and multivariate distribution functions), whereas epistemic (reducible) uncertainty refers to the model and parameter uncertainty of marginals and copula related to the limited size of hydrological records. Model and parameter uncertainty is epistemic as it can be reduced as new information (data and meta-data) becomes available. In particular, the uncertainty of the copula parameters leads commonly to stronger and weaker...
structures of dependence around the point estimates. In a bivariate case, this results in changes of the shape and curvature of the $p$-level curves (Figure 3a). On the other hand, the marginal uncertainty commonly entails a shift of the $p$-level curves along the two axes due to the fluctuations of position, scale, and shape parameters of the marginal distributions (Figure 3b). Both the sources of uncertainty combine randomly (Figure 3c) and must be carefully considered. Moreover, even though the model is perfectly specified, quantile estimates are characterized by the sampling uncertainty related to the size of the analyzed data set. Since large uncertainty usually affects the univariate analysis of extreme values, it is also likely prominent in a multivariate framework, and can be considered the main obstacle to a practical (effective and reliable) use of the quantiles resulting from multivariate (extreme value) frequency analyses. It should be noted that the multiplicity of scenarios lying on a $p$-level region must not be confused with the variability related to the epistemic and sampling uncertainty. Indeed, comparing Figures 2 and 3d, it is clear that exploring a level curve or, equivalently, the conditional distribution defined on it, does not allow the assessment of the uncertainty of its location over the plane.

Based on the above discussion, in this study, we first discuss some key points related to the uncertainty assessment of extreme quantiles in univariate frequency analysis. Then, we show how these concepts can be extended to multivariate distributions, highlighting the difficulty of performing a reliable model specification, and obtaining accurate estimates of extreme multivariate quantiles for small data sets. Simple Monte Carlo procedures are therefore suggested to simulate sets of events that account for both the variability along the $p$-level regions and the epistemic and sampling uncertainty. A methodology to summarize the uncertainty by multidimensional confidence intervals (CIs) is also discussed. The proposed methodology is applied to two case studies already analyzed in the literature. Concluding remarks close the study.

2. Univariate Frequency Analysis: The Uncertainty of the Extreme Quantiles

Some basic concepts of univariate frequency analysis are recalled to introduce joint CIs and multivariate design events. The assessment of the uncertainty that affects the extreme quantiles is strictly related to the scientific inquiry of Klemeš [2000a, 2000b] on the credibility of extrapolated tails of distribution models supported by the two key hypotheses of frequency analysis, namely, that (1) a hydrological variable $X$, such as annual maximum peak flow, is an independent and identically distributed random variable having a (continuous) distribution $F_X$ of a fairly
simple mathematical form, and (2) an n-year record of its historic observations is a random sample from this distribution. The study of Klemes [2000a, 2000b] highlights some critical aspects of the frequency analysis applied to hydrological data such as the increasing use of purely statistical concepts and the neglect of the physics of catchment processes of flood formation [Reed, 2002], thus creating an impression of knowledge where none may exist [Strupczewski et al., 2006].

[11] Klemes [2000b] observed that the choice of the parent distribution of a hydrological variable relies on fitting a set of candidate models to a “probability plot” of the observation records and then selecting the model that minimize some performance criterion. Independent of the estimation technique (e.g., method of the moments, L-moments, maximum likelihood, maximum entropy, etc.), the fitting procedure commonly aims at minimizing the discrepancies between theoretical and observed values corresponding to a fixed probability of (non)exceedance. In other words, the observation records $X_i, i = 1, ..., n$, are treated as error-corrupted values and their empirical frequencies $\hat{F}_i$ as error-free values, thus contradicting the assumption that the observation records are exact values whose frequency is unknown as they are sampled by an unknown distribution $F_X$. Actually, plotting position formulae provide point estimates $\hat{F}_i$ (i.e., the empirical cumulative distribution functions (ECDFs)) of the true probability of the (unknown) CDF of the sorted sample $X_{1:n}, ..., X_{n:n}$. For instance, the Weibull plotting position formula reads as $\hat{F}_i = \frac{E[F_X(X_{i:n})]}{i/n + 1}$. Klemes [2000b] argued that these point estimates do not provide a reliable picture of the underlying $F_X$ as the common small size of hydrological samples prevents the replacement of the unknown irregularly spaced probabilities $P_i$ of the ordered observations $X_i$ with the regularly spaced plotting positions $\hat{F}_i$. Since the values $P_i$ can be seen as an ordered sample from a standard uniform distribution $U(0, 1)$, denoting $P_{1:n}, ..., P_{n:n}$ the order statistics of the values $P_i$, the theory of the order statistics provides the distributions of $P_{i:n}, i = 1, ..., n$, [e.g., Mood

**Figure 3.** The figure shows the effect of the parameter variability on the shape of the p-level curve $F_{XY}(x, y) = 0.9$ for a Gumbel logistic model. (a–c) Effect of the dependence (copula) parameter, and the position and scale parameters of the Gumbel marginal distributions, respectively. (d) Effect of randomly combining the parameter values listed in the legends of Figures 3a–c. The random combination mimics the effect of the estimation uncertainty (see section 3.2 and Appendix C).
According to equation (1), for an infinitely large sample, there is almost a 37% probability that the largest observation is not in its nominal last quantile, i.e., that \( P_{n,n} < \frac{n}{n+1} \) (by using the Weibull plotting position formula) and there is a 28.96% probability that three or more of the five highest \( P_{i,n} \) values are located in the same quantile class, that is to say, for instance, events such as \( P_{n-4,n}, P_{n-2,n}, P_{n-1,n} \) \([\text{Klemeš}, 2000b]\). Therefore, the high probabilities that the extreme observations may easily be dispersed in a very irregular manner can distort the shape of \( F_X \), thus making the extrapolation outside the observed frequencies questionable and no more credible and objective than a “by-eye” extension. An illustration can help better clarify the Klemeš’ claims and their practical consequences. Figure 4a shows the exponential probability chart of \( n = 50 \) annual maximum daily discharge values from the De Vil River. Referring to the end of this discussion for further details about this time series, and adopting a more familiar terminology, the figure displays a simple Q-Q plot in which the reference distribution is the standard exponential. This distribution has been chosen to help visualize light or heavy upper tail behaviors. The Weibull plotting position has been applied to define the ECDF. The slight concave pattern of the points seems to suggest a light tail behavior; however, the largest observation seems to contradict this conclusion. Following the reasoning of Klemeš [2000b], given the large uncertainty of the \( P_i \) values, the \( n \) sorted observations \( X_{i,n} \) can be reasonably associated with a set of \( n \) sorted random numbers \( P_{i,n} \) simulated from a \( U(0,1) \) distribution. Repeating this procedure many times (e.g., 1000), we obtain a family of alternative curves (Figure 4b). For each observation \( X_{i,n} \), these alternative curves provide the possible actual values of \( P_{i,n} \) according to the sampling uncertainty summarized by the distribution in equation (1). The shaded area in Figure 4b describes the 95% CIs for the \( P_{i,n} \) of each \( X_{i,n} \) obtained by equation (1). As stated by Klemeš [2000b] “... such an ‘uncertainty zone’...” is merely an empirically constructed confidence band for the plotting positions and can be obtained analytically, for any given confidence level, from [equation (1)].” As the probability of the largest
observation $X_{a,n}$ can vary between $\approx 0.9$ and more than 0.998, it is evident that making distributional assumptions about the form of $F_X$, especially about the shape of its upper tail, might be a matter of speculation.

[13] Since the approach adopted by Klemes [2000b] challenged the well-established paradigm of error-free plotting position formulae, his inquiry has been sometimes deemed as provocative and cited as a thought provoking criticism to hydrological frequency analysis. However, it can be shown that identical conclusions can be drawn from the classical paradigm of error-free plotting position formulae when it is correctly applied. Following this method, Klemes’ discussion loses its apparently “heretical” nature and falls into the field of the assessment of the sampling uncertainty of extreme quantiles. Among the methods devised for defining the CIs for quantiles and order statistics (see Serinaldi [2009] for an overview), in the classical framework of error-free plotting position, a nonparametric formulation relies on the relationship

$$F_{X_n}(x_p) = \mathbb{I}[X_{a,n} \leq x_p] = \sum_{k=1}^{n} \binom{n}{k} [F_X(x_p)]^k [1 - F_X(x_p)]^{n-k},$$

which is the counter part of equation (1) for the sorted observations $X_{a,n}$. Actually, equation (1) is a special case of equation (2) corresponding to a random variable $X \sim U(0,1)$. However, unlike equation (1), the use of equation (2) requires the knowledge of the parent distribution $F_X$, which is usually unknown. Nevertheless, $F_X$ can be replaced by a simple nonparametric counter part [Serinaldi, 2011]

where $X_{a,n}$ and $X_{n}$ are the order statistics closest to $x$, $n' = n + 1$, $\varepsilon = \frac{x - X_{a,n}}{(X_{n} - X_{a,n})}$, and the term in the second row is the standard linear interpolator. Equation (3) can be easily specialized for variables with support $[0, +\infty)$ (equation (A1)), such as discharge data, or finite support $[0,1]$ (equation (A3)), such as the copula values. The corresponding quantile function of equation (3) is [Hutson, 2002; Serinaldi, 2009]

$$\hat{s}_p(p) = \begin{cases} X_{1,n} + (X_{2,n} - X_{1,n})\log(n'p), & 0 < p \leq \frac{1}{n+1} \\ (1 - \varepsilon)\frac{x}{n+1} + \varepsilon\frac{x}{n+1} & \frac{1}{n+1} < p \leq \frac{n}{n+1} \\ X_{n,n} - (X_{n,n} - X_{n-1,n})\log(n'(1-p)), & \frac{n}{n+1} < p < 1 \end{cases}$$

where $\varepsilon = n'p - \lfloor n'p \rfloor$ and $\lfloor \cdot \rfloor$ denotes the floor function.

[14] The quantile function in equation (4) allows for extrapolation to extreme probabilities, and performs quite well for data from mid to light-tailed distributions. Finite and large sample properties of this quantile estimator are discussed by Hutson [2002]. Equation (4) can be used to simulate alternative realizations of the observations without knowing the parent distribution $F_X$. The procedure is analogous to that used to simulate the alternative sequences of $P_{1:n}$, the only difference being that the uniformly distributed random sequences are transformed via equation (4) to obtain values with the required distribution. Figure 4c shows some alternative simulated patterns along with the 95% CIs, whereas Figure 4d clearly shows that the CIs computed through the two approaches (i.e., Klemes and classical paradigms) are identical. Since the Klemes’ approach allows for computing CIs around the $P_{1:n}$ values for fixed $X_{a,n}$ values, whereas the classical approach around $X_{n,n}$ values for fixed $P_{1:n}$, the only difference between the two methods is the extension of the CIs along the axes.

[15] Finally, we can reveal the true nature of the De Vil River data: they are not real-world observations but simply 50 realizations of a standard exponential random variable $X$ scaled by the relationship $X' = (X + 1) \cdot 100$. Figure 4e shows the exact 95% CIs of the order statistics when the true $F_X$ is used instead of the nonparametric estimators in equation (A1). The figure highlights that the sampling uncertainty is very large even though we know the exact parent distribution. Strajpcewski et al. [2006] already recognized this aspect arguing that “Knowledge of the ‘true’ multiparameter model would not be sufficient to accept it if its parameters were to be estimated from short data.” Therefore, we believe that the criticism of Klemes was essentially focused on the widespread and incorrect idea that accurate point estimates of extreme quantiles are possible by refining the point estimate techniques in spite of the small sample size. On the other hand, the frequency analysis is already equipped by tools allowing for a fair assessment of the reliability of the estimates resulting from small data sets. The divergence of the exact CIs in Figure 4e leads to the same conclusions drawn by Klemes [2000b], with no additional information and a too small sample size, we can only conclude that the “next” largest event will be larger than the observed one. For the data in Figure 4a, the attempt to fit a distribution in order to accurately match the largest values would most likely have led to the choice of a fat-tailed or heavy-tailed distribution, which however could not practically improve the reliability of the extrapolations.
owing to the sampling uncertainty. Therefore, complementing accurate point estimates with realistic CIs that clearly highlight the lack of information is probably the most correct approach to communicate results of hydrological frequency analyses [e.g., Serinaldi, 2009].

3. Multivariate Frequency Analysis: Joint CIs for Extreme Quantiles and Design Scenarios

3.1. Dealing With Frequencies Less Extreme Than the Observed

[16] The discussion in the previous section can be easily extended to multivariate frequency analysis. Even though a small sample size can prevent a reliable extrapolation toward extreme quantiles and statistical inference tools allow the quantification of the uncertainty of such extrapolations, these concepts are sometimes still overlooked in the univariate frequency analyses due to the difficulties of using and communicating uncertainty in practice. However, they have been overlooked completely in multivariate frequency analysis. This seems to be contradictory because, on one hand, it is largely recognized that univariate point estimates of extreme quantiles must be complemented with CIs as they are generally affected by large uncertainty, whereas on the other hand, point estimates resulting from multivariate distributions are taken and applied without any assessment of their reliability. Since every multivariate distribution can be written in terms of its univariate marginals and a copula by the relationship $F_{XY}(x, y) = C(F_X(x), F_Y(y)) = C(u, v)$ [Sklar, 1959], it should be expected that the joint quantiles are affected by further uncertainty, namely the uncertainty related to the marginals and copula. However, despite its importance, this uncertainty assessment is commonly neglected when the analysis moves from univariate to multivariate, and the rapidly increasing literature on multivariate frequency analysis (especially via copulas) does not account for the problem. In a nutshell, the question is: why are 40 annual maximum values considered (correctly) insufficient to reliably estimate 0.99 or 0.998 quantiles, whereas 40 pairs of values are commonly used to obtain quantiles with 0.99 joint probability? This problem has been recognized by Durante and Salvadori [2010] and Salvadori et al. [2011] in their methodological studies, but in the last decade, there has not been any attempt to deal with the assessment of the reliability of the statistical inference of multivariate joint distributions. From our point of view, this is the main obstacle to the operational application of multivariate frequency analysis to real world problems.

[17] Multivariate CIs for joint quantiles can be deduced by applying the same concepts commonly used in the univariate frequency analysis. In particular, simple Monte Carlo algorithms similar to those described in the previous section can be set up. With no loss of generality, the discussion is supported by an illustrative example based on 110 pairs of values simulated by the bivariate logistic model described in section 1 (Figure 5). As shown previously, unlike the univariate case, in a bivariate (multivariate) framework, the plane (hyperspace) can be partitioned in many different ways corresponding with different design scenarios, which in turn are associated with different joint probabilities such as

\[ p = P[X \leq x \cap Y \leq y] = F_{XY}(x, y), \]  
\[ p' = P[X \leq x \cup Y \leq y] = F_{XY}'(x, y), \]  
\[ p^c = P[X > x| Y \leq y] = F_{XY}^c(x, y), \]  
\[ pK = P[F_{XY}(x, y) \leq w] = K_C. \]

[18] These probabilities are just a few examples and are related to different definitions of joint return periods through the general transformation $T = \mu/(1 - p)$ [Salvadori, 2004; Salvadori et al., 2007, 2011]. However, since the formula of $T$ is a simple transformation of $p$ that does not add further information about the degree or rarity of an event, the discussion is focused on the joint probabilities. Equations (5a) and (5b) describe different surfaces whose $p$-level curves are shown in Figure 5 (see also Figure 2), equation (5d) represents a conditional distribution [e.g., Salvadori et al., 2007, pp. 159–161], whereas equation (5d) describes the distribution function of the joint distribution or copula values and is called Kendall distribution or Kendall measure [e.g., Salvadori et al., 2007, pp. 147–148]. $K_C$ measures the probability that a bivariate observation falls in one of the two semispaces of the plane delimited by a $p$-level curve. The Kendall distribution is used not only to define the so-called Kendall return period [Salvadori et al., 2011; Salvadori and De Michele, 2013; Salvadori et al., 2013] but also to assess the goodness-of-fit between the empirical copula and a tested parametric family.

3.1.1. CIs for Kendall Distribution: Crossing the Bridge Between Univariate and Joint CIs

[19] One of the advantages of using the Kendall distribution is that the multivariate fitting problem can be reduced to a univariate one, thus allowing for the application of
standard univariate inference procedures. This also means that we can assess the uncertainty of $K_C$ by the tools described in section 2. Namely, parametric and nonparametric CIs can be defined. However, unlike the univariate case, the procedure involves two steps:

1. Simulation of alternative samples from the theoretical (known or fitted) copula or resampling of the reference observations via bootstrap and subsequent computation of the joint ECDF

$$F_{2,n}(x_i, y_i) = \mathbb{P}[X \leq x_i, Y \leq y_i] = \frac{1}{n+1} \sum_{i=1}^{n+1} I\{X \leq x_i \cap Y \leq y_i\};$$

2. Simulation of random sequences from the $U(0,1)$ distribution to obtain alternative patterns of $K_C$ to be used to construct appropriate CIs according to the Klemeš paradigm.

The first step allows us to account for the uncertainty of estimating the empirical copula, whereas the second step is devised to quantify the uncertainty of $K_C$. Skipping the first step leads to an underestimation of the overall sampling uncertainty related to the finite (and usually small) sample size. Denoting $F_{XY}(X, Y) = C(U, V) = W$. Figure 6a shows the Q-Q plots of joint ECDF values $\hat{w}_{ij} = F_{2,n}(x_{ij}, y_{ij})$ versus $w_i = K_C^{-1}(F_i(\hat{w}_i))$. The figure is analogous to Figure 4c but for the variable $W$. Figure 6 also shows the corresponding nonparametric 95% CIs obtained by bootstrapping the 110 logistic pairs and computing the ECDF of $W$ by equation (A3). It should be noted that the departure from the straight line could easily lead to the rejection of the hypothesis that sample comes from a Gumbel copula even though it is drawn from this copula and the model parameters are perfectly specified. Figure 6b shows the parametric 95% CIs obtained by simulating further samples of size 110 from the parent bivariate logistic model and using the theoretical $K_C$ and its inverse. To highlight the importance of the first step of the algorithm mentioned above, Figures 6c and 6d show the nonparametric and parametric 95% CIs obtained without resampling or simulating alternative bivariate samples but resampling directly the 110 $\hat{w}_{ij}$ values and simulating directly from $K_C$. The underestimation of the uncertainty for the nonparametric CIs is clearly evident and further stresses that the inference on the copula based upon small samples must be done with care. On the other hand, the

Figure 6. Q-Q plots of the variable $W = C(U, V) = F_{XY}(X, Y)$. The reference distribution is the Kendall distribution of the Gumbel-Hougaard copula. (a) Nonparametric 95% CIs obtained by resampling the 50-size sample displayed in Figure 5. (b) Parametric 95% CIs obtained by simulating alternative 50-size samples from the bivariate logistic model described in the text. (c) Nonparametric $K_C$-based CIs obtained by resampling the values of the joint ECDF computed on the sample in Figure 5. (d) Parametric $K_C$-based CIs obtained by simulating from the theoretical Kendall distribution of the Gumbel-Hougaard copula.
parametric CIs are unaffected by the first step since the model is perfectly specified and simulating from the joint distribution or directly from $K_C$ does not change the final results.

### 3.1.2. Joint CIs

[23] To better visualize the impact of the sampling uncertainty on the inference of joint distributions, it is worth moving from $K_C$ to the original two-dimensional space of the bivariate sample. First of all, it should be further highlighted that, similar to the univariate point estimates of $p$ quantiles, a $p$-level curve is a point estimate of the locus of bivariate quantiles with $p$ joint probability. Therefore, as the CIs are one-dimensional objects that describe the uncertainty of the zero-dimensional univariate point estimates, the uncertainty of a one-dimensional object, such as the $p$-level curves of a bivariate distribution, is described by a two-dimensional object, i.e., areas around the $p$-level curves (the same dimensionality ratio holds for higher dimensional distributions). As mentioned in section 1.2, it follows that exploring the $p$-level curves does not enable us to incorporate any type of epistemic uncertainty, but only the multiplicity of (joint) $p$-level point estimates of a multidimensional probability distribution under perfect model specification. On the other hand, an appropriate evaluation of the uncertainty aims at defining how the $p$-level curves and Volpi-Fiori’s conditional distribution fluctuate over the plane and change their shape owing to the finite information content of the analyzed data.

[24] Bivariate CIs around $p$-level curves can give an appropriate picture of the reliability of the joint quantile estimates. In this study, we focus on the sampling uncertainty (i.e., the epistemic uncertainty that is reducible if additional data are available) and parameter estimation uncertainty, which is enough to highlight the difficulty of obtaining operational results from small samples. How to account for model misspecification, which is another source of epistemic uncertainty, is mentioned at the end of this section.

[25] Referring to a $p$-level curve of $F_{XY}$ its uncertainty can be assessed by using a simple Monte Carlo algorithm (ALGO-A):

1. Simulate a bivariate sample from the theoretical (known or estimated) joint distribution with the same size $n$ of the observed sample.
2. Compute the empirical copula $\hat{C}_i(u_i, v_i) = F_{ij}(x_i, y_i) = w_i$ of the simulated sample, its empirical distribution $\hat{F}_j(w_i)$, and the value $K_C(p)$ of the Kendall measure corresponding to the theoretical copula for a specified joint probability of interest $\frac{1}{n+1} < p < \frac{n}{n+1}$ (e.g., $p = 0.99$ if the focus is on the 0.99 $p$-level curve and $n > 100$).
3. Among the $n$ bivariate realizations, select the pair whose $\hat{F}_j(w_i)$ value is closest to the theoretical value $K_C(p)$.
4. Repeat the previous steps $B$ times (e.g., thousands) to obtain a set of $B$ pairs $(x_{pj}, y_{pj})$.

[26] The $B$ pairs $(x_{pj}, y_{pj})$ represent a set of possible realizations characterized by a joint probability $p$ and the uncertainty related to a finite and limited sample size under a correct model specification. The algorithm relies on the fact that a pair $(x_{pj}, y_{pj})$ with $w_j \sim p$ is usually available for $\frac{1}{n+1} < p < \frac{n}{n+1}$. However, simulating $n$-size samples, several pairs can exhibit the same empirical joint probability (i.e., they lay on the same level curve), thus introducing the so-called statistical ties. It follows that the set of available joint ECDF values might not cover the range $(\frac{1}{n+1}, \frac{n}{n+1})$. Therefore, ALGO-A should be used only if the model set up guarantees that at least a pair $(x_{pj}, y_{pj})$ shows a joint ECDF $w_j > p$. However, an alternative algorithm that allows us to overcome this problem is discussed in section 3.2.

[31] Going back to the output of ALGO-A and referring to the bivariate logistic pseudo-observations introduced at the beginning of section 3.1, Figures 7a–7d show 5000 pairs $(x_{pj}, y_{pj})$ resulting from ALGO-A for the 0.8 and 0.99 $p$-level curves of $F_{XY}$ along with the original pseudo-observations and 1000 pairs simulated from Volpi-Fiori’s conditional distribution. The latter set of values falls within the set obtained by ALGO-A. The output of ALGO-A actually represents a two-dimensional region that summarizes the sampling uncertainty of the position of Volpi-Fiori’s conditional distribution, which describes the likelihood of observing pairs with joint probability $p$.

[32] To construct joint CIs, a suitable significance level $\alpha$ has to be assigned to the pairs $(x_{pj}, y_{pj})$. Since the sample is multidimensional, the simple sorting procedures used for constructing CIs for univariate quantiles cannot be applied. Approaches relying on the concepts of depth functions and “bag” plots [Rousseeuw et al., 1999] are also not effective as the uncertainty areas sometimes are not convex (see e.g., Figure 7c). In this respect, the method of highest density regions (HDR) proposed by Hyndman et al. [1996] was used. A highest density region is the smallest region of the sample space containing a given probability. The method is based on kernel smoothers and allows us to define approximate $(1-\alpha)$-level contours and therefore approximate $(1-\alpha)$-level joint CIs. Hyndman et al. [1996] found optimal bandwidths with respect to integrated mean-squared error and showed that the convergence rate of the density estimator has order $B^{-2/3}$. HDR provides an unbiased estimator of the $(1-\alpha)$-level contours that is asymptotically normally distributed with variance inversely proportional to the sample size [Hyndman, 1996]. Moreover, since the pairs $(x_{pj}, y_{pj})$ are generated pseudorandomly from parametric multivariate distributions, the approximation can be made arbitrarily accurate by increasing the sample size $B$ [Hyndman, 1996]. In this study, $B = 5000$ is used.

[33] Figure 7 shows the 95% and 99% regions for the pairs $(x_{pj}, y_{pj})$. Figure 7 clarifies that the CIs of a bivariate quantile are two-dimensional, and a 110-size sample can be sufficient to obtain rather reliable estimates of bivariate quantiles with joint probabilities $p$ and $p'$ equal to 0.8, but is insufficient for quantiles with joint probabilities 0.99. Therefore, making inference on the 0.99 or 0.995 $p$-level curves, as is usually done in the largest part of the hydrological literature dealing with joint distributions, is essentially a matter of speculation similar to the estimation of the 100-year or 200-year peak flow from 40 or 50 (but also 100) annual maximum values with no additional information.

[34] Figure 7 also shows the results of ALGO-A under the hypothesis that the sample size is equal to 1000 (e–h)
and 10,000 (i–l). The figure highlights the progressive decrease of the uncertainty of the \( p \)-level estimates. Namely, at least 10,000 observations are needed to obtain reliable estimates of a 0.99 \( p \)-level curve and corresponding Volpi-Fiori’s distribution. As the sample size increases, the bivariate confidence regions tend to converge to the set of events described by the Volpi-Fiori’s univariate conditional distribution defined on a \( p \)-level curve. Therefore, these scenarios provide a picture of the inherent uncertainty (i.e., the natural variability that leads to use a probabilistic model), but they do not describe any other type of uncertainty. The distinction between different sources of uncertainty is implicit in the study by Volpi and Fiori [2012], who indeed never used the word uncertainty since their discussion only focused on one type of uncertainty (i.e., the inherent uncertainty described by the probabilistic model) under the implicit hypothesis of perfect model specification. However, an explicit distinction helps avoid confusion between the sources of uncertainty, a possibly dangerous underestimate, and ill-posed confidence in the results’ reliability.

These results highlight the rather limited operational value of performing a multivariate frequency analysis on small samples, and how large should be the sample size to attain negligible sampling uncertainty and reliable quantile values. This problem is generally overlooked in hydrological multivariate analyses in spite of its importance. However, in order to move from speculation to practical applications, the uncertainty of the estimates must be assessed and communicated. In this respect, the proposed methodology provides a possible approach.

**Figure 7.** The figure shows the output of ALGO-A. Each panel displays the reference sample (Figure 5), a \( p \)-level curve, the samples simulated by ALGO-A, and the HDR curves that summarize the 95% and 99% joint confidence areas. Each row of panels highlights the reduction of uncertainty related to the increase of sample size.
It should be noted that ALGO-A is general and can be specialized to obtain CIs and uncertainty scenarios for whatever \(p\)-level curves defined on surfaces that describe joint probabilities of every type and, in principle, of every dimension (unless computation limits). Referring for instance to \(p^*\), the joint CIs shown in Figures 7c–7d, 7g–7l have been obtained by slightly modifying ALGO-A as follows:

1. Simulate a bivariate sample from the theoretical (known or estimated) joint distribution with the same size \(n\) of the observed sample.
2. Compute the joint empirical probability \(F_i(x_i) + F_j(y_j) - F_{ij}(x_i,y_j)\) of the simulated sample, its empirical distribution \(F_i(w_i^*)\) corresponding to the theoretical copula for a specified joint probability of interest \(\frac{1}{n^2} < p < \frac{1}{n^2} (e.g., p = 0.99\) if the focus is on the 0.99 \(p\)-level curve and \(n > 100\)).
3. Among the \(n\) bivariate realizations, select the pair whose \(F_i(w_i^*)\) value is closest to the theoretical value \(K^*(p)\).
4. Repeat the previous steps \(B\) times to obtain a set of \(B\) pairs \((x_{ip}, y_{ip})\).
5. Summarize the (1–\(\alpha\))-level confidence regions by HDR.

The only difference with the original version is the use of the function \(K^*\). Unlike the Kendall measure, which depends only on copulas (see equation (5a)), the distribution functions associated to \(F_{XY}^\ast\) and \(F_{X,Y}^\ast\) change for each type of joint/conditional probability defined in a two-dimensional or multidimensional space. Our pragmatic approach was to replace \(K^*\) with the ECDF and quantile function in equations (A3) and (A4) computed on a 10\(^{-6}\)-size sample simulated from the theoretical bivariate distribution.

Finally, even though the model uncertainty is not accounted for in ALGO-A, it can be easily incorporated. Keeping in mind that the model uncertainty in the present context refers to the lack of knowledge of the true shape (family) of marginals and copulas which data come from, let us suppose that exploratory data analyses return a set of joint distributions that are suitable to describe the available data (as they pass goodness-of-fit tests, for instance). To account for the model uncertainty, ALGO-A can be modified by introducing a random selection from the set of candidate models to be used in the first step of the algorithm. Precisely, for each of the \(B\) replicas, one of the candidate models is randomly selected and the bivariate sample is drawn from it. In this way, the resulting \(B\) pairs \((x_{ip}, y_{ip})\) are not simulated from a unique joint distribution but from a pool of models, allowing the exploration of the model space. The random selection of the models can also be further refined introducing a weighting procedure by assigning a larger (smaller) selection probability to the models that are more (less) credible in terms of some information criterion such as Akaike information criterion (AIC) ([Akaie, 1974; Laio et al., 2009]).

Beyond the Gates of the Observed Quantiles

In the previous section, an algorithm to define design scenarios and joint CIs accounting for the sampling uncertainty has been described. However, the illustration relied on a sample of size 110 and joint frequencies 0.8 = 80/100 and 0.99 = 99/100. Therefore, simulating alternative bivariate samples with size 110 easily allows for selecting at least an event with joint probability of non-exceedance close to 0.99 since the values of the empirical copulas are in the range \(\frac{1}{110^2} \approx 0.0009\) to \(\frac{1}{110^2} \approx 0.099\) (overlooking possible statistical ties resulting from pairs of simulated values lying on the same \(p\)-level curve). In other words, as mentioned in section 3.1.2, ALGO-A works only if the focus is on probabilities smaller than the most extreme empirical probabilities corresponding to the sample size (i.e., \(\frac{1}{n^2}\)). However, the most common case in real world analyses involves a small sample size (e.g., less than 100 annual maximum records) and the need to compute quantiles with joint probabilities larger than about 0.9–0.99. This is exactly the same case commonly faced in univariate frequency analysis, in which extreme quantile estimates are obtained by extrapolating a univariate analytical distribution beyond the observed frequencies and the uncertainty of the extrapolated quantiles (or order statistics) is quantified by CIs (see section 2).

Analogous to the univariate analysis, the shortcoming of ALGO-A can be overcome by resorting to fractional order statistics, which can be defined as \(\tilde{Z}_{u,p,n}\), where \(Z\) is a generic random variable and \(n' = n + 1\) ([Stigler, 1977; Hutson, 1999; Serinaldi, 2009]). By using this theory, the distribution function of a generic quantile \(Z_p\) can be written as [Serinaldi, 2011]

\[
F_{Z_p}(z) = 1 - \int F(z; n')dz, (1 - F(z))
\]

where \(I(z; a, b) = \int_0^1 e^{z - t^a - 1} dt/B(a, b)\) is the beta cumulative distribution function (equation (7) may be inverted numerically by a simple zero-crossing algorithm). Therefore, by defining \(Z := C(U, V) = W\), it follows that \(F_Z\) corresponds to \(K_C\), and equation (7) provides the distribution function of the quantiles of \(C\). For example, for a Gumbel-Hougaard copula with parameter \(\theta = 2\), Figure 8 shows the curves described by equation (7) for \(W\) equal to 0.1, 0.5, and 0.9 and three different values of the sample size \(n\). These distributions can be used to build CIs for \(W_p\) and to simulate. Therefore, an algorithm to construct joint CIs around a \(p\)-level curve is straightforward (ALGO-B):

1. For a given (known or fitted) copula and fixed level value \(p\), simulate \(B\) values \(w_{pj}\) from the distribution in equation (7). The simulated samples describe the uncertainty of the copula quantile \(w = p\) related to sample size, or the same, the oscillation of the \(p\)-level curve around the nominal value \(p\).
2. For each \(w_{pj}\), simulate one realization \((u_{pj}, v_{pj})\) from the \(w_{pj}\)-level curve of the theoretical copula. This can be done by using Volpi-Fiori’s conditional distribution or by a simple acceptance-rejection Monte Carlo algorithm (see Appendix B). The set of simulated pairs \((u_{pj}, v_{pj})\) cover an area of the [0, 1]\(^2\) plane corresponding to the possible locations of the \(p\)-curve according to Volpi-Fiori’s conditional distribution under sampling uncertainty.
3. Apply the marginal backward transformations \(x_p = F_{X_p}^{-1}(u_{pj}), y_p = F_{Y_p}^{-1}(v_{pj})\).
4. Define (1–\(\alpha\))-level confidence areas by HDR.

Figures 9a and 9b show the results obtained by applying ALGO-B to a 50-size bivariate logistic sample drawn from the same logistic model used in the previous
As expected, the uncertainty areas around the level curves are similar to those obtained by ALGO-A, which, however, cannot be applied in this case as the most extreme copula quantile in a 50-size sample is \( \frac{50}{50+1} \approx 0.98 \) (excluding possible ties), whereas the 0.99 \( p \)-level quantile is needed. Note that the mode of the cloud of the simulated values does not lie on the 0.99 \( p \)-level curves. This bias reduces as the sample size \( n \) approaches or becomes greater than the minimal value needed to compute the 0.99 quantile, i.e., 100 (for a discussion of these aspects in the univariate case we refer to Serinaldi [2009]).

ALGO-C is as follows:

1. Simulate \( B \) bivariate samples of size \( n \) as \( x_{ij}, y_{ij} \), for \( i = 1, \ldots, n \) and \( j = 1, \ldots, B \), where the pairs \( x_{ij}, y_{ij} \) are pseudorandom realizations from the joint distribution \( F_{XY} \).

Figure 8. Distribution functions of three quantiles (0.1, 0.5, 0.95) of the variable \( W = C(U, V) = F_{XY}(X, Y) \) for three different values of the sample size \( n \).

Figure 9. Output of ALGO-B and ALGO-C. (top) 95% and 99% joint confidence areas for quantiles (parameter estimation uncertainty under correct model specification). The results refer to the uncertainty of the 0.99 \( p \)-level curves of \( F_{XY} \) and \( F_{X|Y} \), for \( n = 50 \) under the correct hypothesis that the sample is drawn from a bivariate Gumbel logistic model.
2. For each $\tilde{\Theta}$ sample, compute the values $\hat{\Theta}$ of the joint distribution parameters $\Theta$ with the chosen estimation method (e.g., maximum likelihood).

3. For each estimated distribution $F_{XY}(x,y; \hat{\Theta})$, simulate one realization $(x_j,y_j)$ from the $p$-level curve via Volpi-Fiori’s distribution or acceptance-rejection algorithms.

4. Define $(1-\alpha)$ level confidence areas around the $p$-level curve by HDR.

Figures 9c and 9d show the output of ALGO-C for the same model and data displayed in Figures 9a and 9b. Comparing top and bottom panels, for the same sample size $n = 50$, the uncertainty corresponding to the quantile estimates tends to be smaller than that of the quantiles and show a different shape especially for $F_{XY}$. This behavior depends on the different nature of quantiles and quantile estimates [Serinaldi, 2009] and is further discussed in section 4. A last remark concerns the checkerboard-like appearance shown by the pairs $(x_j,y_j)$ in Figure 9a. This is a geometric artifact resulting from simulating from around the proper part of the $p$-level curve. A descriptive explanation is provided in Appendix C.

4. Case Studies

To highlight the importance of complementing a multivariate frequency analysis with CIs, two case studies already presented in the literature are examined.

4.1. Case of Volpi and Fiori [2012]

In the first case study, the data used by Volpi and Fiori [2012] are reanalyzed. The data set has been provided by Elena Volpi (personal communication) and comprises 46 pairs of flood volume $V$ and peak $Q$ that characterize 46 hydrographs extracted from 168 historical annual maximum floods of the Tiber River recorded at Ripetta gauging station located in Rome, Italy. The 46 pairs refer to the events exceeding the threshold 1800 m$^3$/s$^{-1}$.

Volpi and Fiori [2012] specified the joint distribution of the data by the copula approach as $F_{VQ}(v,q) = C(F_V(v), F_Q(q))$, in which a Gumbel-Hougaard copula was chosen for $C$ and two Pearson type III (shifted gamma) distributions were selected for modeling $F_V$ and $F_Q$. For the sake of coherence, we use the same model structure. In order to define the CIs of quantile estimates, the parameters of the marginal CDFs are estimated by L-moment method, whereas the copula parameter $\theta$ is estimated by Kendall correlation $\tau_K$ using the relationship $\tau_K(\theta) = (\theta - 1)/\theta$. This choice allows us to check the uncertainty related to a coherent moment-based estimation procedure (in the next case study the maximum likelihood method is tested). Volpi and Fiori [2012] focused on the so-called “OR” joint return period [Salvadori et al., 2007, pp. 148–157] defined as $T^* = \mu/\left(1 - F_{VQ}\right)$, where $\mu$ is the mean interarrival time (3.65 years in the present case). In particular, Volpi and Fiori [2012] computed the $T^* = 200$ years bivariate quantiles, corresponding with the level curve with probability $p = F_{VQ} = 0.98175$. Thus, the methodology described in the previous sections is applied to define the uncertainty around this level curve. For the sake of completeness, we also consider the level curve corresponding with $F_{VQ} = p^*$. In this case, the value 0.98175 for $p^*$ is used as an example with no links to particular design needs. Marginal and joint CIs for quantiles and quantile estimates are shown in Figures 10 and 11.

Focusing on Figure 10a, the side panels show the marginal ECDFs along with the quantile CIs in appropriate axes scale chosen to highlight the uncertainty of the upper tails. The side panels highlight that the large uncertainty which affects the extreme quantiles makes the accurate choice of the plotting position formulae a secondary aspect. Moreover, the great attention to the choice of estimation methods and model selection procedures are also secondary when dealing with small samples, as the CIs tend to overlap [Serinaldi, 2009] and differences can be statistically non-significant. On the other hand, complementing the point estimates with CIs gives a better picture of the actual reliability of the point estimates and a fairer assessment of lack of knowledge resulting from the limited information contained in small samples. This is also confirmed by the width of the Monte Carlo 95% CIs for the model parameters (Table 1). Unfortunately, these CIs are also not often reported in the scientific literature and technical reports.

The main panel of Figure 10a shows the 95% and 99% joint confidence areas corresponding with the 0.98175 $p$-level curve of the bivariate distribution along with observed pairs and a sample simulated from the Volpi-Fiori’s distribution. As mentioned above, the joint CIs describe the uncertainty of defining the exact location of the $p$-level curve and the corresponding Volpi-Fiori’s distribution on the plane due to the small sample size. The joint CIs cover several level curves, thus alerting us to the large variability of the estimate of extreme $p$-level curves. In addition, the samples used to build the joint CIs around the 0.98175 $p$-level curve correspond to the samples simulated from Volpi-Fiori’s distribution under sampling uncertainty (i.e., uncertain location of the $p$-level curve). To better highlight these points, we have computed the univariate and joint CIs under the hypotheses that the size of the observed sample is 1000, 5000, and 10,000 instead of 46. Figures 10b–10d show how the CIs evolve as the sample size increases, converging to the point estimates, namely, the Pearson type III distribution and the 0.98175 $p$-level curve. The figure also highlights that very large samples (e.g., thousands of realizations) are needed to reliably estimate the marginal and joint extreme quantiles. Moreover, the pairs that have been simulated to construct joint CIs might be used as design scenarios characterized by joint probability $p$ and appropriate sampling uncertainty. Indeed, as is shown in Figure 10a, a large number of pairs can be more extreme than those simulated by Volpi-Fiori’s distribution under perfect knowledge conditions.

Figure 11a reproduces Figure 10a, whereas Figures 11b–11d refer to CIs for quantile estimates of the 0.98175 $p$-level curve of $F_{VQ}$ and the CIs of quantiles and quantile estimates of the 0.98175 $p$-level curve of $F_{VQ}$. Similar considerations apply to the other panels. For the present case study, the difference between CIs for quantiles and quantile estimates is not so relevant from a practical point of view, even though the two types of CIs reflect different sources of uncertainty. It should be noted that in principle the CIs of the marginal and joint quantiles can be related in some way; however, this aspect requires further investigation which goes beyond the aim of this study. Nevertheless, in
In this respect, it should be mentioned that each \( p \)-level curve defined on \( F_{VQ} \) embeds several univariate quantiles, namely, all quantiles of \( Q \) and \( V \) with \( F_Q \in (p, 1) \) and \( F_V \in (p, 1) \); therefore, establishing a direct and simple relationship (if any) between univariate and joint CIs is not straightforward.

### 4.2. Case of Yue et al. [1999]

The data set used in this case study is taken from Yue et al. [1999] and has been reanalyzed by Chebana and Ouarda [2011] to define the proper part of the \( p \)-level curves of the joint distribution \( F_{VQ} \). The data comprise 33 pairs of flood volume \( V \) and peak \( Q \) extracted from a daily streamflow record spanning from 1963 to 1995 at the gauging station 061901 near the outlet of the Ashuapmushuan basin located in the Saguenay region in the province of Québec, Canada, at latitude 48.5–50.5°N and longitude 72.5–74.5°W.

---

**Figure 10.** (a) 95% and 99% joint CIs resulting from ALGO-B for the Tiber River data and the \( p \)-level curve corresponding with \( F_{VQ}(x, y) = 0.98175 \). The side panels show the 95% and 99% CIs of the marginal distributions. (b–d) Evolution of the joint confidence area under the hypothesis that the observed sample size is 500, 5000, and 10,000, respectively. The diagrams allow for the assessment of the size required to obtain reliable estimates.
Figure 11. (a) Reproduces Figure 10a and (b–d) CIs for the quantile estimates of the $p$-level curve corresponding with $F_{XY}(x,y) = 0.98175$, and quantiles and quantile estimates of the $p$-level curve corresponding with $F_{XY}(x,y) = 0.98175$.

Table 1. Parameter Values of the Joint Distributions Fitted on the Tiber River Data and Ashuapmushuan Basin Data

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Tiber River</th>
<th>Ashuapmushuan Basin</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>0.64 (–2.37, 2.47) [mm$^3$]</td>
<td>46,387.37 (44466.08, 48482.53) [m$^3$ s$^{-1}$ day$^{-1}$]</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>113.19 (66.58, 188.48) [mm$^3$]</td>
<td>10,184.48 (8702.90, 11656.11) [m$^3$ s$^{-1}$ day$^{-1}$]</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>0.38 (0.24, 0.61) [–]</td>
<td>1798.41 (1731.13, 1835.71) [m$^3$ s$^{-1}$]</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>0.92 (0.56, 1.72) [–]</td>
<td>372.70 (226.51, 581.08) [m$^3$ s$^{-1}$]</td>
</tr>
<tr>
<td>$\theta_c$</td>
<td>6.31 (4.95, 8.01) [–]</td>
<td>6.31 (4.95, 8.01) [–]</td>
</tr>
</tbody>
</table>

Values in parentheses denote the 95% CIs for the parameters. A one-parameter Gumbel-Hougaard copula is used to describe the dependence structure for both data sets. For the Tiber River case study, the marginal distributions of $V$ and $Q$ are modeled by three-parameter Pearson type III distributions $F_V(x; \theta) = \int_0^x \frac{(1-\frac{x^2}{\theta_1})^{-\theta_2}}{\theta_3} \exp \left( -\frac{x^2}{\theta_1} \right) dx$. For the Ashuapmushuan basin case study, $V$ and $Q$ are modeled by two-parameter Gumbel distributions $F_X(x; \theta) = \exp \left( -\exp \left( -\frac{x - \theta_0}{\theta_c} \right) \right)$. 

SERINALDI: MULTIVARIATE FREQUENCY ANALYSIS WITH UNCERTAINTY
whereas two different models were applied to model the copula, namely, a bivariate Gumbel mixed model [Yue et al., 1999] and a Gumbel logistic model [Chebana and Ouarda, 2011]. In this study, the latter is considered. Unlike the previous case study, marginal parameters are estimated by the maximum likelihood method and the copula parameter by the canonical maximum likelihood. The parameter estimates and their Monte Carlo 95% CIs are reported in Table 1. Figure 12 is similar to Figure 11 and shows the marginal and joint distributions for the studied data set. Since Gumbel marginals and Gumbel-Hougaard copula return a so-called bivariate Gumbel logistic distribution, the diagrams look like those displayed in Figure 9, apart from different values of the model parameters. The figure highlights the large uncertainty affecting both univariate and bivariate quantiles. This uncertainty must be taken into account when the results are communicated to provide a fair description of the reliability of the univariate and multivariate quantile point estimates. In particular, it should be noted that the univariate and bivariate CIs related to a 0.99 p-level curve cover several p-level curves with different joint probability. Therefore, even though it is common practice, in dealing with hydrological multivariate frequency analysis, to show level curves and use them to

Figure 12. As for Figure 11, but for the Ashuapmushuan basin data set. CIs are computed for the 0.99 p-level curves of $F_{XY}$ and $F'_{XY}$. 
provide the scenarios characterized by a prescribed joint probability, these values must be seen as only approximate estimates (analogous to the univariate case) because of their large uncertainty. On the other hand, the Monte Carlo algorithms used to construct the confidence areas provide synthetic scenarios that (1) are coherent with the fitted joint distributions, (2) are characterized by the required joint probability, and (3) take into account the sampling and estimation uncertainty.

5. Conclusions

In this study, we have shown how some techniques devised to construct appropriate confidence intervals for quantiles can be extended to a multivariate framework. This stage of the frequency analysis is often overlooked in univariate frequency analysis and has been never tackled in a multivariate framework in spite of its widely recognized importance. Since it is known that the size of samples commonly available in hydrological frequency analysis is insufficient to obtain reliable results, this study explicitly quantifies the effect of some sources of uncertainty and shows that the results usually provided in the literature must be used with care for practical applications. In particular, it should be noted that the $p$-level regions and the corresponding $p$-level scenarios do not quantify epistemic uncertainty but have to be seen as multivariate point estimates. When the sampling uncertainty is accounted for, these $p$-level regions are simply subsets of the $p$-level uncertainty regions, thus leading to a false and possibly dangerous confidence in the results’ reliability. We have shown that the $K_C$ function is affected by sampling uncertainty and the multivariate $p$-level regions corresponding with different values of $p$ can overlap as for univariate quantiles [Serinaldi, 2009]. Therefore, even though accurate inference procedures must always be applied, as for univariate frequency analysis, they cannot solve the problem of the lack of information related to short time series. When the sample is small, many joint distributions and copulas can fit the data adequately and goodness-of-fit tests cannot discriminate between alternative models because of the lack of power. In these cases, choosing a copula and a joint distribution that minimize some performance criterion is secondary compared with the communication of the (large) sampling uncertainty that affects the univariate and multivariate point estimates. On the other hand, effort should be focused upon the retrieval and use of information related to pre-existing general knowledge of the underlying processes that generate the collected data and meta-data, i.e., the specific knowledge [Christakos, 2011; Serinaldi and Kilsby, 2013]. Therefore, general knowledge and specific knowledge should be merged to set up data augmentation techniques that provide samples large enough to allow for accurate point estimates independently of the shape of the joint distribution.

The Monte Carlo algorithms proposed in this study can be applied in principle to joint distributions of every dimension as they rely on the univariate distribution functions (e.g., $K_C$ and $K''$) of the joint probability surfaces (e.g., $P_{XY}$ and $P_{XY''}$). The only limitation is just computational; for instance, as the number of dimensions and the model complexity increase, exploring $p$-level regions and applying acceptance-rejection algorithms can be time expensive and more efficient simulation techniques must be used. In addition, the HDR method used to summarize joint CIs is available up to six dimensions [Hyndman et al., 2012]. However, multivariate frequency analysis of hydrological quantities is almost always limited to a few variables (usually fewer than six) and the uncertainty can be quantified and visualized by focusing on lower dimensional marginals (e.g., pairwise two-dimensional joint distributions).

We also discussed how to incorporate the model uncertainty; however, this type of uncertainty and other sources of uncertainty have not been taken into account in the case studies. It is expected that these sources of uncertainty further increase the variability of the point estimates and they must be suitably processed in order to obtain an exhaustive picture of the overall uncertainty. These remarks further stress the importance of resorting to additional information to obtain reliable estimates.

From an operational point of view, as the multivariate $p$-level event sets returned by the proposed Monte Carlo algorithms describe the distribution function of the $p$-level quantiles under sampling uncertainty, they can be used in multivariate risk analyses. For example, expected damages $D$ for a $p$-level quantile are computed by numerically integrating the damage function over the sampling distribution of the $p$ quantile [e.g., Tung et al., 2006, pp. 435–436]. Therefore, the simulated $p$-level scenarios can be used to perform Monte Carlo integration required to compute the expected damage corresponding for instance to the simultaneous exceedance of critical discharge rates $q_p = \{q_{p1}, q_{p2}\}$ representing possibly the thresholds for damage due to flooding of two residential, business or agricultural areas with joint probability $p = P\{Q_1 > q_{p1}, Q_2 > q_{p2}\}$, which is expressed as

$$
D(q_p, q_{p2}) = \int_{q_{p1}}^\infty D(q_p, q_{p2}) g(q_{p1}|p) dq_{p2}, 
$$

where $D(q_p, q_{p2})$ is the damage function given known flow capacities of the hydraulic structures $q_{p2} = \{q_{p1}, q_{p2}\}$, and $g(q_{p1}|p)$ is the sampling joint density function of the bivariate $p$-level quantiles. The sampling uncertainty can also be combined with the inherent uncertainty and the hydraulic uncertainty associated with the flow-carrying capacities to obtain the annual expected damage cost [Tung et al., 2006, pp. 436]. More generally, the simulated $p$-level combinations can be used to evaluate the effects of different hydrological loads on an infrastructure adopting an ensemble approach [Volpi and Fiori, 2012], thus allowing for the choice of the combination (with joint probability $p$) that is the most critical in terms of structure size or optimal in terms of cost-benefit criteria.

Appendix A: Extended ECDFs

For random variables defined in $[0, \infty)$, The ECDF estimator in equation (3) and the corresponding quantile function in equation (4) specialize as follows:
\[
\hat{F}_X(x) = \begin{cases} 
\frac{x}{(n'x_{1,n})}, & 0 < p \leq \frac{1}{n + 1} \\
(1 - \epsilon) \frac{j}{n + 1} + \epsilon \frac{k}{n + 1}, & \frac{1}{n + 1} < p < \frac{n}{n + 1} \\
\frac{x - x_{n,n} + \epsilon}{1 - x_{n,n} + \frac{n}{n + 1}}, & \frac{n}{n + 1} \leq p < 1 
\end{cases}
\]

and

\[
\hat{F}_X(p) = \begin{cases} 
\epsilon x_{1,n}, & 0 < p \leq \frac{1}{n + 1} \\
(1 - \epsilon) x_{[\epsilon p] + 1,n} + \epsilon x_{[\epsilon p] + 1,n}, & \frac{1}{n + 1} < p < \frac{n}{n + 1} \\
x_{n,n} + \epsilon(1 \cdot x_{n,n}), & \frac{n}{n + 1} \leq p < 1 
\end{cases}
\]

[70] For random variables defined in [0, 1], the ECDF estimator in equation (3) and the corresponding quantile function in equation (4) become

\[
\hat{F}_X(x) = \begin{cases} 
\frac{x}{(n'x_{1,n})}, & 0 < x \leq x_{1,n} \\
(1 - \epsilon) \frac{j}{n + 1} + \epsilon \frac{k}{n + 1}, & x_{j,n} < x < x_{k,n}, j < k = 1, \ldots, n, \\
\frac{x - x_{n,n} + \epsilon}{1 - x_{n,n} + \frac{n}{n + 1}}, & x \geq x_{n,n}
\end{cases}
\]

and

\[
\hat{x}_P(p) = \begin{cases} 
\epsilon x_{1,n}, & 0 < p \leq \frac{1}{n + 1} \\
(1 - \epsilon) x_{[\epsilon p] + 1,n} + \epsilon x_{[\epsilon p] + 1,n}, & \frac{1}{n + 1} < p < \frac{n}{n + 1} \\
x_{n,n} + \epsilon(1 \cdot x_{n,n}), & \frac{n}{n + 1} \leq p < 1 
\end{cases}
\]

[71] Data that lie on a given \( p \)-level curve can be simulated directly by sampling from Volpi-Fiori’s distribution; however, since this distribution needs the definition of an auxiliary curvilinear coordinate system as well as the integration of the joint density along the \( p \)-level curve [Volpi and Fiori, 2012], we suggest an alternative based on a simple acceptance-rejection algorithm that reads as follows for the \( p \)-level curves corresponding to \( F_{XY}^1 \):

[72] 1. Sample \( B \) realizations \( u_j \) with \( j = 1, \ldots, B \), from a uniform distribution defined in \([0, 1] \).

[73] 2. Compute the corresponding values \( v_j \) such that \((u_j, v_j) \in L^p_{\text{ALGO-A}} \). For Archimedean copulas, \( v_j \) can be computed by using explicit relationships [Salvadori et al., 2007, pp. 146]. For the \( p \)-level curves of \( F_{XY}^1 \) and for copulas different from Archimedean, since \( p \)-level curves are monotonic, simple zero-cross finding algorithms can be applied.

[74] 3. Draw \( B \) auxiliary realizations \( \xi_j \) from a uniform distribution defined in \([0, 1] \).

[75] 4. Compute the copula density \( c(u_j, v_j) \) for the \( B \) pairs \((u_j, v_j) \).

[76] 5. Reject the pairs \((u_j, v_j) \) such that \( \frac{c(u_j, v_j)}{\max_j \{ c(u_j, v_j) \} } < \xi_j \).

[77] To simulate along a \( p \)-level curve corresponding to \( F_{XY}^1 \), the values \( u_j \) at the first step have to be sampled from a uniform distribution defined in \([0, p] \).

Appendix B: Simulation of \( p \)-Level Scenarios by Acceptance-Rejection Algorithm

Appendix C: Explanation of Checkerboard-Like Appearance

[78] Figures 9a and 10a show that the samples \((x_{p,j}, y_{p,j})\) exhibit a characteristic checkerboard-like appearance. This texture results from simulating along the \( p \)-level curves of a given joint distribution. Indeed, CIs for quantiles measure the uncertainty of defining the exact \( p \) value of a given level curve. ALGO-A and ALGO-B allow us to explore the neighbor of a level curve corresponding to the nominal joint probability \( p \). Given the shape of the copula density (Figure 1), simulations fall around the proper part of the \( p \)-level curves yielding the checkerboard-like appearance. Figure 13 shows that the same effect can be obtained by overlapping an “L”-shaped texture which mimics points belonging to the proper part of a \( p \)-level curve. The checkerboard-like appearance does not emerge for the samples drawn from ALGO-C. Since ALGO-C explores the parameter uncertainty, in this case the simulated pairs \((x_{p,j}, y_{p,j})\) come from copulas with slightly different parameters and different shapes of the level curves (see e.g., Figure 3d); therefore, this shape variability
Figure 13. Illustration of the checkerboard-like appearance exhibited by the output samples of ALGO-A for the p-level curves of $F_{XY}$.

prevents the emergence of regular spatial patterns like the checkerboard texture.

[79] Acknowledgments. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) “UK Infrastructure Transitions Research Consortium” grant EP/I01344X/1, EPSRC “Flood MEMORY: Multi-Event Modelling Of Risk & recovery” grant EP/K013513/1, and Willis Research Network. The author thanks Elena Volpi for kindly providing the Tiber River data and for the fruitful discussion, Chris G. Kilshy (Newcastle University, UK) for his useful remarks, the Associate Editor, and three anonymous reviewers for their comments and criticisms. The analyses were performed in R [R Core Team, 2012] by using the contributed packages Cairo [Urbaneck and Horner, 2011], CDVine [Schepsmeier and Brechmann, 2012], copula [Yan, 2007; Kojadinovic and Yan, 2010], evd [Stephenson, 2002], hdrcde [Hyndman et al., 2012], HH [Heiberger, 2012], KernSmooth [Wand, 2012], nacopula [Hofert and Mächler, 2011], rgdal [Keitt et al., 2012], and scatterplot3d [Ligges and Mächler, 2003]. The authors and maintainers of this software are gratefully acknowledged.

References


Keitt, T. H., R. Bivand, E. Pebesma, and B. Rowlingson (2012), rgdal: Bindings for the Geospatial Data Abstraction Library, R package version 0.7–18.


