Multivariate quantiles in hydrological frequency analysis

F. Chebana* and T.B.M.J. Ouarda

Industrial Chair in Statistical Hydrology/Canada Research Chair on the Estimation of Hydrometeorological Variables,
INRS-ETE, 490 rue de la Couronne, Quebec (QC),
Canada G1K 9A9

***Corresponding author:** Tel: (418) 654-2542
Fax: (418) 654-2600
Email: fateh_chebana@ete.inrs.ca

April 14th 2009

Revised version submitted to Environmetrics

Abstract

Several hydrological phenomena are described by two or more correlated characteristics. These dependent characteristics should be considered jointly to be more representative of the multivariate nature of the phenomenon. Consequently, probabilities of occurrence cannot be estimated on the basis of univariate frequency analysis (FA). The quantile, representing the value of the variable(s) corresponding to a given risk, is one of the most important notions in FA. The estimation of multivariate quantiles has not been specifically treated in the hydrological FA literature. In the present paper, we present a new and general framework for local FA based on a multivariate quantile version. The multivariate quantile offers several combinations of the variable values that lead to the same risk. A simulation study is carried out to evaluate the performance of the proposed estimation procedure and a case study is conducted. Results show that the bivariate estimation procedure has an analogous behaviour to the univariate one with respect to the risk and the sample size. However, the dependence structure between variables is ignored in the univariate case. The univariate estimates are obtained as special combinations by the multivariate procedure and with equivalent accuracy.

1. Introduction and literature review

Serious economic and social consequences are generally associated to extreme hydrological events, such as floods, storms and droughts. It is hence of high importance to develop the appropriate models for the prediction of such events. Frequency analysis (FA) procedures are commonly used tools for the analysis of extreme hydrological events. Relating the magnitude of extreme events to their frequency of occurrence is the principal aim of FA. This relationship can be obtained through the use of probability distributions (Chow et al., 1988).

Generally, hydrological events are characterized by several correlated variables, for instance, flood volume, peak and duration (e.g., Ashkar, 1980; Yue et al., 1999; Shiau, 2003; De Michele et al., 2005; Zhang and Singh 2006); storm duration and intensity (e.g., Salvadori and De Michele, 2004); and drought volume, duration and magnitude (e.g., Ashkar et al., 1998; Kim et al., 2003). Multivariate FA has recently attracted increasing attention and the importance of jointly considering all variables that characterize an event was clearly pointed out. However, the quantile notion was not studied appropriately in hydrological FA in a multivariate context. The univariate aspects of FA have been studied extensively, see e.g. Stedinger and Tasker (1986), Cunnane (1987) and Rao and Hamad (2000).

Justifications for adopting the multivariate framework to treat extreme events were discussed in several studies. In bivariate FA, Yue et al. (1999) concluded that single-variable hydrological FA can only provide limited assessment of extreme events. Indeed, univariate FA cannot provide a complete assessment of the probability of occurrence if the underlying hydrological event is described by a set of correlated random variables. The joint study of the probabilistic characteristics of such events, through their joint distribution, leads to a better understanding of the phenomenon. It was also outlined in Shiau (2003) that multivariate FA requires considerably more data and more sophisticated mathematical analysis. Univariate FA can be useful when only one random variable is significant for design purposes or when the two random variables are less dependent. However, the analysis of each random variable separately cannot reveal the significant relationship between them. Therefore, it is of importance to jointly consider the underlying random variables. Salvadori et al. (2007) pointed out that if one variable is significant in the design process, then univariate FA may be applied. Otherwise, univariate FA cannot provide complete assessment of the probability of occurrence.

The multivariate hydrological FA literature mainly treated one or more of the following three elements: (1) showing the importance and explaining the usefulness of the multivariate framework, (2) fitting the appropriate multivariate distribution (copula and marginal distributions) in order to model extreme events, and estimating the corresponding parameters, and (3) defining and studying bivariate return periods. The quantile is an important and extensively studied notion in univariate hydrological FA. However it was not properly addressed in the multivariate FA framework. For a random variable that represents the magnitude of an event that occurs at a given time and at a given site, the quantile function expresses the magnitude of the event in terms of its exceedence or non-exceedence probability. These probabilities are also associated to return periods. The goal of FA is to obtain reliable estimates of the quantiles corresponding to return periods of specific relevance (Rao and Hamed, 2000).

The usual univariate quantile can be extended to the multivariate setting in several ways (see Serfling, 2002 and Belzunce et al., 2007). One of the main difficulties of multivariate quantile extensions is related to the interpretation of the obtained quantile values. The objectives of the present paper are to introduce multivariate quantiles in hydrological FA, to adapt them to the resolution of hydrological problems, to interpret their significance and to study their properties. We also propose an estimation procedure for multivariate quantiles and establish the link with the univariate framework. To reach the above objectives, the multivariate quantile version presented by Belzunce et al. (2007) is adopted in the present paper. It possesses several advantages: it is simple, intuitive, interpretable and probability-based (rather than analytic, algebraic or geometric).

The paper is organized as follows. In Section 2, we present a short review of multivariate quantiles in the statistical literature. Section 3 presents multivariate quantiles in hydrology with an adaptation of the proposed procedure to floods in Section 4. Section 5 contains a simulation study. We present some properties of multivariate quantiles including a comparison between univariate and bivariate quantiles in Section 6. Section 7 contains an application of the procedure to a case study. Conclusions and directions for future work are reported in the last section.

2. Multivariate quantiles in statistical literature

In the statistical literature, several studies proposed to extend the well-known univariate quantile to higher dimensions. Serfling (2002) presented a review and a classification of some of these multivariate quantile versions. According to this classification, there are two major categories of multivariate quantiles: vector- and real-valued quantiles.

The vector-valued category contains four classes:

- *Multivariate quantiles as inversions of mappings:*

In the univariate setting, a quantile is defined as an inversion of the corresponding cumulative distribution function. For a random vector X having an absolutely continuous distribution F on $\mathbb{R}^d$ $d > 1$, a multivariate quantile is defined as the inverse of the mapping (see Koltchinskii and Dudley, 1996):

$$t \to -G_F(t) = E\left\{(X-t)/\|X-t\|\right\} \text{ from } \mathbb{R}^d \text{ to } \mathbb{R}^d \quad (1)$$

- *Multivariate quantiles based on norm minimization:*

This kind of multivariate quantiles is developed by Abdous and Theodorecu (1992) and Chaudhuri (1996). This extension corresponds to the following characteristic of the univariate *p*th-quantile (see Ferguson, 1967):

The quantile corresponds to the value of θ that minimizes $E\left\{|Z-\theta| + (2p-1)(Z-\theta)\right\}$ for a random variable Z with $E|Z| < \infty$. Several forms of the function to be optimized lead to several multivariate quantile functions.

- *Multivariate quantiles based on depth functions:*

One of the quantile features is that it is defined through “order statistics” by ordering the sample. Depth functions are mainly introduced to define an outward ordering in a multivariate sample. Hence, multivariate quantile functions can be defined through the use of depth functions. The reader is referred to Zuo and Serfling (2000) for a review regarding depth-functions and to Chebana and Ouarda (2008) for an adaptation and application in hydrology.

- *Data-based multivariate quantiles based on gradients:*

This version extends the property of the median which minimizes over θ the function $D(\theta) = \sum_i |X_i - \theta|$, or equivalently, it is a zero of the gradient $S(\theta) = -\sum_i \text{sng}(X_i - \theta)$.

Extension to the multivariate context considers various choices of $D(.)$ and the corresponding gradients $S(.)$, for example $D_1(\theta) = \sum_i \|X_i - \theta\|_1$ (see Hettmansperger et al., 1992).

In the real-valued quantile category, we find only one class. It is related to *the generalized quantile processes* defined as follows. Let P be a probability distribution on $\mathbb{R}^d$, C a subclass of Borel sets and λ a real-valued function, then this quantile function is given by:

$$U(p) = \inf\left\{\lambda(c); c \in C : P(c) \geq p\right\} \qquad (2)$$

Generalized quantile processes were introduced by Einmahl and Mason (1992). Some examples and applications are given in Serfling (2002). This version is more complex than the above ones, since it is general and valid even for discrete random variables.

Recently, Belzunce et al. (2007) defined another bivariate vector-valued quantile version. This version is not included in the review by Serfling (2002) and is focused on the bivariate context. Let (X,Y) be an absolutely continuous random vector and $p \in]0,1[$. The pth bivariate quantile set or bivariate quantile curve for the direction ε is defined as:

$$Q_{X,Y}(p,\varepsilon) = \left\{(x,y) \in \mathbb{R}^2 : F_\varepsilon(x,y) = p\right\} \qquad (3)$$

where $F_\varepsilon(x,y)$ is one of the following probabilities :

$F_{\varepsilon++}(x,y) = \Pr\left\{X \geq x,\ Y \geq y\right\}, F_{\varepsilon+-}(x,y) = \Pr\left\{X \geq x,\ Y \leq y\right\}, F_{\varepsilon--}(x,y) = \Pr\left\{X \leq x,\ Y \leq y\right\}$ and $F_{\varepsilon-+}(x,y) = \Pr\left\{X \leq x,\ Y \geq y\right\}$.

Note that equation (3) describes four quantile curves. Each one of these curves corresponds to one of the four quadrant events: Simultaneous exceedence $\left\{X \geq x,\ Y \geq y\right\}$, exceedence-non-exceedence $\{X \geq x, Y \leq y\}$, non-exceedence-exceedence $\{X \leq x, Y \geq y\}$ and simultaneous non-exceedence $\{X \leq x, Y \leq y\}$.

3. Multivariate quantiles in hydrology

In the present section we focus on the bivariate case for simplicity and clarity. However, all the elements of the developments can be defined and obtained in higher dimensions. In the bivariate case, we assume that X and Y are two random variables with joint distribution F, marginal distributions F_X and F_Y respectively and copula C (copulas are presented in Appendix A1). The variables X and Y represent the characteristics of a hydrological phenomenon.

3.1. Quantiles

In multivariate FA the focus was made on the multivariate return period (e.g., Shiau, 2003 and Salvadori et al., 2007). To our knowledge, the notion of multivariate quantiles is not employed in hydrology. The bivariate quantile version given in (3) is selected to be employed in the present paper. Aside from its simplicity and intuitivity, this quantile version does not require any symmetry assumption and the bivariate distribution (copula and margins) appears in its evaluation. Furthermore, this quantile version is probability-based (convenient for risk evaluation) rather than analytical or geometrical. In other words, the bivariate quantile (3) is a curve corresponding to any combination (*x*,*y*) that satisfies $F_\varepsilon(x, y) = p$ (an infinity of combinations).

Using Skalr's result (equation A1), expression (3) can be simplified. It can be obtained for the uniform margins and then transformed using the univariate marginal quantile function and the copula. Indeed, for instance when considering the event $\{X \le x, Y \le y\}$, the quantile curve can be expressed as follows:

$$Q_{X,Y}(p) = \left\{(x, y) \in \mathbb{R}^2 \text{ such that } x = F_X^{-1}(u), y = F_Y^{-1}(v); u, v \in [0,1] : C(u,v) = p\right\} \qquad (4)$$

In the bivariate setting, among the four simultaneous events described above, the simultaneous exceedence $\{X \ge x,\ Y \ge y\}$ and simultaneous non-exceedence $\{X \le x,\ Y \le y\}$ would be of interest in hydrology. This is mainly so because of the positive correlation, generally observed between the variables *X* and *Y*. Salvadori et al. (2007, page 127) indicated that, when investigating droughts, the event $\{X \le x,\ Y \le y\}$ could be of interest, whereas the event $\{X \ge x,\ Y \ge y\}$ is important if floods are considered. On the other hand, it is indicated in the literature (e.g., Shiau, 2003 and Salvadori et al., 2007), that the event $\{X \ge x,\ Y \ge y\}$ is of interest especially when the focus is on the evaluation of return periods. However, when the focus is on the evaluation of quantiles, the event $\{X \le x,\ Y \le y\}$ is of more interest, just as it is the case in the univariate setting (e.g., Hosking and Wallis, 1997).

The quantile curve is composed of two parts: the naïve part (tail) and the proper part (central). The naïve part is composed of two segments starting at the end of each extremity of the proper part. In the remainder of the paper, the term "quantile curve" refers to the proper part of the curve, unless indicated otherwise. The usual univariate quantiles are special cases of the bivariate quantile curve given in (3). The univariate quantiles represent the extreme points of the proper part of the bivariate quantile curve as illustrated in Figure 1. More details and explanations regarding these elements are given in Section 6.

For convenience, the following notations are employed throughout the paper: QC_p is the bivariate quantile curve associated to a risk p of the considered event on variables X and Y; $Q_{x,y}(p)$ represents a point (a combination) of the curve QC_p; $QC_x(p)$ and $QC_y(p)$ are the coordinates of the point $Q_{x,y}(p)$, that is $Q_{x,y}(p)=\left(QC_x(p),QC_y(p)\right)$. The univariate quantiles are denoted as $QD_X(p)$ and $QD_Y(p)$ when directly evaluated and $QL_X(p)$ and $QL_Y(p)$ when deduced as extreme values from the bivariate quantile curve. These notations are illustrated in Figure 1 for the non-exceedence event.

3.2. Quantile estimation procedure

In practice the true quantile is unknown and hence should be estimated. One can proceed by fitting a bivariate distribution, estimating its parameters and then obtaining the estimated quantile curve. More explicitly, given a bivariate sample the procedure is composed of the following steps:

1. Fit a multivariate distribution to the data set:
    a. Fit a copula to the data set;
    b. Fit marginal distributions for each variable separately;
2. Estimate the distribution parameters:
    a. Estimate the parameters of the copula of step 1.a;
    b. Estimate the parameters of each marginal distribution of step 1.b;

3. Specify the event of interest according to the phenomenon being studied and the specific application (e.g., $\{X \le x, Y \le y\}$);
4. Estimate the different quantile combinations $Q_{x,y}(p)$ that constitute the quantile curve for a given risk p in (0,1);
5. Select the appropriate combination(s) for the specific application.

To deal with step 1 in the described procedure, goodness-of-fit tests are required for the copula as well as for the marginal distributions. More precisely, these statistical tests deal with composite null hypotheses and focus on a specific parametric class of distributions (copula and margins). In a composite null hypothesis, we assume, for instance, that the copula belongs to the logistic Gumbel class and the marginal distributions are in the Generalized Extreme Value class. Such tests are well-known in the literature for univariate distributions. For instance, the empirical cumulative distribution function given by Cunnane (1978) can be used. Some statistical tests (numerical or graphical) have also been developed to treat copula's goodness-of-fit (see, e.g. Genest and Rivest, 1993; Fermanian, 2005 and Genest et al., 2009).

Once the parametric class of distributions (copula and margins) is identified from step 1, the corresponding parameters should be estimated. To estimate the distribution parameters (step 2), several methods exist in the literature especially in the univariate setting: For instance, the method of moments, the maximum likelihood method (e.g., Johnson, et al., 1995), the generalized method of moments (e.g., Ashkar and Ouarda, 1996), the *L*-moments method (Hosking and Wallis, 1997), the generalized maximum likelihood approach (Martins and Stedinger, 2000) and mixed methods (Chebana et al., 2008). Regarding the parameters of copulas, general estimation methods, such as the maximum likelihood method and the method of moments, can be applied. For instance, in the case of bivariate Archimedean copulas, Genest and Rivest (1993) employed a method of moments based on Kendall's tau coefficient to estimate the dependence parameter.

Note that the procedure presented above is parametric which is commonly used in hydrological FA. Nonparametric approaches have been employed in hydrological FA in the univariate context (see e.g., Adamowski and Feluch, 1990; Ouarda et al., 2001). However, Singh and Strupczewski (2002) reported that nonparametric methods are of limited use for the hydraulic design of major structures.

Even though the above estimation procedure is presented in the bivariate setting, it can be defined when more than two variables are involved to characterize the phenomenon. In the following we state the required elements as well as some difficulties that may arise when the multivariate setting is considered. Let $\left(X_1,...,X_d\right)$ be a random vector defined on $\mathbb{R}^d$, $d \geq 1$, with joint distribution F and marginal distributions $F_1,...,F_d$. In this setting, Sklar's theorem expresses the existence of a copula C that meets the condition $F(x_1,...,x_d) = C\left(F_1(x_1),..., F_d(x_d)\right)$ for real $x_1,...,x_d$. To define the analogous of the quadrant events, let I and J be two subsets, which may be empty, that constitute a partition of $\left\{1,2,...,d\right\}$. The events of interest are of the form $E_{I,J}^d = \left\{X_i \leq x_i,\, X_j \geq x_j \text{ for } i \in I, j \in J\right\}$. The number of these events is 2^d (four in the bivariate setting). The quantile version (given in (3) or (4)) can be defined in a multivariate context instead of the bivariate one. For instance, assume the event of interest is $E^d_{I,\varnothing} = \{X_1 \le x_1, \ldots, X_d \le x_d\}$. Then the corresponding multivariate quantile can be given by:

$$Q_{X_1,\ldots,X_d}(p) = \left\{(x_1,\ldots,x_d) \in R^d \text{ such that } x_j = F_j^{-1}(u_j); u_j \in [0,1], j = 1,\ldots,d : C(u_1,\ldots,u_d) = p\right\} \quad (5)$$

For d = 3, there are 8 possible joint events $E^3_{I,J}$, such as $E^3_{\{2\},\{1,3\}} = \{X_1 \ge x_1, X_2 \le x_2, X_3 \ge x_3\}$. The corresponding multivariate quantiles represent, for each event, a surface in a three-dimensional space.

Therefore, all theoretical elements required to define the procedure in a d-dimensional space are available. However, in practice, some difficulties may arise. The effective modeling of the multivariate copula is an important element of the analysis. Indeed, even though some well-known classes of Archimedean copulas and extreme value copulas are available in the multivariate setting, they are not convenient to model complex dependence structures. Defining and fitting other kinds of copulas, for $d \ge 3$, is a subject of interest and continuous development. The number of parameters, to be estimated for the copula and each marginal distribution grows quickly with the dimension d and hence increases the related uncertainty. The complexity of the considered copula and the numerical difficulties encountered in the bivariate setting, such as the resolution of equation (3), become even more important when the dimension of the problem increases.

4. Adaptation to floods

The multivariate quantile estimation procedure may be applied to several hydrological phenomena, such as droughts, storms and floods. In this section, the multivariate procedure is adapted to flood events. That consists in specifying the variables of interest, identifying the appropriate copula and the marginal distributions, estimating their parameters and stating the quantile curves more explicitly.

4.1. Flood characteristics

Floods are mainly described through three variables obtained from the corresponding hydrograph, that is their volume *X*, peak *Y* and duration *Z*. Figure 2 illustrates a typical flood hydrograph with these characteristics. It was shown in several studies that flood peak and volume are highly correlated as well as flood volume and duration, but flood peak and duration are not significantly correlated (see e.g., Yue el al., 1999). In the present section the bivariate volume and peak vector (*X*,*Y*) is considered.

4.2. Bivariate distribution

In the literature, flood peaks and flood volumes are often marginally represented by a Gumbel distribution (e.g. Yue et al., 1999 and Shiau, 2003). The cumulative distribution function for a random variable *X* following a Gumbel distribution is given by:

$$F_X(x) = \exp\left\{-\exp\left(-\tfrac{x-\beta_X}{\alpha_X}\right)\right\},\ x \text{ real},\ \alpha_X > 0 \text{ and } \beta_X \text{ real} \qquad (6)$$

Archimedean copulas represent convenient multivariate models to describe the dependence structure for hydrological flood events (e.g. Salvadori and De Michele, 2004). More precisely, Zhang and Singh (2006) showed the superiority of the Gumbel logistic copula for modeling flood volume and peak dependence. The copula representing the Gumbel logistic model is expressed according to the following formula:

$$C_\gamma(u,v) = \exp\left\{-\left[(-\log u)^\gamma + (-\log v)^\gamma\right]^{1/\gamma}\right\},\ \gamma \geq 1 \text{ and } 0 \leq u,v \leq 1 \qquad (7)$$

where γ is the dependence parameter. The Gumbel logistic copula C_γ is an Archimedean copula with generator function $\psi(t)=\left(-\log t\right)^\gamma, 0<t<1$. It is also an extreme value copula with dependence function $A(t)=\left((1-t)^\gamma+t^\gamma\right)^{1/\gamma}$ (see Appendix A1).

4.3. Estimation of the parameters

Several methods are available in the literature to estimate the parameters α_X and β_X of the marginal Gumbel distribution, for instance, the *L*-moment method (Hosking and Wallis, 1997) and the maximum likelihood method (e.g., Johnson et al., 1995).

The parameter γ of the Archimedean copula C_γ can be expressed as a function of the correlation coefficient ρ and the Kandall's tau coefficient. Gumbel and Mustafi (1967) expressed γ as a function of the correlation coefficient ρ as:

$$\gamma=\frac{1}{\sqrt{1-\rho}},\quad 0\le\rho<1 \tag{8}$$

Genest and Rivest (1993) provided the equation of γ as a function of the Kandall's tau coefficient $\tau=4E\left[F(X,Y)\right]-1$:

$$\gamma=1+\frac{\tau}{1-\tau} \tag{9}$$

4.4. Bivariate quantile curves

For an Archimedean copula with a generator function φ and a given value of $p\in\left]0,1\right[$, the quantile curve given by (3) corresponding to the event $\left\{X\le x, Y\le y\right\}$ is given by:

$$QC_p = \left\{ Q_{x,y}(p) = \left(QC_x(p), QC_y(p) \right); \varphi\left(F_Y(QC_y(p)) \right) = \varphi\left(p \right) - \varphi\left(F_X(QC_x(p)) \right) \right\} \qquad (10)$$

according to the notation given in Section 3.1. More explicitly, for the Gumbel logistic copula, the generator φ should be replaced by $\psi(t) = \left(-\log t\right)^{\gamma}$ in expression (10) and both F_X and F_Y by the expression of equation (6).

In expression (10), the event being considered is the simultaneous non-exceedence for both variables *X* and *Y*. Other events can also be of interest in hydrology and are studied in the literature (e.g. Salvadori et al., 2007), such as $\{X \geq x\,,\ Y \geq y\}$, $\{X \leq x \text{ or } Y \leq y\}$ and $\{X \geq x \text{ or } Y \geq y\}$. The corresponding quantile curves can be obtained using some probabilistic manipulations and depend only on the copula and the marginal distributions. We have, for instance,

$$\Pr\{X \geq x\,,\ Y \geq y\} = 1 - F_X(x) - F_Y(y) + C\left(F_X(x), F_Y(y)\right)$$

$$\begin{aligned} \Pr\{X < x \text{ or } Y < y\} &= 1 - \Pr\{X \geq x\,,\ Y \geq y\} \\ &= F_X(x) + F_Y(y) - C\left(F_X(x), F_Y(y)\right) \end{aligned}$$

Without loss of generality, in the present paper we consider the simultaneous exceedence and non-exceedence events $\{X \geq x\,,\ Y \geq y\}$ and $\{X \leq x\,,\ Y \leq y\}$ respectively.

5. Simulation study

In order to evaluate the performance of the proposed procedure, a simulation study is carried out. In this section we present the generation procedure, the performance evaluation criteria and the obtained results.

5.1. Generated samples

The simulations deal with floods. We generate M= 10 000 samples representing the volume (X) and peak (Y) variables with sample sizes n = 30 and 60. According to Section 4, the generated samples are from a bivariate distribution composed by Gumbel margins and a Gumbel logistic copula given in (6) and (7) respectively. The considered parameters of the marginal distributions are $\alpha_X = 300.22$, $\beta_X = 1239.80$ and $\alpha_Y = 15.85$, $\beta_Y = 51.85$. For comparison purposes, the parameter of the Gumbel logistic copula is taken to be $\gamma = 1, 1.414, 3.162$ (equivalent to $\rho = 0, 0.5, 0.9$ according to (8)). The sample generation is based on the algorithm developed by Ghoudi et al. (1998) and presented in Appendix A1.

The parameters are estimated using the L-moment method (Hosking and Wallis, 1997) and equation (8). The quantiles (bivariate and univariate) are obtained for the values of the risk p = 0.9, 0.99 and 0.995. The considered events are the simultaneous non-exceedence and exceedence events as indicated at the end of Section 4.4.

5.2. Performance evaluation criteria

Given the nature of bivariate quantiles (curves), the usual performance evaluation criteria do not apply and should be adapted. Basically, the evaluation consists in the assessment of the distance between the true and estimated quantile curves. In the present case, the quantile curve is a function. Consequently, the notation $(x, G_p(x))$ for the quantile curve is convenient for the definition of the evaluation criteria and will be adopted. Let M be the number of simulation repetitions, and let $\hat{G}_p^{[m]}(x)$ be a coordinate of the *m*th repetition of the estimated quantile for *p* ($0 < p < 1$). Then, the corresponding *point-wise* relative error is given by:

$$R_p^{[m]}(x) = \frac{\hat{G}_p^{[m]}(x) - G_p(x)}{G_p(x)} \tag{11}$$

Note that these relative differences represent vertical point-wise distances between the underlying curves. To be more interpretable, the point-wise relative errors (11) should be summarized with respect to *x* and *m*. For *x*, we consider distances or norms in functional spaces such as the L^r distances with $r \geq 1$. The L^r distances are defined as $\|f - g\|_r = \left(\int_S |f - g|^r \, d\lambda \right)^{1/r}$. They represent distances between two functions *f* and *g* on a given space *S* with a positive measure λ (see, e.g., Jones, 1993, Chapter 10). The L^1, L^2 and L^∞ are the most commonly used particular cases. It is shown (see, e.g., Jones, 1993) that $\| \; \|_r \leq \| \; \|_{r'}$ for $1 \leq r \leq r'$. Note that the L^1 distance is more intuitive and more representative than L^2 and L^∞. However, it is theoretically more complex to handle. Generally, estimations in FA are evaluated using relative bias (*RB*) and relative root-mean-square-errors (*RRMSE*). The use of L^1, L^2 and L^∞ distances does not allow to evaluate the *RB*. To evaluate the *RB*, we propose criteria based on the following relative integrated error:

$$RIE^{*[m]}(p) = \frac{1}{L_p} \int_{QC_p} R_p^{[m]}(x)\,dx, \qquad 0 < p < 1, \;\; m = 1, \ldots, M \tag{12}$$

where L_p is the length of the proper part of the true quantile curve QC_p.

The integral $RIE^{*[m]}(p)$ cannot define a norm since it may have negative values. The "pseudo-norm" associated to $RIE^{*[m]}(p)$ is denoted by L^{1*} since it is similar to L^1.

Regarding the RRMSE, the pseudo-norm L^{1*} is not appropriate since its values may be very small whereas the estimated and true curves are very different. Hence, it is convenient to evaluate the RRMSE on the basis of the following L^{1} distance given by:

$$RIE^{[m]}(p) = \frac{1}{L_p} \int_{QC_p} \left| R_p^{[m]}(x) \right| dx , \qquad 0 < p < 1, \;\; m = 1,\ldots,M \tag{13}$$

In (12) and (13), the length L_p is assumed to be non null. A discussion related to this point is presented in Section 6.

Then, in order to evaluate the estimation error on the M generated samples, the *RB* and the *RRMSE* are respectively based on $RIE^{*[m]}(p)$ and $RIE^{[m]}(p)$ and are given by:

$$RB(p) = 100 \frac{1}{M} \sum_{m=1}^{M} RIE^{*[m]}(p) \quad \text{and} \quad RRMSE(p) = 100 \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(RIE^{[m]}(p) \right)^2} \tag{14}$$

The *RB* based on $RIE^{*[m]}(p)$ is useful to indicate whether there is an over- or under-estimation.

5.3. Simulation results

Before presenting the simulation results, we illustrate true and estimated quantile curves as well as the value of the relative errors given in (11) for p = 0.9 and $\gamma = 1.414$. The relative errors are presented in Table 1 whereas Figure 3 illustrates the corresponding true and estimated quantile curves for two generated samples of size n = 30. For both samples, the relative errors associated to the univariate quantiles, directly evaluated or as extreme points, are similar for each variable. The bivariate quantile curve is over-estimated for the first sample and under-estimated for the second. We observe that the values of Table 1 are in agreement with the curves in Figure 3. We conclude that the relative errors reflect the obtained results. Therefore, the *RB* based on L^{1*}

can be employed, jointly with the *RRMSE* based on L^1, as convenient criteria in the multivariate FA setting. Note that all possible combined criteria (RB and RRMSE) and “norms” (L^{1*}, L^1 and L^2) were considered. They are not presented since they were not judged to provide additional information.

The univariate quantiles can be evaluated directly or as extreme points of the bivariate quantile curve. Note that similar errors do not imply similar estimated values obtained by both approaches. That is $(\hat{x} - x) \approx (\hat{y} - y)$ does not lead to $\hat{x} \approx \hat{y}$ unless we have $x \approx y$. Hence, it is useful to compare the true values of the univariate quantiles using both approaches. This is the object of Table 2 for the non-exceedence event using the parameters α_X, β_X, α_Y, β_Y and $\gamma = 1, 1.414, 3.162$. The corresponding relative differences are very low especially in the dependent cases. Therefore, in the remainder of the paper, the evaluations using both methods are considered to be almost equivalent.

Tables 3 and 4 illustrate the simulation results and correspond respectively to $n = 30$ and $n = 60$ results for both simultaneous exceedence and non-exceedence events. Several conclusions related to the variation of the relative errors can be deduced from these results with respect to different factors including sample size, dependence parameter, risk p and the type of events.

From the results, it can be seen that L^{1*} can be considered as a convenient indicator to evaluate the simulation performance. We observe that the RB is very small as it does not exceed 0.6% for the exceedence event and it is less than 1% for the non-exceedence event. It is generally observed in hydrological FA that the errors expressed in terms of the criteria RB and RRMSE for the quantiles are lower than those of the parameters. The low values of the quantile estimation RB can be explained by the effect of the compensation of parameter errors.

In general, we observe that the values of the performance criteria increase with respect to the risk p in both univariate and bivariate settings with some exceptions. In the univariate setting, the observed exception is related to the RRMSE of X when p increases from 0.99 to 0.995 in the simultaneous exceedence event when $n = 30$ (Table 3). However, when n takes the values 60, the RRMSE of X has an increasing behaviour (Table 4). Hence, this exception is due to the short sample size. In the univariate framework, the increase of the error with respect to the risk p is well known. This behaviour can be explained by the fact that a quantile corresponding to a small risk is close to the central body of the distribution. Therefore, an important part of the data contributes to its estimation. In the bivariate setting, exceptions are observed in the simultaneous non-exceedence event for both values of n (Tables 3 and 4). These exceptions are not due to the sample size as in the univariate case. They can be explained on the basis of the "curse of dimensionality". The curse of dimensionality means, in the present context, that the central part of a multivariate distribution contains little probability mass and samples tend to fall in the tails of the distribution. To explain this aspect, we consider a uniform distribution on the unit hypercube in $\mathbb{R}^d$ and we denote f_d as the fraction of the volume of the hypercube contained in the unit hypersphere. When the dimension d varies from 1 to 7, the fraction f_d takes respectively the values 1, 0.79, 0.52, 0.31, 0.16, 0.08 and 0.04 (Scott, 1992, Chapter 1). We observe that f_d decreases rapidly with respect to the dimension d. For more details and examples, the reader is referred to Scott (1992). In the present case where $d = 2$, the variation in the part of the data that contributes to the estimation is very small. Hence the *RRMSE*, which is expected to increase with respect to p, is seen to decrease slightly (less than 0.5%). This situation arises in the non-exceedence event for $\gamma = 1.414$ and $\gamma = 1$ if p increases from 0.99 to 0.995 (Tables 3 and 4).

The relative error variations are negligible for the univariate estimation with respect to the values of the dependence parameter γ. The reason is that the marginal distributions are not affected by the copula and the copula has always the same values in its extreme points, that is $C(u, 1) = u$ and $C(1,\mathrm{v}) = \mathrm{v}$ for all $u, \mathrm{v} \in I$ (see Appendix A1). However, for the bivariate setting we have two different situations. For the non-exceedence event, the RRMSE increases with respect to the dependence parameter γ. Whereas for the exceedence event, and for both values of n, the RRMSE can be considered constant with a slight increase of the RB with respect to γ. Figure 4 helps to explain the difference of behaviour of the RRMSE for the exceedence and non-exceedence events. It illustrates the three true quantile curves corresponding to the three values of γ for both exceedence and non-exceedence events. When comparing Figures 4b and 4c, we observe that the three curves are closer to each other in the exceedence event than in the non-exceedence event. Furthermore, we observe that the exceedence event curves are bounded by the zero axes and are shorter than those of the non-exceedence event.

When comparing the results of bivariate and univariate quantile estimation, we observe that:

- In the exceedence event: the RBs of X and Y are similar whereas the RRMSE related to Y quantiles is significantly larger than the RRMSE of X especially for small n and large p. The RB of the bivariate estimation is larger than the *RB* of Y. The RRMSEs of Y and the bivariate estimation are very close with slight differences of around 1%.

- In the non-exceedence event: the values of the RB and RRMSE related to X and Y are similar. The values of these criteria for the bivariate estimation are larger than those of each one of the univariate cases. The differences between them decrease when *n* increases.

In both cases of events, the RB and RRMSE of bivariate estimation are relatively larger than those of univariate estimation. This can be explained by the fact that the bivariate estimation includes the errors from the parameters of the marginal distributions as well as the errors from the dependence parameter of the copula. It is important to note that the *RB* and *RRMSE* are evaluated differently for the univariate and the bivariate settings.

We observe from Tables 3 and 4 that the sample size *n* has an effect on the results in several ways. First, the estimation becomes more accurate when *n* increases from 30 to 60 for all situations. Second, in the univariate setting, the increasing behaviour of the RB and RRMSE with respect to *p* is better respected when *n* increases. Third, in the non-exceedence event, the differences of RB and RRMSE between the bivariate and univariate estimations are reduced.

Finally, we conclude that the bivariate estimation procedure performs comparably to the univariate procedure in terms of RB and RRMSE behaviour with respect to the risk *p* and the sample size *n*. However, the univariate estimation does not consider the variation in the dependence structure of the phenomenon. In terms of the relative error values, it is important to notice that the univariate and bivariate procedures employ different performance criteria.

6. Multivariate quantile properties

In this section we present a set of general statements that are useful to explain bivariate quantiles as well as their relation with univariate quantiles. In order to reduce space, we only treat the simultaneous non-exceedence event. Other events can be treated similarly.

For p in]0, 1[, recall that $QD_X(p) = F_X^{-1}(p)$ and $QD_Y(p) = F_Y^{-1}(p)$ are the marginal quantiles and $\left(QC_x(p), QC_y(p)\right)$ is one of the bivariate quantile combinations such that $\Pr\left\{X \leq QC_x(p), Y \leq QC_y(p)\right\} = p$. The bivariate quantile gives several possible scenarios $\left(QC_x(p), QC_y(p)\right)$ which all lead to the same risk *p*. Note that the values of the coordinates $QD_x(p)$ and $QD_y(p)$ vary in opposite directions, to preserve the same risk *p*. We have necessarily $QD_X(p) \leq QC_x(p)$ and $QD_Y(p) \leq QC_y(p)$, since all quantities $\Pr\left\{X \leq QD_X(p)\right\}$, $\Pr\left\{Y \leq QD_Y(p)\right\}$ and $\Pr\left\{X \leq QC_x(p), Y \leq QC_y(p)\right\}$ are equal to p. In other words, the marginal quantiles correspond to the extreme scenarios related to the event: the smallest $QD_x(p)$ and the largest $QD_y(p)$ and vice versa. More explicitly, the univariate quantiles correspond to the particular combinations $\left(QD_X(p), \infty\right)$ and $\left(\infty, QD_Y(p)\right)$ in (3). Indeed, $QD_X(p)$ is defined such that $\Pr\left\{X \leq QD_X(p)\right\} = \Pr\left\{X \leq QD_X(p), Y < \infty\right\} = p$ where the maximum value of *Y* can be infinitely large, see illustration in Figure 1. On the basis of the bivariate distribution used in the simulation section with $\gamma = 1.414$, Figure 5 illustrates the corresponding quantile curves for three values of *p*. We observe from Figure 5 that the quantile curves are composed by two parts: a central part which corresponds to the “proper” combinations and a tail part which corresponds to the “naïve” combinations where the curve is constant. It is important to identify the first combination where the curve is constant on each axis as illustrated in Figure 5.

When p is very close to 0 and 1 in the exceedence and non-exceedence events the corresponding proper part of the quantile curve becomes respectively the combinations $\left(F_X^{-1}(0), F_Y^{-1}(0)\right)$ and $\left(F_X^{-1}(1), F_Y^{-1}(1)\right)$. It can be obtained since $C(0,0)=0$ and $C(1,1)=1$. The criteria given in (12) and (13) cannot be defined since the length *Lp* is null. If necessary, these criteria should be evaluated as "distances" between points instead of curves. The values $F_X^{-1}(0)$ and $F_X^{-1}(1)$ represent the support extremities of the marginal distribution of *X*. Note that $F_X^{-1}(0)$ and $F_X^{-1}(1)$ may be infinite. For a given sample, the combinations $\left(F_X^{-1}(0), F_Y^{-1}(0)\right)$ and $\left(F_X^{-1}(1), F_Y^{-1}(1)\right)$ represent respectively $\left(\min_i(X_i), \min_i(Y_i)\right)$ and $\left(\max_i(X_i), \max_i(Y_i)\right)$.

In the non-exceedence event $\{X \le x, Y \le y\}$, small values of the risk p correspond to a large number of possible scenarios in the practical sense but not in the mathematical sense. When the risk p increases, the number of such scenarios decreases, and hence the quantile curve becomes shorter. Therefore, the univariate and the bivariate quantile combinations become closer as points in the bidimensional space as illustrated in Figure 5.

In the multivariate context, for a given problem, we have "one" joint event, e.g. $\{X \le x, Y \le y\}$, to which we associate "one" risk level to be evaluated. However, in the univariate setting, for the same problem, each variable needs to be treated separately. That induces several events, e.g. $\{X \le x\}$ and $\{Y \le y\}$, and hence possibly several risk levels to evaluate. Furthermore, some events cannot be expressed in the univariate context. That situation occurs generally when the events are not of "rectangular" form. The univariate context can only provide the bounds of each variable without any information about the shape of the relation between the variables. Figure 6 illustrates a specific situation of an ellipse and a rectangle where the bounds of X and Y are the same for both shapes. The ranges corresponding to these two situations can be described precisely in the multivariate context.

It is clear that flood peak and volume values obtained by single-variable FA are significantly different from those obtained using the bivariate distribution. As it was indicated previously, floods are naturally multivariate phenomena. As a consequence, bivariate modeling is more realistic than the univariate one. This means that the realistic quantile values are those obtained from the bivariate distribution. Figure 7 illustrates this fact on a specific case. It presents the true 0.99-quantile curve of the bivariate distribution used in the simulation section with $\gamma = 1.414$ and the non-exceedence event. The realistic extreme combinations (volume, peak) corresponding to p = 0.99 are (x_1, y_2) = (2621, 176) and (x_2, y_1)= (3589, 125). However, the univariate quantile values are (volume, x_1= 2621) and (peak, y_1= 125). Therefore, the combination of the univariate values (x_1, y_1) = (2621, 125) corresponds to another risk p' smaller than p = 0.99 and hence may lead to the wrong conclusions. Note that the values x_2= 3589 and y_2=176 do not appear explicitly in the univariate context. The values x_2 and y_2, as univariate quantiles, correspond to a risk p''=1 since they represent the largest values of each variable. Hence, the combination (x_2, y_2) corresponds also to the risk p''=1. Table 5 quantifies the differences between univariate and bivariate quantile evaluations. This example shows that univariate estimation results should be used cautiously.

Numerical difficulties can be encountered to obtain combinations of the bivariate quantile from the equation $F(x, y) = p$. First, the resolution of this equation requires more running time when considering high values of p and especially when doing simulations. The reason is related to the thinness of the grid of the unit square that represents the range of copulas. The grid step is selected according to the value of p. For instance, in the present study, when $p = 0.9$, the grid step is 0.01 however this step becomes 0.003 for $p = 0.99$ and 0.0015 for $p = 0.995$. Second, when the copula representing the extreme event is not Archimedean, the bivariate quantile scenarios can be more complex to obtain. This difficulty occurs also in the univariate setting for some distributions.

7. Case study

The data set used in this case study is taken from Yue et al. (1999) and concerns the Ashuapmushuan basin located in the Saguenay region in the province of Québec, Canada. The flood volume (X) and peak (Y) were extracted from a daily streamflow data set from 1963 to 1995. The gauging station 061901 is near the outlet of the basin, at latitude 48.69°N and longitude 72.49°W. This region is characterized by a high spring-snowmelt flood season.

The Gumbel marginal distribution (6) often represents well extreme events such as flood peak and volume (e.g. Yue et al., 1999 ; Shiau, 2003). It is also shown in the present study that the Gumbel distribution can be selected as a marginal distribution for both peak and volume (see

Figure 8a,b). The corresponding parameter estimates are obtained using the probability weighted moment method, or equivalently the L-moment method, and are given by $\hat{\alpha}_X = 46262.0, \hat{\beta}_X = 10295.6$ and $\hat{\alpha}_Y = 1258.3, \hat{\beta}_Y = 291.4$. These estimates are very close to the ones obtained by Yue et al. (1999) using the method of moments.

Furthermore, according to Figure 8c, the Gumbel logistic copula (7) can be selected to fit the dependence structure of the data set on the basis of the function K and its estimation given in Appendix A1. The correlation coefficient between X and Y is $\rho = 0.60$ which leads to the estimation of the corresponding parameter $\hat{\gamma} = 1.57$ using (8).

The bivariate quantile curves are obtained using the procedure proposed in Section 3. Figure 9 presents the quantile curves corresponding to the simultaneous non-exceedence event with different risk values (p = 0.9, 0.99 and 0.995). In the simultaneous exceedence event the quantile values corresponding to large values of p are negative for the volume and the peak. This is statistically possible since the Gumbel distribution is defined for real values of the variable (equation (6)). However, physically this is not possible since the volume and peak are positive characteristics. In the univariate setting the exceedence event quantile associated to large values of p corresponds to the left tail of the Gumbel distribution. The left tail of the Gumbel distribution is generally of less interest in hydrology. Note that Yue et al. (1999), who used the data set of the present case study, treated only the non-exceedence event. In Figure 9, we present also some possible combinations including the univariate ones for the non-exceedence event for p = 0.9, 0.99 and 0.995. As indicated in Section 6, we can identify from Figure 9 the two parts that compose each quantile curve and also values of some particular combinations including the extreme ones.

8. Conclusions and future work

In the present paper we introduced the notion of multivariate quantile in hydrological FA. The extension of the quantile notion to high dimensions leads to several multivariate quantile versions. The selected version is simple, intuitive, probability-based and interpretable. Even though, the focus was on the bivariate context, the study can be conducted in higher dimensions with the appropriate adaptations. The bivariate quantile version developed in this study is a curve composed by several combinations with the same risk. The univariate estimated quantiles, correctly combined, are particular cases corresponding to the extreme scenarios of the bivariate quantile curve. Depending on the available resources and the nature of the project, one or more convenient scenarios may be selected. Hence, aside from being more accurate and realistic, the bivariate setting offers more flexibility to designers than the univariate framework. A parametric quantile estimation procedure is proposed. It was evaluated on the basis of a simulation study. The proposed procedure was also applied to a real world case study.

Results show that the estimation procedure performs better for large sample sizes in all considered situations. The univariate estimation does not take into account the dependence structure between variables and should be used cautiously. The relative errors of both bivariate and univariate estimations are of the same order of magnitude with similar behaviours with respect to sample size and risk. The multivariate procedure provides univariate quantile estimates that are very close to those obtained directly using the univariate procedure and also with equivalent precisions. Note that the performances of univariate and bivariate procedures are evaluated on the basis of different criteria. The main differences between univariate and bivariate estimations are conceptual.

Even though several insights are brought to the multivariate FA through the present study, other remaining issues deserve to be developed in future work such as:

- Study the impact of different factors that may have significant effects on estimation performances. This includes, for instance, the estimation method of the distribution parameters and the selection of the multivariate distribution.
- Develop a nonparametric estimation procedure and compare its results with the parametric one.
- Associate, to the estimated quantile curve, the corresponding confidence interval.
- Consider other classes of copulas since not all hydrological phenomena are necessarily modeled with Archimedean copulas.
- Develop regional multivariate FA models, such as the index-flood model, in order to treat the estimation in sites with short records or ungauged sites. Note that Chebana and Ouarda (2007) proposed discordancy and homogeneity statistical tests in the multivariate framework. These tests can be considered as a first step towards a regional estimation procedure.

Acknowledgments

Financial support for this study was graciously provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Canada Research Chair Program. The authors would like to thank Barbara Martel for her assistance. The authors wish also to thank the Editor-in-Chief and the anonymous reviewer whose comments helped improve the quality of the paper.

Appendix:

In this appendix, we present two useful notions for multivariate FA: copulas and bivariate return periods.

A1. Copulas

To describe the dependence structure between two or more random variables, the notion of copula is employed. It is independent of the marginal distributions and hence the marginal distributions may belong to different classes of distributions. Copulas have recently received increasing attention in various science fields (see for instance Nelsen, 2006). A function C: $I \times I \rightarrow I$ (I = [0, 1]) is said to be a copula if the following conditions are fulfilled :

- for all $u, v \in I$: $C(u, 0) = 0$, $C(u, 1) = u$, $C(0, \mathrm{v}) = 0$, and $C(1,\mathrm{v}) = \mathrm{v}$;
- for all $u_1, u_2, \mathrm{v}_1, \mathrm{v}_2 \in I$ $u_1 \leq u_2$ and $\mathrm{v}_1 \leq \mathrm{v}_2 : C(u_2,\mathrm{v}_2) - C(u_2,\mathrm{v}_1) - C(u_1,\mathrm{v}_2) + C(u_1,\mathrm{v}_1) \geq 0$

Sklar's theorem (Sklar, 1959) provides the relationship between a bivariate distribution on one hand and the corresponding copula and marginal distributions on the other hand. Sklar's result states that there exists a copula C such that:

$$F(x, y) = C\left(F_X(x),\ F_Y(y)\right) \text{ for all real } x \text{ and } y \qquad \text{(A1)}$$

where F is the joint distribution of X and Y and F_X and F_Y are their marginal distributions respectively. In addition, if F_X and F_Y are continuous, the copula C is unique.

Two classes of copulas are of particular interest in statistical and hydrological literature: Archimedean and Extreme Value (EV) copulas. A bivariate Archimedean copula is characterized by the expression:

$$C(u,v)=\psi^{-1}\left(\psi(u)+\psi(v)\right), \quad 0<u,v<1 \qquad \text{(A2)}$$

where the generator $\psi(.)$ is a convex decreasing function satisfying $\psi(1)=0$.

The class of EV copulas is defined on the basis of a dependence function *A* through the formula given by Pickands (1981) as:

$$C(u,v)=\exp\left\{\left(\log u+\log v\right)A\left(\frac{\log u}{\log u+\log v}\right)\right\}, \quad 0<u,v<1 \qquad \text{(A3)}$$

where the dependence function *A* is convex and defined on [0, 1] with $\max\{t,1-t\}\le A(t)\le 1$.

According to Genest and Rivest (1993), an Archimedean copula, with a generator function ψ, is characterized by the following function:

$$K_{\psi}(z)=z-\frac{\psi(z)}{\psi'(z)} \qquad \text{(A4)}$$

which can be estimated by:

$$\widehat{K}(z)=\frac{1}{N}\sum_{i=1}^{N}1_{[w_i\le z]} \quad \text{where} \quad w_i=\frac{1}{N-1}\sum_{t=1}^{N}1_{[x_1^t<x_1^i,\,x_2^t<x_2^i]}, \; i=1,..,N \qquad \text{(A5)}$$

for a given bivariate sample $(x_1^1,x_2^1),(x_1^2,x_2^2),...,(x_1^N,x_2^N)$.

The functions K_{ψ} and $\widehat{K}$ can be used for the fitting of an Archimedean copula to a bivariate sample.

To generate bivariate samples from the Gumbel logistic copula (7), we consider the algorithm developed by Ghoudi et al. (1998). For a bivariate vector (X, Y) following an extreme value copula (A3) with dependence function A and margins F_X and F_Y, the algorithm is summarized as follows. Let V_1 and V_2 be uniform random variables and Z be a random variable with a cumulative distribution function G_Z and probability density function g_Z given by $G_Z(z) = z + z(1-z)A'(z)/A(z),\ 0 \le z \le 1.$ This algorithm consists of the following steps:

1. Simulate Z;
2. Given Z, take $W = V_1$ with probability $p(Z)$ and $W = V_1 V_2$ with probability $1-p(Z)$, where $p(z) = z(1-z)A''(z)/\left(A(z)g_Z(z)\right)$;
3. Put $U_1 = W^{Z/A(Z)}$ and $U_2 = W^{(1-Z)/A(Z)}$;
4. Set $X = F_X^{-1}(U_1)$ and $Y = F_Y^{-1}(U_2)$

A2. Bivariate return period

The notion of return period for hydrological extreme events is commonly used in hydrological FA. The return period of a given event is defined as the mean of the probability of its occurrence as indicated e.g. in Rao and Hamed (2000). Note that the return period concept is an estimation of the probability or the risk whereas the quantile is the value of the variable leading to this risk.

The bivariate extreme hydrological event distributions and corresponding return periods have been extensively studied e.g. in Shiau (2003) and Salvadori et al. (2007). For instance, for the event $\{X > x, Y > y\}$, Salvadori et al. (2007) defined the bivariate return period as the positive number $T^{\wedge}_{x,y}$ given by

$$T^{\wedge}_{x,y} = \frac{1}{\Pr\{X > x, Y > y\}} \tag{A6}$$

Adapted definitions are also given for other events. The above definition concerns the annual maximum series. For partial duration series, definitions are also available in the literature, e.g. in Shiau (2003) or Salvadori et al. (2007). Finally, relationships between univariate return periods and the joint return period are also derived in Salvadori et al. (2007).

Bibliography

Abdous, B. and Theodorecu, R. (1992) Note on the spatial quantile of a random vector. *Statist. Probab. Lett.,* **13**, 333-336.

Adamowski, K. and Feluch, W. (1990) Nonparametric flood frequency analysis with historical information. *ASCE J. Hydraul. Engng*., **116**, 1035–1047.

Ashkar, F. (1980) *Partial duration series models for flood analysis*. PhD thesis, Ecole Polytechnique of Montreal, Montreal, Canada.

Ashkar, F. and Ouarda, T.B.M.J. (1996) On some methods of fitting the generalized Pareto distribution. *J. Hydrol.* **177**, pp. 117–141.

Ashkar, F.; El Jabi, N. and Issa, M. (1998) A bivariate analysis of the volume and duration of low-flow events. *Stoch. Hydrol. Hydraul.*, **12**, 97-116.

Belzunce, F.; Castaño, A.; Olvera-Cervantes, A. and Suárez-Llorens A. (2007) Quantile curves and dependence structure for bivariate distributions. *Comput. Statist. Data Anal.*, **51**, Issue 10, 5112-5129.

Chaudhuri, P. (1996) On a geometric notion of quantiles for multivariate data, *J. Amer. Statist. Assoc.*, **91**, 862-872.

Chebana, F. and Ouarda, T.B.M.J. (2007) Multivariate L-moment homogeneity test. *Water Resour. Res.*, **43**(8), W08406, doi:10.1029/2006WR005639.

Chebana, F. and Ouarda, T.B.M.J. (2008) Depth and homogeneity in regional flood frequency analysis. *Water Resour. Res.*, **44**, W11422, doi:10.1029/2007WR006771.

Chebana, F.; El Adlouni, S. and Bobée, B. (2008) Halphen distributions: mixed estimation methods and comparison. [In French] Research report R-994 ISBN: 978-2-89146-579-3, 83 pages.

Chow, V.T.; Maidment, D.R. and Mays, L.R. (1988) *Applied Hydrology*. McGraw-Hill, New York.

Cunnane, C. (1978) Unbiased plotting positions—a review. *J. Hydrology*, **37**, 205–222.

Cunnane, C. (1987) Review of statistical models for flood frequency estiamtion. *Hydrologic frequency modeling*, V. P. Singh, ed., Reidel, Dordrecht, The Netherlands, 49–95.

De Michele, C.; Salvadori, G.; Canossi, M.; Petaccia, A. and Rosso, R. (2005) Bivariate Statistical Approach to Check Adequacy of Dam Spillway. *J. Hydrologic Engrg.*, **10**, 50-57.

Einmahl, J. H. J. and Mason, D. M. (1992) Generalized quantile processes. *Ann. Statist.*, **20**, 1062-1078.

Fermanian, J-D. (2005) Goodness-of-fit tests for copulas. *J. Multivariate Anal.,* **95**, 119-152.

Ferguson, T.S. (1967) *Mathematical statistics: a decision theoretic approach*, Academic Press, New York.

Genest, C. and Rivest, L-P. (1993) Statistical Inference Procedures for Bivariate Archimedean Copulas. *J. Amer. Statist. Assoc.*, **88**, 1034-1043.

Genest, C. ; Rémillard, B. and Beaudoin, D. (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, **44**, In Press.

Ghoudi, K.; Khoudraji, A. and Rivest, L-P. (1998) Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.*, **26**, 87-197.

Gumbel, E.J. and Mustafi, C.K. (1967) Some analytical properties of bivariate extreme distributions. *J. Amer. Statist. Assoc.*, **62**, 569–588.

Hettmansperger T. P.; Nyblom, J. and Oja, H. (1992) On multivariate notions of sign and rank, in: Y. DODGE (ed.), L1-Statistical analysis and related methods, North-Holland, Amsterdam, 267–278.

Hosking, J. R. M. and Wallis, J. R. (1997) *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press.

Johnson, N. L.; Kotz, S. and Balakrishnan, N. (1995) *Continuous univariate distributions*. Vol. 1 & 2. Second edition. John Wiley & Sons, Inc., New York.

Jones, F. (1993) *Lebesgue integration on Euclidean space*. Jones and Bartlett Publishers, Boston, MA.

Kim, T.; Valdés, J. B. and Yoo, C. (2003) Nonparametric Approach for Estimating Return Periods of Droughts in Arid Regions. *J. Hydrologic Engrg.,* **8**, 237-246.

Koltchinskii, V. and Dudley, R. M. (1996) On spatial quantiles, unpublished manuscript.

Martins, E.S. and Stedinger, J.R. (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.*, **36** (3), 737-744.

Nelsen, R. B. (2006) *An introduction to copulas*. Springer-Verlag, New York. Second edition.

Ouarda, T. B. M. J.; Haché, M.; Bruneau, P. and Bobée, B. (2000) Regional Flood Peak and Volume Estimation in Northern Canadian Basins. *ASCE J. Cold. Reg. Engrg.,* **14**, 176-191.

Ouarda, T. B. M. J.; Girard, C.; Cavadias, G. S. and Bobée, B. (2001) Regional flood frequency estimation with canonical correlation analysis. *J. Hydrology,* **254,** 157-173.

Pickands, J. (1981) Multivariate extreme value distributions. In *Bulletin of the International Statistical Institute*: Proceedings of the 43rd Session (Buenos Aires), pp. 859-878. Voorburg, Netherlands: ISI.

Rao, A.R. and Hamed, K.H. (2000) *Flood Frequency Analysis*. CRC Press, Boca Raton.

Salvadori, G.; De Michele, C.; Kottegoda, N.T. and Rosso, R. (2007) *Extremes in Nature: An Approach Using Copulas.* Springer

Salvadori, G. and De Michele, C. (2004) Analytical calculation of storm volume statistics involving Pareto-like intensity-duration marginals. *Geophys. Res. Lett.,* **31**, L04502.1-L04502.4.

Scott, D. W. (1992) *Multivariate Density Estimation, Theory, Practice and Visualization*. Wiley New York.

Serfling, R. (2002) Quantile functions for multivariate analysis: approaches and applications. *Statist. Neerlandica,* **56**, 214-232.

Shiau, J. T. (2003) Return period of bivariate distributed extreme hydrological events. *Stoch. Environ. Res. Risk. Assess.,* **17**, 42-57.

Singh, V. P. and Strupczewski, W. G. (2002) On the status of flood frequency analysis. *Hydrol. Process.*, **16**, 3737-3740.

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

Stedinger, J.R. and Tasker, G. (1986) Regional hydrologic analysis, 2, Model-error estimators, estimation of sigma and log Pearson type 3 distributions. *Water Resour. Res.*, **22**, 1487-1499.

Yue, S.; Ouarda, T. B. M. J.; Bobée, B.; Legendre, P. and Bruneau, P. (1999) The Gumbel mixed model for flood frequency analysis. *J. Hydrology,* **226**, 88-100.

Zhang, L. and Singh, V. P. (2006) Bivariate Flood Frequency Analysis Using the Copula Method. *J. Hydrologic Engrg.,* **11**, 150-164.

Zuo, Y. and Serfling, R. (2000) General notions of statistical depth function. *Ann. Statist.,* **28** 461–482.

Table 1: Relative errors (%) of the 0.9-quantile estimations corresponding to two generated samples. The univariate quantiles are evaluated directly and as extreme points of the bivariate quantile curves. The relative errors $RIE^*(p)$ of the bivariate quantiles are evaluated using equation (11)

	1st sample	2nd sample
$RIE^*(p)$ for QC_p	2.44	-3.21
Relative error for QL_X	3.41	-3.78
Relative error for QD_X	2.81	-3.77
Relative error for QL_Y	-1.69	0.20
Relative error for QD_Y	-2.19	0.15

Table 2: Comparison of the true values of the univariate quantiles evaluated directly and as extreme points of the bivariate quantile curve using the parameters α_X, β_X, α_Y, β_Y and γ =1, 1.414, 3.162 for the non-exceedence event.

			Direct	As extreme point	Relative difference* (%)
γ = 1	p = 0.9	X	1915.40	1945.50	1.5712
		Y	87.52	89.1071	1.8154
	p =0.99	X	2620.90	2652.30	1.1999
		Y	124.76	126.42	1.3307
	p =0.995	X	2829.70	2896.50	2.3595
		Y	135.79	139.31	2.5959
γ = 1.414	p = 0.9	X	1915.40	1923.20	0.4073
		Y	87.52	87.93	0.4706
	p =0.99	X	2620.90	2629.20	0.3168
		Y	124.76	125.20	0.3514
	p =0.995	X	2829.70	2852.80	0.8153
		Y	135.79	137.01	0.8970
γ = 3.162	p = 0.9	X	1915.40	1915.50	0.0029
		Y	87.52	87.52	0.0034
	p =0.99	X	2620.90	2620.90	0.0025
		Y	124.76	124.77	0.0027
	p =0.995	X	2829.70	2830.30	0.0223
		Y	135.79	135.82	0.0245

* Relative difference = 100 (As extreme point-Direct)/Direct

Table 3: Relative errors (%) of univariate quantiles evaluated directly and relative errors (%) of the bivariate quantile curve for the simultaneous non-exceedence and exceedence events when n=30

			p=0.9		***p=0.99***		***p=0.995***	
			RB	*RRMSE*	*RB*	*RRMSE*	*RB*	*RRMSE*
Exceedence	$\gamma = 3.162$	QD_X	0.09	6.35	-0.39	9.75	-0.34	9.62
		QD_Y	0.10	8.63	-0.44	15.13	-0.06	16.38
		Biv^*	0.92	9.32	2.24	14.90	2.94	16.17
	$\gamma = 1.414$	QD_X	-0.15	6.35	-0.43	9.77	-0.58	9.65
		QD_Y	-0.21	8.71	-0.40	15.96	-0.32	18.19
		Biv^*	0.53	10.16	1.19	15.78	1.40	17.53
	$\gamma = 1$	QD_X	0.02	6.41	-0.40	9.76	-0.46	9.68
		QD_Y	0.04	8.60	0.04	16.13	-0.26	18.77
		Biv^*	0.67	9.98	0.89	15.59	0.79	17.58
Non-exceedence	$\gamma = 3.162$	QD_X	-0.02	7.09	0.04	9.52	-0.04	9.96
		QD_Y	-0.02	8.17	0.03	10.55	-0.01	10.91
		Biv^*	0.59	13.04	0.65	16.66	0.46	16.76
	$\gamma = 1.414$	QD_X	0.03	7.09	-0.04	9.36	0.11	9.93
		QD_Y	0.16	8.27	-0.02	10.43	0.20	10.88
		Biv^*	0.40	12.32	0.27	15.63	0.36	15.07
	$\gamma = 1$	QD_X	-0.04	7.16	-0.12	9.44	0.03	9.99
		QD_Y	0.06	8.26	0.15	10.63	0.02	11.04
		Biv^*	-0.09	11.09	-0.09	13.47	-0.16	12.88

*The *RB* and *RRMSE* are evaluated using respectively $RIE^{*[m]}(p)$ and $RIE^{[m]}(p)$

Table 4: Relative errors (%) of univariate quantiles evaluated directly and relative errors (%) of the bivariate quantile curve for the simultaneous non-exceedence and exceedence events when n=60

			p=0.9		***p=0.99***		***p=0.995***	
			RB	*RRMSE*	*RB*	*RRMSE*	*RB*	*RRMSE*
Exceedence	$\gamma = 3.162$	QD_X	0.06	4.48	-0.00	7.50	-0.11	8.01
		QD_Y	0.06	6.08	0.12	11.25	-0.11	12.62
		Biv^*	0.57	6.88	1.93	11.79	2.11	12.86
	$\gamma = 1.414$	QD_X	-0.01	4.50	0.02	7.54	-0.11	7.99
		QD_Y	-0.05	6.06	0.25	11.28	0.19	13.12
		Biv^*	0.44	7.49	1.38	11.90	1.57	13.35
	$\gamma = 1$	QD_X	0.05	4.50	-0.07	7.47	-0.18	7.94
		QD_Y	-0.07	6.06	0.00	11.36	0.11	13.13
		Biv^*	0.39	7.31	0.67	11.46	0.85	12.81
Non-exceedence	$\gamma = 3.162$	QD_X	-0.01	5.08	0.04	6.65	0.04	7.05
		QD_Y	-0.01	5.87	0.00	7.38	0.03	7.78
		Biv^*	0.41	9.59	0.46	11.95	0.37	12.40
	$\gamma = 1.414$	QD_X	0.03	5.08	-0.05	6.62	0.04	6.94
		QD_Y	-0.04	5.90	0.02	7.30	-0.03	7.66
		Biv^*	0.14	8.99	0.26	11.30	0.09	10.94
	$\gamma = 1$	QD_X	0.07	5.00	0.03	6.61	-0.00	7.06
		QD_Y	0.04	5.78	0.05	7.38	0.02	7.76
		Biv^*	0.07	7.90	0.03	9.68	-0.12	9.41

*The *RB* and *RRMSE* are evaluated using respectively $RIE^{*[m]}(p)$ and $RIE^{[m]}(p)$

Table 5: Comparison of univariate and bivariate flood 0.99-quantiles on the basis of the bivariate distribution used in the simulation with $\gamma = 1.414$

Obtained values	Q =125 (m^3/s)		V =2621 (day.m^3/s)	
	V single	V joint	Q single	Q joint
Associated values	2621	3589	125	176
Relative differences	-26.97%=(2621-3589)/3589		-28.98%=(125-176)/176	

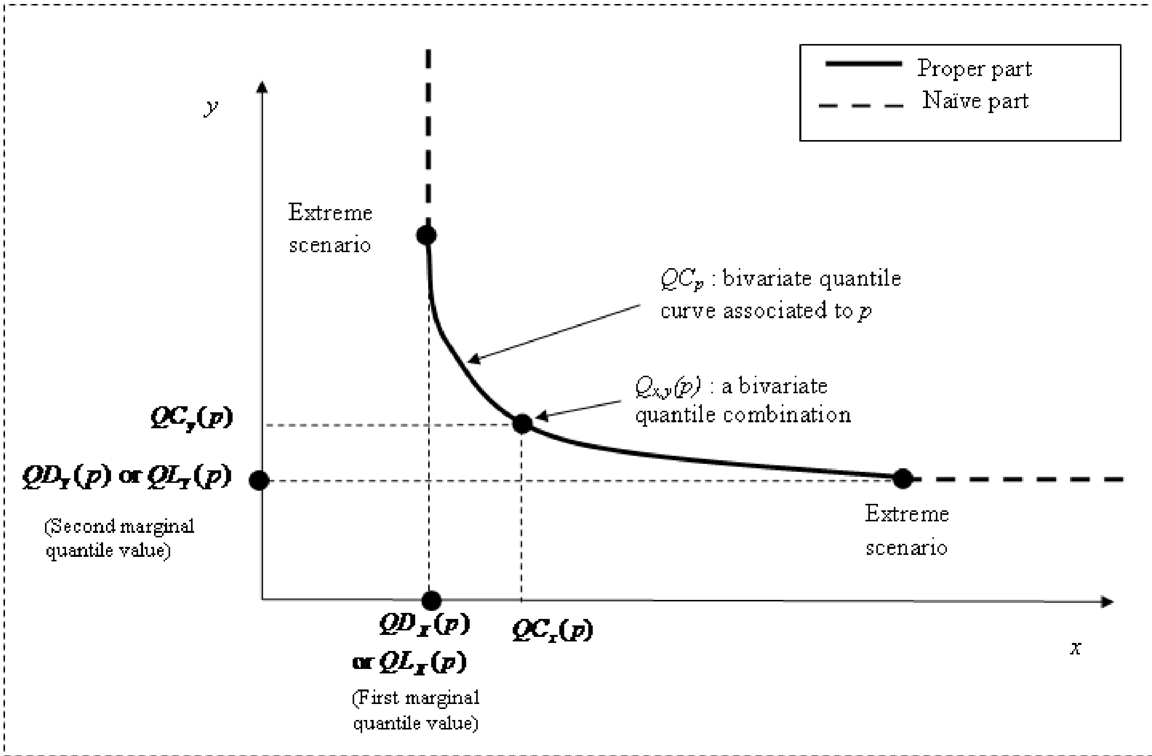


Figure 1: Illustration of the bivariate and univariate quantiles corresponding to the non-exceedence event

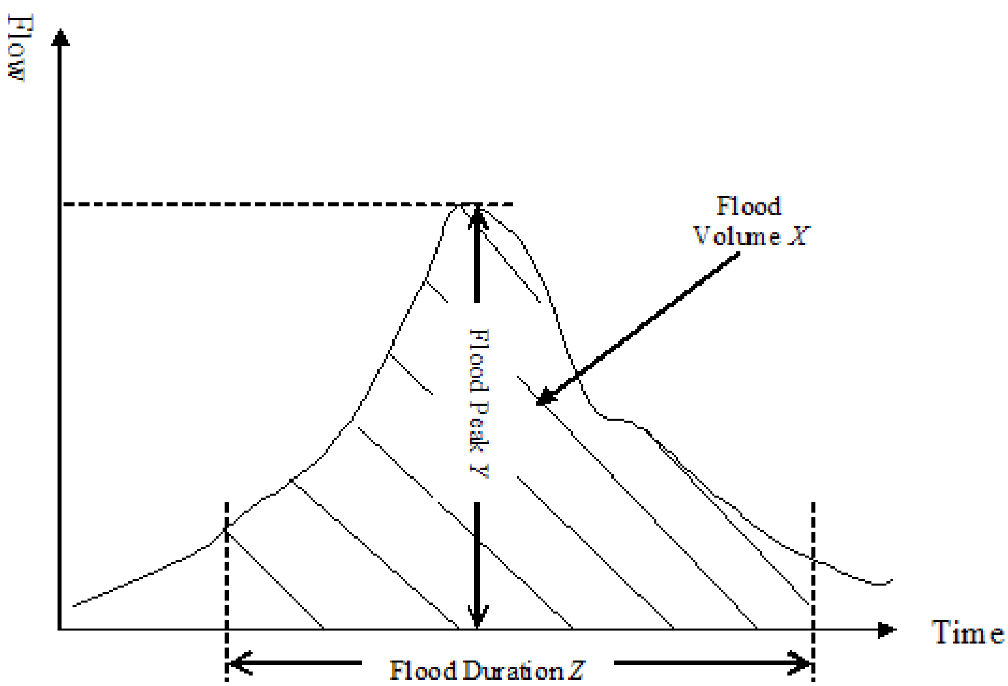


Figure 2: Typical flood hydrograph

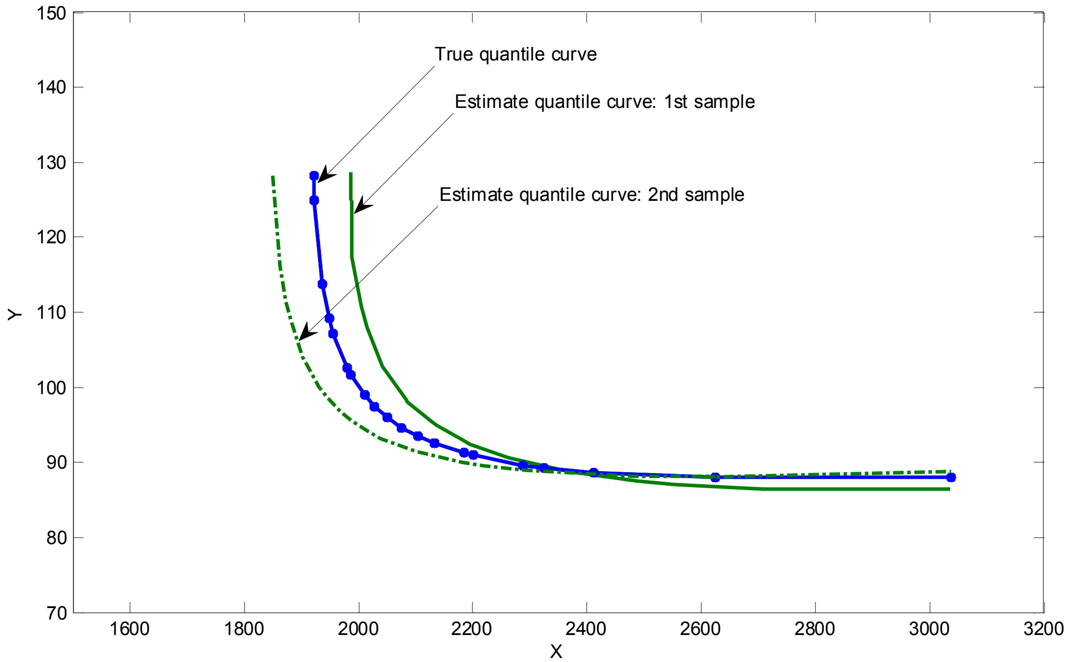
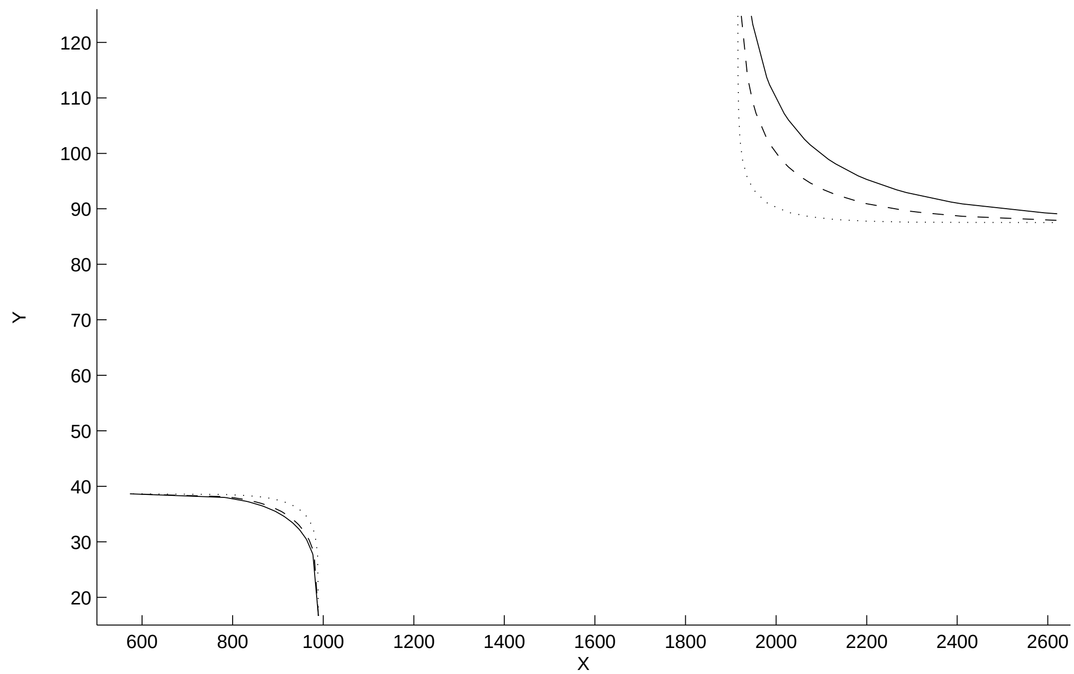
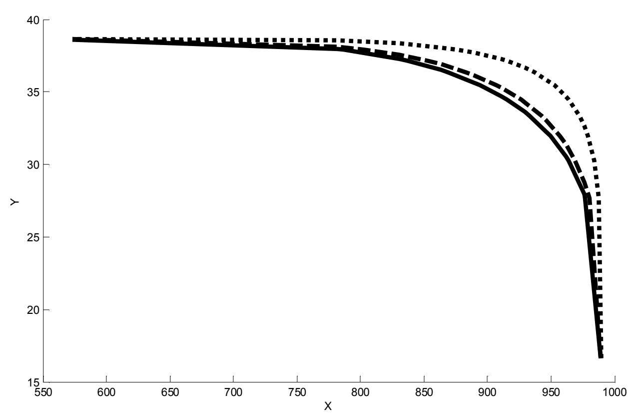
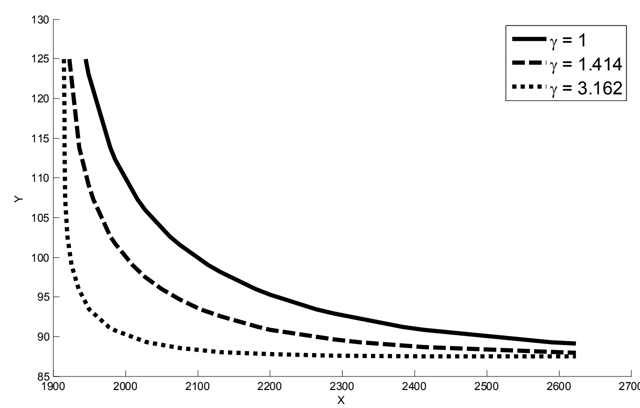


Figure 3: Illustration of estimated and true 0.9-quantile curves for two generated samples from the bivariate distribution used in the simulation with $\gamma = 1.414$ and n = 30 for the non-exceedence event



a)





b) c)

Figure 4: Illustration of the evolution of the quantile curves of the bivariate distribution used in the simulations with respect to γ for a) both exceedence and non-exceedence events b) exceedence event and c) non-exceedence event. The evaluated quantiles correspond to p =0.9. Figures b) and c) are zoomed in from Figure a). The axes x and y represent the flood volume and flood peak respectively.

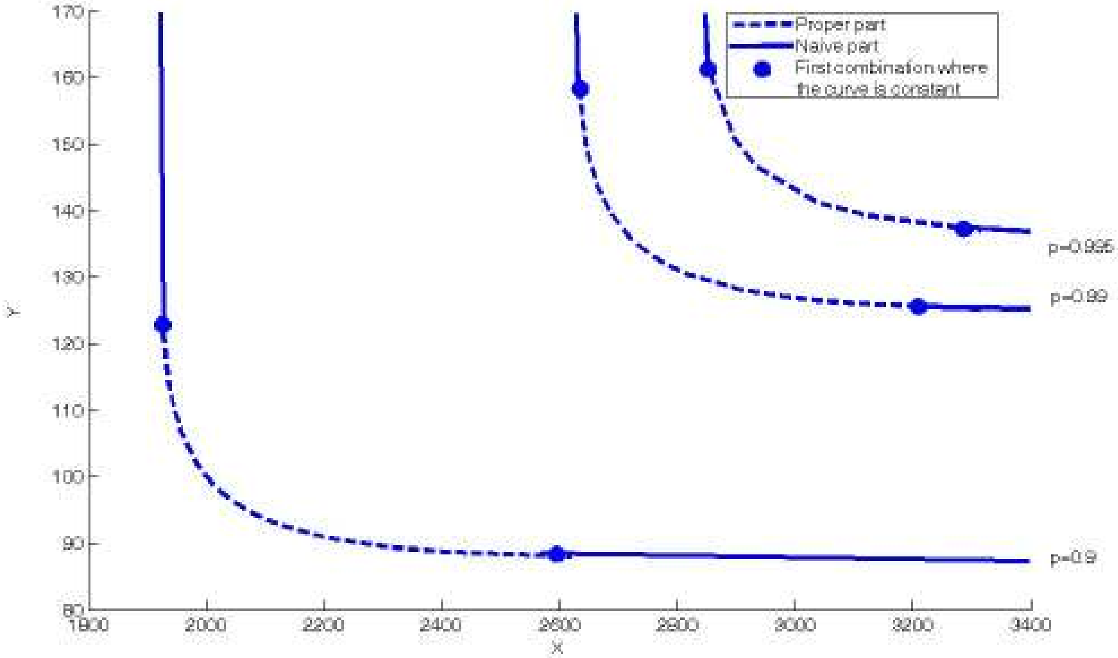


Figure 5: Illustration of quantile curves corresponding to three values of p from the bivariate distribution used in the simulation with $\gamma = 1.414$ and the non-exceedence event

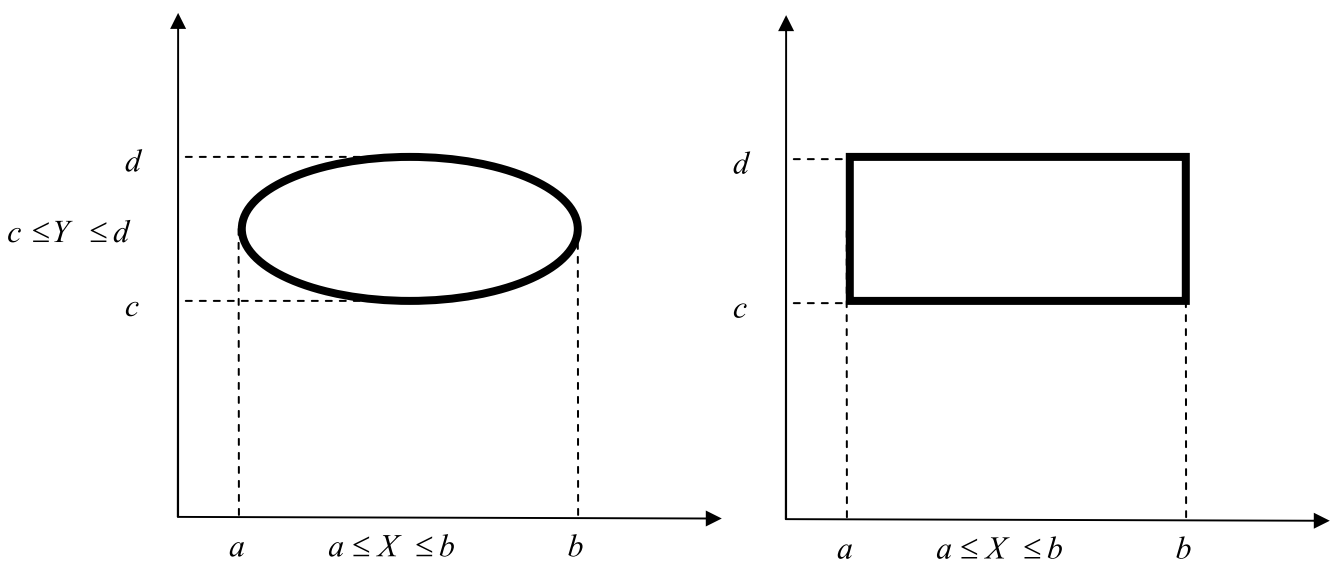


Figure 6: Illustration of the fact that the univariate modeling is limited and cannot provide a complete assessment of complex events

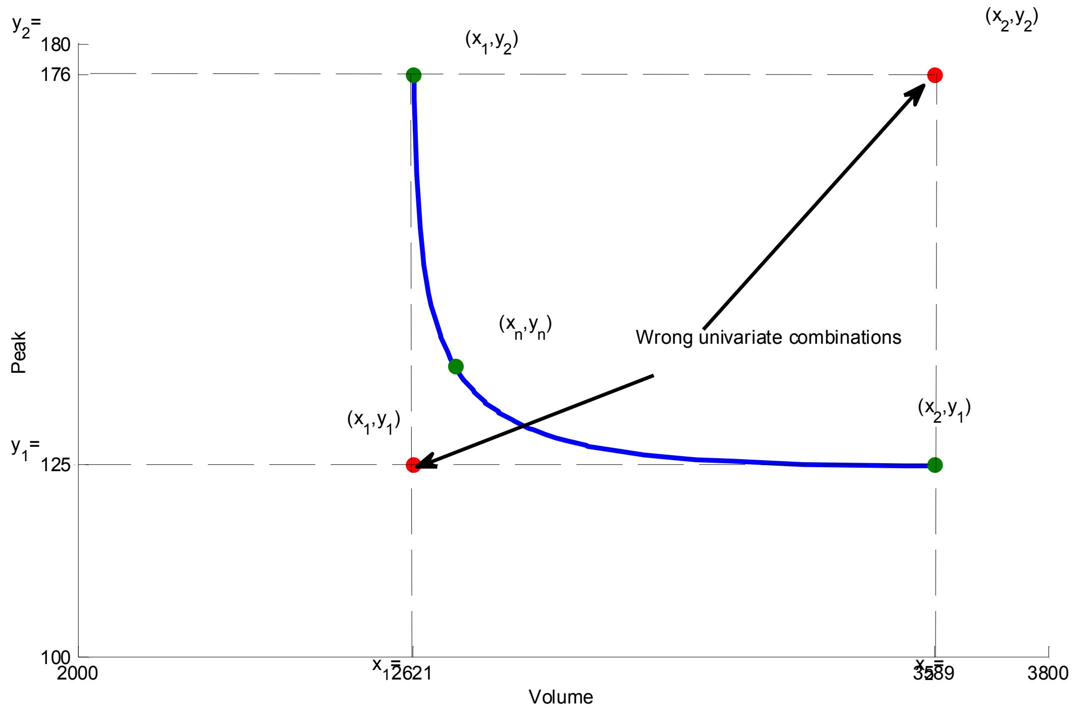
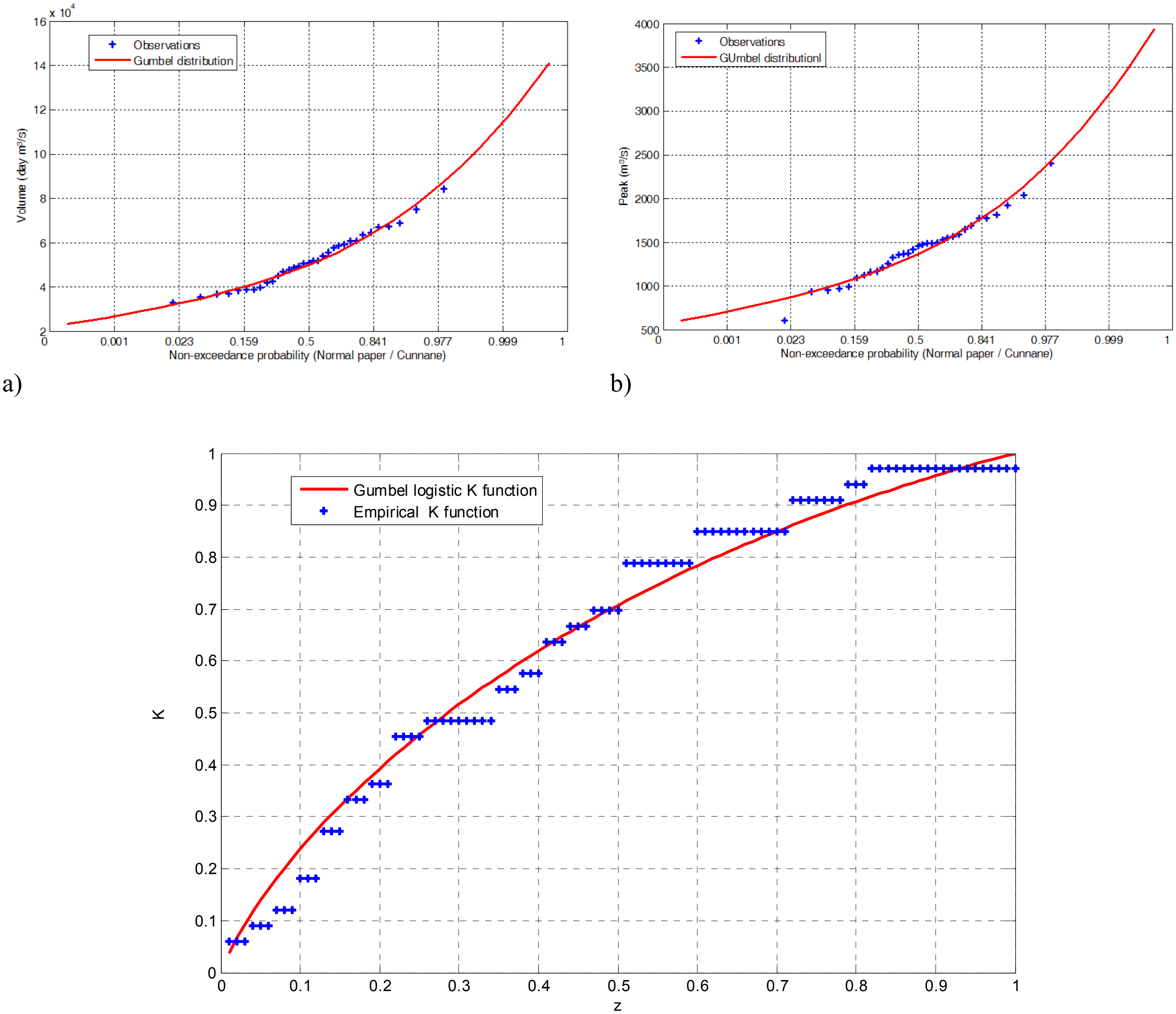


Figure 7: Illustration of univariate and bivariate 0.99-quantile combination values on the basis of the bivariate distribution used in the simulation with $\gamma = 1.414$ for the non-exceedence event



a) b)

c)

Figure 8: Bivariate distribution fitting to the case study data set a) Peak b) Volume and c) Copula

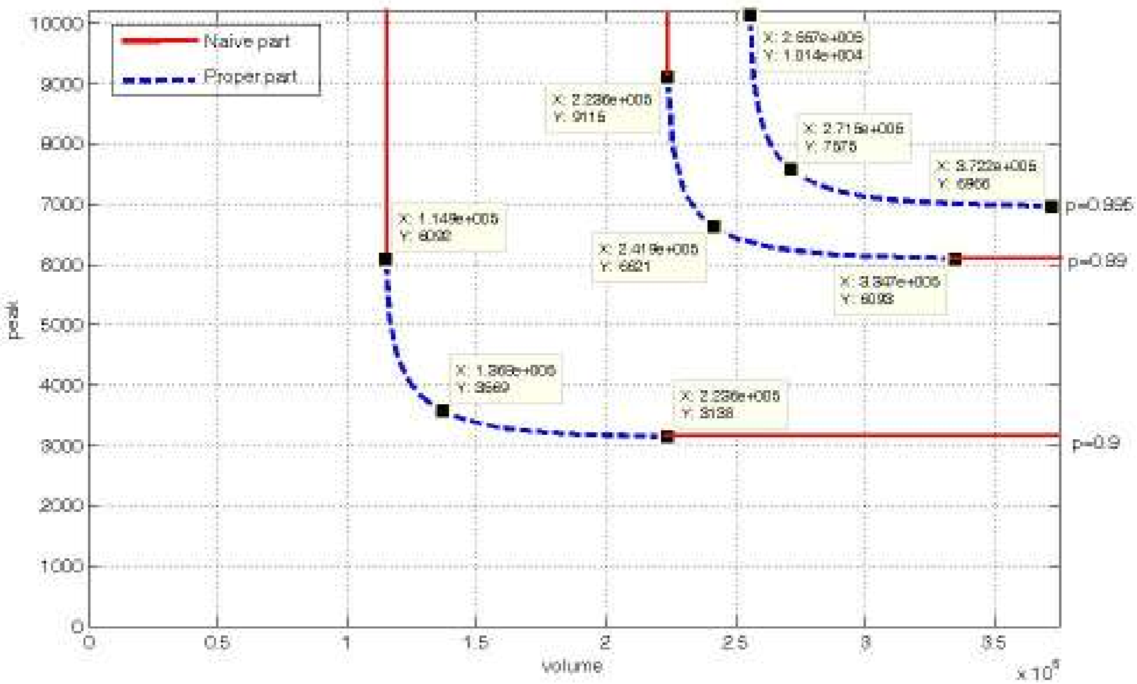


Figure 9: Quantile curves corresponding to the case study