Bivariate index-flood model for a northern case study

M.-A. Ben Aissia* [1]

F. Chebana [1]

T. B. M. J. Ouarda [1,2]

P. Bruneau [3]

M. Barbet [3]

[1] Statistics in Hydrology Working Group. Canada Research Chair on the Estimation of Hydrological Variables, INRS-ETE,

490, de la Couronne, Quebec, (Quebec) Canada, G1K 9A9.

[2] Institute Center for Water and Environment (iWATER)

Masdar Institute of Science and Technology,

PO Box 54224, Abu Dhabi, UAE.

[3] Hydro-Québec Équipement

855, Ste-Catherine East, 19th Floor, Montreal (Quebec) Canada, H2L 4P5.

* Corresponding author: mohamed.ben.aissia@ete.inrs.ca

October 2013

Abstract

Floods, as extreme hydrological phenomena, can be described by more than one correlated characteristic such as peak, volume and duration. These characteristics should be jointly considered since they are generally not independent. For an ungauged site, univariate regional flood frequency analysis (FA) provides a limited assessment of flood events. A recent study proposed a procedure for regional FA in a multivariate framework. This procedure represents a multivariate version of the index-flood model and is based on copulas and multivariate quantiles. The performance of the proposed procedure was evaluated by simulation. However, the model was not tested on a real-world case study data. In the present paper, practical aspects are investigated jointly for flood peak (Q) and volume (V) of a data set from the Côte-Nord region in the province of Quebec, Canada. The application of the proposed procedure requires the identification of the appropriate marginal distribution, the estimation of the index flood and the selection of an appropriate copula. The results of the case study show a good performance of the regional bivariate FA procedure. This performance depends strongly on the performance of the two univariate models and more specifically the univarite model of Q. Results show also the impact of the homogeneity of the region on the performance of the univariate and bivariate models.

1. Introduction and literature review

A flood can be described as a multivariate event whose main characteristics are peak, volume and duration. Thus, the severity of a flood depends on these characteristics, which are mutually correlated (Ashkar 1980, Yue et al. 1999, Ouarda et al. 2000, Yue 2001, Shiau 2003, De Michele et al. 2005, Zhang and Singh 2006, Chebana and Ouarda 2009, Chebana and Ouarda 2011). These studies show that these variables have to be jointly considered.

The use of joint probabilistic behaviour of correlated variables is necessary to understand the probabilistic characteristic of such events. Yue et al. (1999) used the bivariate Gumbel mixed model with standard Gumbel marginal distributions to represent the joint probability distribution of flood peak and volume, and flood volume and duration. Ouarda et al (2000) were first to study the joint regional behaviour of flood peaks and volume. To model flood peak and volume, Yue (2001) and Shiau (2003) used the Gumbel logistic model with standard Gumbel marginal distributions. Recently, copulas have been shown to represent a useful statistical tool to model the dependence between variables. To model flood peak and volume with Gumbel and Gamma marginal distribution respectively Zhang and Singh (2006) used the copula method, bivariate distributions of flood peak and volume, and flood volume and duration in frequency analysis (FA). Using the Gumbel–Hougaard copula, Zhang and Singh (2007) derived trivariate distributions of flood peak, volume and duration in FA.

Generally, the record length of the available streamflow data at sites is much shorter than the return period of interest and in some cases, there may not be any streamflow record at these sites. Consequently, local frequency estimation is difficult and/or not reliable. Regional FA is hence commonly used to overcome this lack of data. It is based on the transfer of available data from other stations within the same hydrologic region into a site where little or no data are available. The regional FA procedure was investigated with different approaches by several authors including Stedinger and Tasker (1986), Rocky Durrans and Tomic (1996), Nguyen and Pandey (1996), Hosking and Wallis (1997), Alila (1999, 2000) and Ouarda et al. (2001). GREHYS (1996a, 1996b) presented an intercomparison of various regional FA procedures.

In the literature, flood FA can be classified into four classes according to the univariate/multivariate and local/regional aspects. The local-univariate and regional-univariate classes were widely studied in the literature (Singh 1987, Wiltshire 1987, Burn 1990, Hosking and Wallis 1993, Hosking and Wallis 1997, Alila 1999, Ouarda et al. 2006, Nezhad et al. 2010). Recently, researchers have been increasingly interested in the multivariate case and many studies treated the problem of local-multivariate flood FA (Yue et al. 1999, Yue 2001, Shiau 2003, De Michele et al. 2005, Grimaldi and Serinaldi 2006, Zhang and Singh 2006, Chebana and Ouarda 2011). However, multivariate regional FA has received much less attention (Ouarda et al. 2000, Chebana and Ouarda 2007, Chebana and Ouarda 2009, Chebana et al. 2009).

The two main steps of the regional FA are the delineation of hydrological homogeneous regions and regional estimation (GREHYS 1996a). In the multivariate case, the delineation of hydrological homogeneous regions was treated by Chebana and Ouarda (2007). They proposed discordancy and homogeneity tests that are based on multivariate L-moments and copulas. Chebana et al. (2009) studied the practical aspects of these tests.

In univariate-regional FA, different methods were proposed to estimate extreme quantiles such as regressive models and index-flood models (e.g. GREHYS 1996a, 1996b). Chebana and Ouarda (2009) proposed a procedure for regional FA in a multivariate framework. The proposed procedure represents a multivariate version of the index-flood model. In this method, it is assumed that the distribution of flood characteristics (flood, peak or volume) at different sites within a given flood region is the same except for a scale parameter. Chebana and Ouarda (2009) adopted the multivariate quantile as the curve formed by the combination of variables corresponding to the same risk (Chebana and Ouarda 2011). In order to model the dependence between variables describing the event they employed the copula. In the present paper, practical aspects of the proposed procedure by Chebana and Ouarda (2009) are studied. Real data sets from sites in the Côte Nord region in the northern part of the province of Quebec, Canada are used. Flood peak and volume are the two variables studied jointly in the present study.

The next section presents the theoretical background, including the bivariate modelling, univariate index-flood model and multivariate quantiles. The “Multivariate Index-flood Model” section details the methodology of the adopted procedure with an emphasis on practical aspects. The case study section presents the study procedure as well as the obtained results. Concluding remarks are presented in the last section.

2. Background

In this section, the background elements to apply the index-flood model in the multivariate regional FA procedure are presented. Bivariate modelling including copulas and marginal distributions, univariate index-flood model and multivariate quantiles are briefly described.

II.2.1. Bivariate flood modelling and copulas

In bivariate modelling, a joint bivariate distribution for the underlying variables has to be obtained. According to Sklar's theorem (1959), the bivariate distribution is composed of a copula and two margins which are not necessarily similar.

In the remainder of the paper, we denote F_X and F_Y respectively the marginal distribution functions of given random variables X and Y, and F_{XY} the joint distribution function of the vector (X,Y).

a) Copula

Due to its ability to overcome the limitation of classical joint distributions, copulas have received increasing attention in various fields of science (see e.g. Nelsen 2006). Copulas are used to describe and model the dependence structure between the two random variables. A copula is an independent function of marginal distributions. For more details on copula functions, see for instance Nelsen (2006), Chebana and Ouarda (2007) and Salvadori et al. (2007). According to Sklar's (1959) theorem, we can construct the bivariate distribution F_{XY} with margins F_X and F_Y by:

$$F_{XY}(x,y) = C\left[F_X(x), F_Y(y)\right] \text{ for all real x and y} \quad (1)$$

When F_X and F_Y are continuous, the copula C is unique.

Different classes of copulas are studied in the literature such as the Archimedean, Elliptical, Extreme Value (EV), Plackette and Farlie-Gumbel-Morgenstern (FGM) copulas (see e.g. Nelsen 2006, Salvadori et al. 2007). The use of a copula requires the estimation of its parameters as well as goodness-of-fit procedures. In addition, since in hydrology we are particularly interested by the risk, the tail dependence of copulas is also a factor to take into account.

Copula parameter estimation: Assuming the unknown copula C belongs to a parametric family $C_0 = \{C_\theta : \theta \in R^q\}; q \geq 2$. The estimation of the parameter vector $\boldsymbol{\theta}$ is the first step to deal with. In the case of one-parameter bivariate copula, a popular approach consists of using the method of moment-type based on the inversion of Spearman's ρ and Kendall's τ. Demarta and McNeil (2005) have shown that such approach may lead to inconsistencies. The maximum pseudo-likelihood (MPL) approach is shown to be superior to the other ones (Besag 1975, Genest et al. 1995, Shih and Louis 1995, Kim et al. 2007) in which the observed data are transformed via the empirical marginal distributions to obtain pseudo-observations on which the maximum-likelihood approach is based to estimate the associated copula parameters (Genest et al. 1995). The advantage of this approach is that it can provide greater flexibility than the likelihood approach in the representation of real data. It consists in maximizing the log pseudo-likelihood:

$$\log L(\theta) = \sum_{i=1}^{n} \log c_\theta\left(\hat{U}_i\right) \qquad (2)$$

where c_θ denotes the density of a copula $C_\theta \in C_0$, and $\hat{U}_k = \left(\hat{U}_{kX}, \hat{U}_{kY}\right)^T$ are the pseudo-observation obtained from $\left(X_k, Y_k\right)^T$ given by:

$$\hat{U}_{kl} = R_{kl} \big/ (n+1), \quad k = 1,...,n; \quad l = X \text{ or } Y \qquad (3)$$

with R_{kX} being the rank of X_k among $X_1,\ldots,X_n$ and R_{lY} being the rank of Y_l among $Y_1,\ldots,Y_n$.

Goodness-of-fit test: The most important step in copula modelling is the copula selection by the goodness-of-fit test. Formally, one wants to test the hypotheses:

$$H_0 : C \in C_0 \quad \text{against} \quad H_1 : C \notin C_0 \tag{4}$$

Due to the novelty of copula modelling in flood FA, there is no common goodness-of-fit test for copulas. One of the most commonly used goodness-of-fit tests and valid only for Archimedean copulas is the graphic test proposed by Genest and Rivest (1993) based on the K function given by

$$K_\phi(u) = u - \frac{\phi(u)}{\phi'(u)} \quad 0 < u < 1 \tag{5}$$

where ϕ is the generator function of the Archimedean copula. The K function can be estimated by

$$\begin{aligned} \hat{K}(u) &= \frac{1}{N}\sum_{i=1}^{N} 1_{[w_i \le u]} \quad \text{where} \\ w_i &= \frac{1}{N-1}\sum 1_{[u_1^t < u_1^i, u_2^t < u_2^i]}, \ \mathrm{i} = 1,..,\mathrm{N} \end{aligned} \tag{6}$$

for a given bivariate sample $(u_1^1, u_2^1), (u_1^2, u_2^2), \ldots, (u_1^N, u_2^N)$. Genest and Revest (1993) have shown that $\hat{K}$ is a consistent estimator of K under weak regularity conditions. Note that Archimedean copulas are widely employed in hydrology and particularly to model flood dependence.

Recently, a relatively large number of goodness-of-fit tests were proposed (see e.g. Charpentier 2007, Genest et al. 2009, for extensive reviews). Genest et al. (2009) carried out a power study to evaluate the effectiveness of various goodness-of-fit tests and recommended a test based on a parametric bootstrapping procedure which makes use of the Cramer-von Mises statistic S_n (S_n goodness-of-fit test) :

$$S_n = \int n\left\{C_n(u,v) - C_{\theta_n}(u,v)\right\}^2 dC_n(u,v) \qquad (7)$$

where C_n is the empirical copula calculated using n observation data. and $C_{\theta n}$ is an estimation of C obtained assuming $C \in C_0$. The estimation $C_{\theta n}$ is based on the estimator θ_n of θ such as the maximum pseudo-likelihood estimator given in (2).

b) AIC for copula

In some cases, results of the goodness-of-fit testing show that more than one copula provide a good fit to the data set. To select the most adequate copula, we use the AIC (Akaike's information criterion) proposed by Kim et al. (2008) in the context of copulas:

$$\begin{aligned} AIC &= -2\log(L(\hat{\theta}; X, Y)) + 2r; \\ L(\hat{\theta}; X, Y) &= \sum \log\left\{ c\left(F_X(X), F_Y(Y), \hat{\theta}\right)\right\} \end{aligned} \qquad (8)$$

where $\hat{\theta}$ is the estimation of the copula parameter vector $\boldsymbol{\theta}$, r is the dimension of $\boldsymbol{\theta}$ and c is the copula density.

The copula which has the lowest AIC value is the most adequate copula for the data set.

c) Marginal distributions

To selection of the most appropriate marginal distribution (for X and for Y). The choice of the appropriate distribution is based on the Chi-square goodness-of-fit test, graphics and selection criteria (AIC see e.g. Akaike (1973) and BIC see e.g. Schwarz (1978)). For parameter estimation, a number of methods are available in the literature to estimate marginal distribution parameters; such as, the method of moments, the maximum likelihood method and the L-moments method.

II.2.2. Univariate Index-flood model

Introduced by Dalrymple (1960), the index-flood model was used initially for regional flood prediction. It is also used to model other hydrological variables including storms and droughts (e.g. Pilon 1990, Hosking and Wallis 1997, Hamza et al. 2001, Grimaldi and Serinaldi 2006). This model is based on the assumption of the homogeneity of the considered region and all the sites in the region have the same frequency distribution function apart from a scale parameter specific to each site. Let N_s be the number of sites in the region. The model gives the quantile $Q_i(p)$ corresponding to the non-exceedance probability p at site i as:

$$Q_i(p) = \mu_i q(p), \quad i = 1,....,N_s \quad \text{and} \quad 0 < p < 1 \qquad (9)$$

where μ_i corresponds to the index flood and q is the regional growth curve.

The index flood parameter μ_i can be estimated using a number of approaches (Hosking and Wallis 1997). For instance, Brath et al. (2001) used three models of estimating the index flood parameter. These models are multi-regression model, rational model and geomorphoclimatic model. They show that best results are given by considering the multi-regression model of the form:

$$\hat{\mu}_i = a_0 A_1^{a_1} A_2^{a_2} A_3^{a_3} ... A_{np}^{a_n} \qquad (10)$$

in which a_i are coefficients to be estimated, and A_i,…, A_{np} represent an appropriate set of morphological and climatic characteristics of the basin such as watershed area and slope of the main channel.

II.2.3. Multivariate quantiles

Unlike to the well-known univariate quantile, the multivariate quantile has received less attention in hydrology. Despite that, a few studies proposed multivariate quantile versions. For details, the reader is referred to Chebana and Ouarda (2011). The p^{th} bivariate quantile curve for the direction ε is defined as:

$$q_{XY}\left(p,\varepsilon\right)=\left\{\left(x,y\right)\in R^2 : F\left(x,y\right)=p\right\} \tag{11}$$

with $p \in I$ is the risk and *F(x,y)* is the bivariate cumulative distribution function given by:

$$F\left(x,y\right)=\Pr\left\{X \le x, Y \le y\right\} \tag{12}$$

which represents the probability of the simultaneous non-exceedance event. Other events can also be considered (see Chebana and Ouarda 2011 for more details).

The bivariate quantile in (10) is a curve corresponding to an infinity of combinations (*x*,*y*) that satisfies $F\left(x,y\right)=p$. For the event $\left\{X \le x, Y \le y\right\}$, using (2) and (10), the quantile curve can be expressed as follows:

$$q_{XY}\left(p\right)=\left\{\begin{matrix}\left(x,y\right)\in R^2 \quad \text{such that} \quad x=F_X^{-1}\left(u\right), \\ y=F_Y^{-1}\left(v\right);\ u,v\in\left[0,1\right]:C\left(u,v\right)=p\end{matrix}\right\} \tag{13}$$

The index-flood model used in this paper is based on (12). The resolution of (12), using copula and margin distribution, gives an infinity of combinations (x,y). These combinations constitute the corresponding quantile curve. The main properties of the index-flood model are (see Chebana and Ouarda 2011 for more details):

1. The marginal quantiles are special cases of the bivariate quantile curve. Indeed, they correspond to the extreme scenarios of the proper part related to the event;

2. The bivariate quantile curve is composed of two parts: naïve part and proper part. The proper part is the central part whereas the naïve part is composed of two segments starting at the end of each extremity of the proper part;
3. When the risk p increases, the proper part of the bivariate quantile becomes shorter.

3. Multivariate index-flood model in practice

The following procedure is proposed by Chebana and Ouarda (2009) and represents a complete multivariate version of regional FA. Since Chebana and Ouarda (2009) represent a theoretical study, we propose in the present paper a methodology of application of this procedure on a real world case study. The multivariate index-flood model regional estimation requires the delineation of a homogeneous region.

The step of delineation of a homogeneous region is treated by Chebana and Ouarda (2007) in the multivariate case. Based on multivariate *L*-moments, they proposed statistical tests of multivariate discordancy D and homogeneity H. The practical aspects of these tests are studied in Chebana et al. (2009).

The estimation procedure of the extreme event by the multivariate index-flood model is developed by Chebana and Ouarda (2009). It consists in extending the index-flood model to a multivariate framework using copula and multivariate quantiles. In this step, the homogeneity of the region is assumed. Indeed, non-homogeneous sites must be removed in the first step.

Let N' be the number of sites in the homogeneous region with record length n_i at site i, $i=1,\ldots,N'$. The goal is to estimate, at the target site l, the bivariate and marginal quantiles corresponding to a risk p.

Let $\left(x_{ij}, y_{ij}\right)$ for $i=1,\ldots,N'; j=1,\ldots,n_i$, be the data where x and y represent the observations of the considered variables. Let q_p be the regional growth curve which represents a quantile curve common to the whole region.

The complete procedure of determination of the bivariate quantile curve for an ungauged site is described as follows:

1. Identify the homogeneous region to be used in the estimation as follows: to identify and remove discordant sites, apply the multivariate discordancy test D and check the homogeneity of the remaining sites by the homogeneous test H. In practice, it's very difficult to find an exactly homogeneous region. According to Hosking and Wallis (1997), approximate homogeneity is sufficient to apply a regional FA, in the multivariate framework, this procedure was developed by Chebana and Ouarda (2007) and results will be used in this paper.
2. Assess the location parameters μ_{iX} and μ_{iY} $i=1,\ldots,N'$ and standardize the sample (x_{ij},y_{ij}), $j=1,\ldots,n_i$ to be:

$$x'_{ij} = \frac{x_{ij}}{\mu_{iX}}, y'_{ij} = \frac{y_{ij}}{\mu_{iY}} \qquad (14)$$

3. Select the bivariate distribution which is composed of a copula and two margins. In this step, our goal is to identify adequate marginal distributions and copula for the whole region to fit the standardized data $\left(x_{ij}', y_{ij}'\right)$. This step is described as follows:

a) Collect the data from the homogeneous region to get a sample $(x_k", y_k")$ $k=1,...,n;\ n=\sum_{i=1}^{N'} n_i$. This sample will be used to select the marginal distributions and copula.

b) Identify the adequate marginal distributions (for *X* and for *Y*) using the AIC, BIC and graphical criteria.

c) Select the adequate copula using the graphic test proposed by Genest and Rivest (1993) and the AIC criterion.

4. For each site *i*, *i=1,…,N'*, estimate the parameters of marginal distributions and copula family selected in step 3. For the copula family, the MPL method is used to estimate the copula parameter. However, for marginal distributions, the estimation method depends on the marginal distribution. Let $\hat{\theta}_k^{(i)}$ be the estimator of the k^{th} parameter from the standardized data of the *i*th site *k*=1,…, *s*; *s* is the number of parameters to be estimated, $i=1,...,N'$. Obtain the weighted regional parameter estimators:

$$\hat{\theta}_k^r = \frac{\sum_{i=1}^{N'} n_i \hat{\theta}_k^{(i)}}{\sum_{i=1}^{N'} n_i}, \quad k=1,...,s \tag{15}$$

5. For a given value of risk *p*, estimate different combinations of the estimated growth curve $\hat{q}_{x,y}(p)$ from (12) using the fitted copula with the corresponding weighted regional parameter $\hat{\theta}_k^{(R)}$with *k*=1,…, *s*.

6. Estimate the index flood parameter by a multivariate multiple regression model

$$\log(\mu) = E \times \log(A) + \varepsilon \tag{16}$$

where μ is the index flood vector, A is the matrix of watershed physiographic characteristics, E is the matrix of coefficients to estimate and ε is the error. The estimation of index flood can be separated into two steps:

a) Choice of physiographic characteristics: the aim of this step is to select, from a list of physiographic characteristics, the optimal set of physiographic characteristics to be considered in the model. Here, the order of characteristics in the selected set is important. The method of multivariate stepwise regression based on the Wilks statistics was used (see e.g. Rencher 2003).

b) Estimation of the coefficients E: the method of maximum likelihood is used (Meng and Rubin 1993).

7. Multiply each growth curve combination with the vector of index flood of the target l: μ_{lX} and μ_{lY}

$$\left(\hat{Q}^r_{xy}\left(p\right)\right)_l = \begin{pmatrix} \mu_{lX} \\ \mu_{lY} \end{pmatrix} \hat{q}_{xy}\left(p\right), \quad 0 < \mathrm{p} < 1 \tag{17}$$

Hence, the obtained result in (16) is an estimation of the bivariate regional quantile associated to the risk p.

To evaluate the performance of the regional FA models, Hosking and Wallis (1997) suggested an assessment procedure that involves generation of regional average L-moments through a Monte Carlo simulation. This procedure is based on the Jackknife resampling procedure (e.g. Chernick 2012). It consists in considering each site as an ungauged one by removing it temporarily from the region and estimating the bivariate and univariate regional quantiles for various nonexceedance probabilities p in the simulations. This is similar, for instance, to Ouarda et al (2001) in the regional frequency analysis context. At the *mth* repetition, the regional growth curves and the site *i* quantiles are computed.

As indicated in Chebana and Ouarda (2009), the performance of the corresponding bivariate regional FA model cannot be evaluated on the basis of the usual performance evaluation criteria. The evaluation is based on the deviation between the regional and local quantile estimated curves. The quantile curve is denoted by $(x, G_p(x))$. The relative error between the regional and local quantile curves is given by:

$$R_p(x) = \frac{G_p^r(x) - G_p^l(x)}{G_p^l(x)} \quad (18)$$

where exponents *r* and *l* referring respectively to regional and local quantile curves.

This relative difference represents vertical point-wise distances between the two quantile curves. In order to evaluate the estimation error for a site *I*, Chebana and Ouarda (2009) proposed the bias and root-mean-square error respectively given by

$$B_i(p) = \frac{100}{M}\sum_{m=1}^{M} REI_m{*}(p) \text{ and } R_m(p) = 100\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(REI_m(p)\right)^2} \quad (19)$$

where *M* is the number of simulations, REI* and REI are the two relative integrated error of the simulation *m* defined respectively by

$$REI{*}(p) = \frac{1}{L_p}\int_{QC_p} R_p(x)\,dx,\ 0 < p < 1 \quad (20)$$

$$REI(p) = \frac{1}{L_p}\int_{QC_p} \left|R_p(x)\right| dx,\ 0 < p < 1 \quad (21)$$

with L_p is the length of the proper part of the true quantile curve QC_p for the risk *p*.

To summarize these criteria over the sites of the region, it is possible to average them to obtain the regional bias, the absolute regional bias and the regional quadratic error given respectively by

$$
\begin{aligned}
RB(p) &= \frac{1}{N'}\sum_{i=1}^{N'} B_i \\
ARB(p) &= \frac{1}{N'}\sum_{i=1}^{N'} |B_i| \\
RRMSE(p) &= \frac{1}{N'}\sum_{i=1}^{N'} R_i
\end{aligned}
\qquad (22)
$$

4. Case study

The application of the index-flood model in a multivariate regional FA framework concerns a regional data set of interest for the Hydro-Québec Company. The two main flood characteristics, that is, volume V and peak Q are jointly considered. These flood features are random by definition since they are based on the flood starting and ending dates. The latter are obtained using an automatic method which consists in the analysis of cumulative annual hydrographs by adjusting the slopes with a linear approximation (e.g. Ben Aissia et al. 2012). The employed data is used in Chebana et al. (2009). They are from sites in the Côte Nord region in the northern part of the province of Quebec, Canada. The number of sites in the region is N=26 stations with record lengths n_i between 14 and 48 years. More information about the data is given in Table 1. Figure 1 presents the geographical location and the correlation coefficient between Q and V for the underlying sites.

II.4.1. Study procedure:

The procedure of the study is composed of the following eight steps:

1. Delineate the homogeneous region;
2. Assess the location parameters μ_{iV} and μ_{iQ} for i = 1,…, N' given by (13);

3. Select a family of regional multivariate distributions to fit the standardized data of the whole region;
4. For each site in the homogeneous region, estimate the parameters of the marginal distributions and copula family. Estimate the regional parameter estimator $\hat{\theta}_k^{(R)}$ by (14);
5. Estimate different combinations of the estimated growth curve $\hat{q}_{v,q}(p)$ from (12);
6. Estimate the index flood by a multiregression model (15);
7. Using (16), estimate the bivariate regional quantiles associated to the risk *p*;
8. For each flood characteristic, estimate the univariate regional growth curve and using (8) estimate the univariate regional quantile;
9. Evaluate the performance of the regional models (univariate and bivariate) by Monte Carlo simulation.

II.4.2. Result and discussion

In this section, results of the application of the adopted procedure are presented. First, results of the multivariate homogeneity study are briefly presented followed by the results of the index-flood regional estimation.

Discordancy and homogeneity

The employed data is the same used in Chebana et al. (2009) and the discordancy and homogeneity results are presented in that reference and in Table 1.The sites that may be discordant have a large discordancy value. Results show that:

- Sites 2 and 16 are discordant for *V*;

- Site 2 or sites 2 and 3 are discordant for *Q*;
- Sites 2 and 21 are discordant for (*V*,*Q*).

The two sites 2 and 21 are eliminated to allow application of the respective homogeneity test. Table 2 presents the homogeneity test values for the region for *V*, *Q* and (*V*,*Q*) after removing the two discordant sites (2 and 21). From Table 2, according to the statistic *H*, we conclude that the region is homogeneous for *V*, heterogeneous for *Q* and could be homogeneous for (*V*,*Q*).

Identification of marginal distributions

In regional FA, a single frequency distribution is fitted from the whole standardized data. In general, it will be difficult to get a homogeneous region, consequently there will be no single "true" marginal distribution that applies to each site (Hosking and Wallis 1997). Therefore, the aim is to find a marginal distribution that will yield accurate quantile estimates for each site. The scale factor of this marginal distribution changes from one site to another.

Figure 2 shows that the adequate marginal distributions are Gumbel for *Q* and GEV for *V*. Results for the appropriate marginal distributions are in agreement with those of similar studies e.g. Cunnane and Nash (1971) and De Michele and Salvadori (2002).

Identification of copula

Table 1 indicates that the dependence between *V* and *Q* varies from 0.34 to 0.82 while Figure 1 shows that the dependence variability is scattered over the entire study area. The graphic test based on the *K* function (5) with the estimate (6) is applied for the three

Archimedean copulas: Gumbel, Frank and Clayton. This test leads to fitting the Frank copula to the bivariate data for the studied region. The illustration of this fitting is presented in Figure 3.

The AIC and p-value of the S_n goodness-of-fit test described earlier and proposed by Kojadinovic and Yan (2009) are also calculated for the commonly considered copulas in hydrology. However, direct results show that none of the commonly used copulas in hydrology can be accepted. Even though, the graphic test based on the K function indicates excellent fitting with Frank copula, the S_n goodness-of-fit test rejects this copula, as well as the other ones being considered. First, the reason may be that numerical tests tend to be narrowly focused on a particular aspect of the relationship between the empirical copula and the theoretical copula and often try to compress that information into a single descriptive number or test result (see e.g. NIST 2013). Second, the test is widely and successfully applied to at-site hydrological studies which is not the case for regional studies where the total sample size is very large (here n=714). The performance of S_n goodness-of-fit test could be affected when the sample size is large as indicated in Genest et al. (2009). In addition, in terms of application, Vandenberghe et al. (2010) indicated limitation of this test for long sample size like in rainfall. Therefore, to overcome this situation, this test is applied to the data series of each site separately. This is justified since basically regional FA assumes the same distribution in each site apart from a scale factor (see e.g. Hosking and Wallis 1997, Ouarda et al. 2008). However, according to Hosking and Wallis (1997), it is difficult in practice to have a single distribution which provides a good fit for each site. The goal is hence to find a distribution that will yield accurate quantile estimates for all sites. For the present case-study, results (Table 3) show that Frank is the most accepted copula in the study sites (accepted by the S_n goodness-of-fit test for 20 sites and sorted best by AIC for 17 sites among 24 sites). Frank copula has already been shown to be adequate to model the dependence between flood *V* and *Q* in a number of hydrological studies (see e.g. Grimaldi and Serinaldi 2006). Finally, based on the above arguments (at-site Goodness-of-fit selection, regional graphic test based on the *K* function, regional and at-site AIC, hydrological literature), the Frank copula is selected for the present case-study. Therefore, the appropriate copula is *Frank* defined by:

$$C_\gamma(u,v) = \frac{1}{\ln\gamma}\ln\left[1+\frac{(\gamma^u-1)(\gamma^v-1)}{\gamma-1}\right]; \quad 0\le\gamma; \quad 0<u,v<1 \qquad (23)$$

where γ is the parameter to be estimated. The choice of the adequate copula is in agreement with those of similar studies e.g Lee et al. (2012).

Estimation of parameters associated to margins and copula

Parameters of marginal distributions and copula for each site and their corresponding confidence intervals are presented in Figure 4 while Table 4 showing the regional parameters of the marginal distributions and copula determined by (14). The MPL is employed for the copula parameter. For the Gumbel distribution, μ and σ represent, respectively, the location and scale parameters whereas for the GEV distribution, μ, σ and *k* represent respectively the location, scale and shape parameters. The ML method is used to estimate the Gumbel parameters while the generalized ML (Martins and Stedinger 2000) is used to estimate the GEV parameters.

Index flood estimation

To estimate the index flood $\hat{\mu}_Q$ of the peak and $\hat{\mu}_V$ of the volume, we use the multiregression m#odel described by (9). The available morphologic and climatic characteristics, used as explicative or input variables in the model are: watershed area in km^2 (*BV*), mean slope of the watershed in % (*BMBV*), percentage of forest in % (*PFOR*), percentage of area covered by lakes in % (*PLAC*), annual mean of total precipitation in mm (*PTMA*), summer mean of liquid precipitation in mm (*PLME*), degree days above zero in degree Celsius (*DJBZ*), absolute value of mean of minimum temperatures in January ($Tmin_{jan}$), February ($Tmin_{feb}$), March ($Tmin_{mar}$) and April ($Tmin_{apr}$), absolute value of mean of maximum temperatures in January ($Tmax_{jan}$), February ($Tmax_{feb}$), March ($Tmax_{mar}$) and April ($Tmax_{apr}$), and mean of cumulative precipitation in January ($PRCP_{jan}$), February ($PRCP_{feb}$), March ($PRCP_{mar}$) and April ($PRCP_{apr}$).

The selection of the significant variables to be included in model (9) is based on the stepwise method. Which led to the selection of *BV*, $Tmin_{jan}$, $Tmax_{fev}$ and $PRCP_{fev}$. The model coefficients are estimated by the ML method. Then, the model built is given by:

$$\begin{aligned} \hat{\mu}_Q &= -4.05 \cdot BV^{0.09} \cdot T\min_{jan}{}^{-1.33} \cdot T\max_{feb}{}^{1.04} \cdot PRCP_{feb}{}^{0.79} \\ \hat{\mu}_V &= 6.68 \cdot BV^{1.00} \cdot T\min_{jan}{}^{-3.31} \cdot T\max_{feb}{}^{1.55} \cdot PRCP_{feb}{}^{0.14} \end{aligned} \quad (1)$$

Note that *BV* is already selected in similar studies (e.g. Brath et al. 2001) which is not the case for $Tmin_{jan}$, $Tmax_{feb}$ and $PRCP_{feb}$.

Model performance is evaluated by the following criteria: coefficient of determination (R^{2*}), relative root-mean-square error ($RRMSE^*$) and mean relative bias (MRB^*) defined by:

$$R^{2*} = 1 - \frac{\sum_{i=1}^{N'} \left(\hat{\chi}_i - \chi_i\right)^2}{\sum_{i=1}^{N'} \left(\chi_i - \bar{\chi}\right)^2} \quad (2)$$

$$RRMSE^* = 100 \sqrt{\frac{1}{N'-1} \sum_{i=1}^{N'} \left(\frac{\hat{\chi}_i - \chi_i}{\chi_i}\right)^2} \quad (3)$$

$$MRB^* = 100 \frac{1}{N'} \sum_{i=1}^{N'} \left(\frac{\hat{\chi}_i - \chi_i}{\chi_i}\right) \quad (4)$$

with $\hat{\chi}_i$ and χ_i represent the estimated and calculated (mean of observed data in underling site) index flood respectively, and *N'* is the number of sites.

The criteria R^{2*}, $RRMSE^*$ and MRB^* are evaluated on the basis of a cross-validation of the model with Jackknife. Results are presented in Table 5. The obtained values of R^2 are higher than 0.95 which shows the high performance of the built model in (9). This performance is confirmed by the low values of R*RMSE* and *MRB* in Table 5.

Bivariate and univariate growth curve estimation

The bivariate regional growth curve is estimated for each risk value *p* by (12) and by using the regional parameters of the bivariate distribution. On the other hand, univariate regional growth curves of *V* and *Q* are estimated directly using regional parameters of marginal distributions. Figure 5 shows the univariate and bivariate estimated growth curves corresponding to nonexceedance probabilities *p* =0.9, 0.95, 0.99, 0.995 and 0.999 as well as the quantile curve in the unit square and the marginal distributions for *Q* and *V*. Univariate regional growth curves of *V* and *Q* are also presented in Table 6. Univariate and bivariate quantiles can be assessed by multiplying growth curves by the corresponding index flood (16).

Model performances

As described above, the accuracy of the quantile estimates of the three regional models: univariate of V (V-model), univariate of Q (Q-model) and bivariate of (V,Q) (VQ-model) is assessed using a Monte Carlo simulation procedure. The record lengths of the simulated sites are assumed to be the same as those of observed data and the number of simulations is set to be M=500. To illustrate these results, we present in Figure 6 the univariate and bivariate quantiles of three sites derived from one simulation (M=1) and from the sample data, as well as quantile curves in the unit square and the local and regional marginal distributions of Q and V. Figure 6 shows that, generally, the performance of the two univariate models and the bivariate model decrease with the risk level and depends on the discordancy values. Indeed, for Mistassibi (Figure 6 a) the performance of the V-model is higher than that of the Q-model which is in harmony with the two discordance values of V and Q and with the difference between marginal distributions (local and regional) of Q and V in the side panels. The performance of the bivariate model depends mainly on marginal distributions. Indeed, a small difference in the marginal distribution leads to possible wide shifts in the quantile curve. However, the unit square curves indicate very less effect. Figure 7 illustrates the bivariate quantiles (Regional and the 500 simulations) corresponding to a nonexceedance probability of p=0.9 for the Petit Saguenay station. Figure 7 shows that, in the Petit Saguenay station, the simulated bivariate quantile curves form a surface which includes (but not in the middle) the regional bivariate quantile curve. Table 7 presents the univariate and bivariate model performances of the corresponding nonexceedance probability p = 0.90, 0.95, 0.99, 0.995 and 0.999. The univariate and bivariate model performances in each site are presented in Figure 8.

Table 7 shows that the V-model performs well, since all performance criteria are less than 16% for all values of p. However, the performance of the Q-model is lower compared to that of the V-model where for instance, for p = 0.999, the RRMSE is larger than 21%. This conclusion can also be drawn from Figure 8 where the performance criteria of the Q-model are clearly higher than those of the V-model for all values of p. This conclusion can be explained by the fact that the region is heterogeneous for Q. On the other hand, the performance of the VQ-model is, generally, somewhat lower than the Q-model. This conclusion is confirmed by Figure 8 where we see a close performance criteria for the VQ-model and Q-model. One can explain this by the fact that the univariate quantiles are special cases of bivariate quantiles, since they correspond to the extreme scenario of the proper part related to the event. Then the performance of the univariate models has an effect on the performance of the bivariate model. Since the performance criteria of the Q-model are higher than those of the V-model then effects of the Q-model performance on the QV-model is more important than the effects of the V-model performance. On the other hand, from Figure 8 we observe that the performance behaviour criteria of the VQ-model and Q-model are similar to those of Gumbel parameters (Figure 4 a), especially for the scale parameter (σ). Consequently, a variation of the Gumbel parameters has an effect on the Q-model performance and therefore an effect on the VQ-model performance.

Performance criteria corresponding to the VQ-model are less than 19% for the highest considered risk level p = 0.999 (Table 7). Values of these performance criteria are larger than those obtained by Chebana and Ouarda (2009). Indeed, unlike their simulation study, the performance of the bivariate model is affected by the error of the index flood estimation as well as parameter estimations. Generally the performance criteria increase with the value of the risk p (Table 7 and Figure 8). An exception is recorded between p=0.995 and p=0.999 where performance criteria of the VQ-model are higher for p=0.995. This finding can be explained by the curse of dimensionality in the multivariate context, where the central part of a distribution contains little probability mass compared to the univariate framework (for more details see Scott 1992, Chebana and Ouarda 2009).

In order to further explain the results, we plot in Figure 9 the RRMSE of each model (for p=0.99) with respect to the corresponding discordancy values. Ideally we should find an increasing relation between the RRMSE of each model and the corresponding discordance. This relation is observed only for the V-model (Figure 9a) since the studied region is homogeneous for V, heterogeneous for Q and could be homogeneous for (V,Q). To find out other factors that have an impact on the model performance, we present in Figure 10 the RRMSE of the VQ-model (for p=0.99) with respect to watershed area and the correlation between V and Q. Figure 10a shows that high RRMSE values are seen for small watersheds whereas Figure 10b shows that sites with $\rho(V,Q) > 0.6$ have a good performance (RRMSE of the order of 10%) with the exception of Godbout (site number 15) which has ρ=0.75 and high RRMSE. Godbout is one of the four sites that have a high value of the Gumbel scale parameter and a high RRMSE of the Q-model and the VQ-model.

The quantile curve, for a given risk p, leads to infinite combinations of (Q,V) associated to the same return period. However, they could be not equal in practice or in practical point of view (Chebana and Ouarda 2011). Recently, Volpi and Fiori (2012) proposed a methodology to identify a subset of the quantile curve according to a fixed probability percentage of the events, on the basis of their probability of occurrence; see Volpi and Fiori (2012) for more details. As an illustrative example, the Chamouchouane station is considered. Figure 11 presents the curves and the limits with probability (1-α)=0.95.

5. Conclusions and perspectives

The procedure for regional FA in a multivariate framework is applied to a set of sites from the Côte-Nord region in the northern part of the province of Quebec, Canada. This procedure is proposed by Chebana and Ouarda (2009) and represents a multivariate version of the index-flood model. It is based on copulas and multivariate quantiles. Chebana and Ouarda (2009) evaluated the proposed model based on a simulation study. In the present paper, practical aspects of this model are presented and investigated jointly for the flood peak and volume of the considered data set.

Results show that the appropriate fitted marginal distributions are Gumbel for Q and GEV for V as well as the Frank copula for their dependence structure. The multi-regressive proposed method to estimate the index flood is shown to lead to a high performance. The performance of the two univariate models is in accordance with the quality of the region (homogeneity test). Indeed, the studied region is homogenous for V and heterogeneous for Q where the performance of the V-model is higher than that of the Q-model. The high performance of the V-model is confirmed by a relation between their performance criteria and the discordance values of V in each site whereas the low performance of the Q-model is mainly caused by the variation of the marginal distribution parameters. This is a logical consequence of the heterogeneity of the region for Q. The performance of the two univariate models increases with the risk level *p*. For the bivariate model, the performance criteria are less than 19% which indicates the high performance of the proposed procedure to estimate bivariate quantiles at ungauged sites. This performance increases, generally, with the risk level *p* and is affected by the performance of the Q-model. Results show also that high values of the performance criteria of the bivariate regional model are seen for small watershed and for sites with low correlation between *V* and *Q*. From this study it is concluded that a good performance of the bivariate model requires good performance of the two univariate models. This means that we should have a homogeneous region for both univariate variables.

The considered method estimates the bivariate quantile as combinations that constitute the quantile curve for a given risk level *p*. A method to select the appropriate combination(s) for a specific application is of interest and should be developed in future efforts. Furthermore, the adaptation of the model to the estimation of other hydrological phenomena such as drought and the consideration of others homogenous regions can be conducted by considering the appropriate distributions, copulas and events to be studied.

ACKNOWLEDGEMENTS

The authors thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and Hydro-Québec for the financial support.

REFERENCES

Akaike, H., 1973. "Information measures and model selection." *Information measures and model selection*, 50: 277-290.

Alila, Y., 1999. "A hierarchical approach for the regionalization of precipitation annual maxima in Canada." *Journal of Geophysical Research D: Atmospheres*, 104(D24): 31645-31655.

Alila, Y., 2000. "Regional rainfall depth-duration-frequency equations for Canada." *Water Resources Research*, 36(7): 1767-1778.

Ashkar, F., 1980. Partial duration series models for flood analysis. Montreal, Qc, Canda, Ecole Polytech of Montreal.

Ben Aissia, M. A., F. Chebana, T. B. M. J. Ouarda, L. Roy, G. Desrochers, I. Chartier and É. Robichaud, 2012. "Multivariate analysis of flood characteristics in a climate change context of the watershed of the Baskatong reservoir, Province of Québec, Canada." *Hydrological Processes*, 26(1): 130-142.

Besag, J., 1975. "Statistical Analysis of Non-Lattice Data." *Journal of the Royal Statistical Society. Series D (The Statistician)*, 24(3): 179-195.

Brath, A., A. Castellarin, M. Franchini and G. Galeati, 2001. "Estimating the index flood using indirect methods." *Estimation de l'indice de crue par des méthods indirectes*, 46(3): 399-418.

Burn, D. H., 1990. "Evaluation of regional flood frequency analysis with a region of influence approach." *Water Resources Research*, 26(10): 2257-2265.

Charpentier, A., Fermanian, J.-D., Scaillet, O., 2007. The estimation of copulas: Theory and practice. New York, *Risk Books*.

Chebana, F. and T. B. M. J. Ouarda, 2007. "Multivariate L-moment homogeneity test." *Water Resources Research*, 43.

Chebana, F. and T. B. M. J. Ouarda, 2009. "Index flood–based multivariate regional frequency analysis." *Water Resources Research*, 45.

Chebana, F. and T. B. M. J. Ouarda, 2011. "Multivariate quantiles in hydrological frequency analysis." *Environmetrics*, 22(1): 63-78.

Chebana, F., T. B. M. J. Ouarda, P. Bruneau, M. Barbet, S. E. Adlouni and M. Latraverse, 2009. "Multivariate homogeneity testing in a northern case study in the province of Quebec, Canada." *Hydrological Processes*, 23(12): 1690-1700.

Chernick, M. R., 2012. "The jackknife: A resampling method with connections to the bootstrap." *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2): 224-226.

Cunnane, C. and J. E. Nash, 1971. "Bayesian estimation of frequency of hydrological events." *Internationa Association of Hydrological Sciences Publications*, 100: 47-55.

Dalrymple, T., 1960. Flood-frequency analyses. Washington, D.C., *U.S. G.P.O.*

De Michele, C. and G. Salvadori, 2002. "On the derived flood frequency distribution: analytical formulation and the influence of antecedent soil moisture condition." *Journal of Hydrology*, 262(1-4): 245-258.

De Michele, C., G. Salvadori, M. Canossi, A. Petaccia and R. Rosso, 2005. "Bivariate statistical approach to check adequacy of dam spillway." *Journal of Hydrologic Engineering*, 10(1): 50-57.

Demarta, S. and A. J. McNeil, 2005. "The t copula and related copulas." *International Statistical Review*, 73(1): 111-129.

Durrans, S. R. and S. Tomic, 1996. "Regionalization of low-flow frequency estimates: an Alabama case study." *Journal of the American Water Resources Association*, 32(1): 23-37.

Genest, C., K. Ghoudi and L. Rivest, 1995. "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions." *Biometrika*, 82(3): 543-552.

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

Genest, C. and L.-P. Rivest, 1993. "Statistical Inference Procedures for Bivariate Archimedean Copulas." *Journal of the American Statistical Association*, 88(423): 1034-1043.

GREHYS, 1996a. "Presentation and review of some methods for regional flood frequency analysis." *Journal of Hydrology*, 186(1-4): 63-84.

GREHYS, 1996b. "Inter-comparison of regional flood frequency procedures for Canadian rivers." *Journal of Hydrology*, 186(1-4): 85-103.

Grimaldi, S. and F. Serinaldi, 2006. "Asymmetric copula in multivariate flood frequency analysis." *Advances in Water Resources*, 29(8): 1155-1167.

Hamza, A., T. B. M. J. Ouarda, S. R. Durrans and B. Bobée, 2001. "Development of scale invariance and tail models for the regional estimation of low-flows." *Canadian Journal of Civil Engineering*, 28(2): 291-304.

Hosking, J. R. M. and J. R. Wallis, 1993. "Some statistics useful in regional frequency analysis." *Water Resources Research*, 29(2): 271-281.

Hosking, J. R. M. and J. R. Wallis, 1997. Regional frequency analysis : an approach based on L-moments. Cambridge, *Cambridge University Press*.

Kim, G., M. J. Silvapulle and P. Silvapulle, 2007. "Comparison of semiparametric and parametric methods for estimating copulas." *Computational Statistics and Data Analysis*, 51(6): 2836-2850.

Kim, J.-M., Y.-S. Jung, E. Sungur, K.-H. Han, C. Park and I. Sohn, 2008. "A copula method for modeling directional dependence of genes." *BMC Bioinformatics*, 9(1): 225.

Kojadinovic, I. and J. Yan, 2009. "A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems." *Statistics and Computing*: 1-14.

Lee, T., R. Modarres and T. B. M. J. Ouarda, 2012. "Data based analysis of bivariate copula tail dependence for drought duration and severity." *Hydrological Processes*: in press.

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

Meng, X. L. and D. B. Rubin, 1993. "Maximum likelihood estimation via the ecm algorithm: A general framework." *Biometrika*, 80(2): 267-278.

Nelsen, R. B., 2006. An introduction to copulas. New York, *Springer*.

Nezhad, M. K., K. Chokmani, T. B. M. J. Ouarda, M. Barbet and P. Bruneau, 2010. "Regional flood frequency analysis using residual kriging in physiographical space." *Hydrological Processes*, 24(15): 2045-2055.

Nguyen, V.-T.-V. and G. Pandey, 1996. A new approach to regional estimation of floods in Quebec. Proceedings of the 49th Annual Conference of the CWRA, Quebec City, *Collection Environnement de l'Université de Montréal*.

NIST, 2013. e-Handbook of Statitical Methods. http://www.itl.nist.gov/div898/handbook/pmd/section4/pmd44.htm, [accessed september 2013].

Ouarda, T. B. M. J., K. M. Bâ, C. Diaz-Delgado, A. Cârsteanu, K. Chokmani, H. Gingras, E. Quentin, E. Trujillo and B. Bobée, 2008. "Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study." *Journal of Hydrology*, 348(1–2): 40-58.

Ouarda, T. B. M. J., J. M. Cunderlik, A. St-Hilaire, M. Barbet, P. Bruneau and B. Bobée, 2006. "Data-based comparison of seasonality-based regional flood frequency methods." *Journal of Hydrology*, 330(1-2): 329-339.

Ouarda, T. B. M. J., C. Girard, G. S. Cavadias and B. Bobée, 2001. "Regional flood frequency estimation with canonical correlation analysis." *Journal of Hydrology*, 254(1-4): 157-173.

Ouarda, T. B. M. J., M. Hache, P. Bruneau and B. Bobee, 2000. "Regional flood peak and volume estimation in northern Canadian basin." *Journal of Cold Regions Engineering*, 14(4): 176-191.

Pilon, P. J., 1990. The Weibull distrubution applied to regional low-flow frequency analysis. Regionalization in Hydrology, Ljubljana, *IAHS Publ.*

Rencher, A. C., 2003. Methods of Multivariate Analysis, *John Wiley & Sons, Inc.*

Salvadori, G., N. Carlo De Michele, T. Kottegoda and R. Rosso, 2007. Extremes in Nature: An Approach Using Copula. Dordrecht, *Springer, p292*.

Schwarz, G., 1978. "Estimating the Dimension of a Model." *Annals of Statistics*, 6(2): 461-464.

Scott, D. W., 1992. Multivariate density estimation : theory, practice, and visualization. New York, *Wiley, p265*.

Shiau, J. T., 2003. "Return period of bivariate distributed extreme hydrological events." *Stochastic Environmental Research and Risk Assessment*, 17(1-2): 42-57.

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

Singh, V. P., 1987. Regional flood frequency analysis. Dordrecht, *D. Reidel Publishing, p419*.

Sklar, A., 1959. "Fonction de répartition à n dimensions et leurs margins." *Institut Statistique de l'Université de Paris*, 8: 229-231.

Stedinger, J. R. and G. D. Tasker, 1986. "Regional Hydrologic Analysis, 2, Model-Error Estimators, Estimation of Sigma and Log-Pearson Type 3 Distributions." *Water Resources Research*, 22(10): 1487-1499.

Vandenberghe, S., N. E. C. Verhoest and B. De Baets, 2010. "Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall." *Water Resources Research*, 46(1): W01512.

Volpi, E. and A. Fiori, 2012. "Design event selectionin bivariate hydrological frequency analysis." *Hydrological science journal*, 57(8): 1-10.

Wiltshire, S. E. (1987). "Statistical techniques for regional flood-frequency analysis. [electronic resource]."

Yue, S., 2001. "A bivariate extreme value distribution applied to flood frequency analysis." *Nordic Hydrology*, 32(1): 49-64.

Yue, S., T. B. M. J. Ouarda, B. Bobée, P. Legendre and P. Bruneau, 1999. "The Gumbel mixed model for flood frequency analysis." *Journal of Hydrology*, 226(1-2): 88-100.

Zhang, L. and V. P. Singh, 2006. "Bivariate flood frequency analysis using the copula method." *Journal of Hydrologic Engineering*, 11(2): 150-164.

Zhang, L. and V. P. Singh, 2007. "Trivariate Flood Frequency Analysis Using the Gumbel-Hougaard Copula." *Journal of Hydrologic Engineering*, 12(4): 431-439.

Tables and Figures

Tables

Table 1: Discordancy statistic for each site (Chebana et al. 2009).

#	Site name	*BV* (Km^2)	n_i	(*V*,*Q*) correlation coefficient	Discordancy statistic		
					V	Q	(V,Q)
1	Petit Saguenay	729	24	0.50	0.80	0.40	1.09
2	Des Ha Ha	564	19	0.73	3.60	4.44	3.88
3	Aux Écorces	1120	34	0.5	0.16	2.42	0.69
4	Pikauba	489	34	0.34	0.89	1.16	1.22
5	Métabetchouane	2270	30	0.54	1.22	1.23	1.59
6	Petite Péribonka	1090	31	0.62	0.26	0.45	0.98
7	Chamouchouane (Ashuapmushuan)	15 300	43	0.70	0.13	0.14	0.26
8	Mistassibi	8690	39	0.52	0.32	0.78	0.88
9	Mistassini	9620	43	0.52	0.62	0.19	0.53
10	Manouane	3720	23	0.39	0.55	0.47	2.38
11	Valin	740	31	0.42	0.40	0.46	2.37
12	Ste-Marguerite	1100	21	0.48	1.50	0.55	1.30
13	DesEscoumins	779	19	0.49	1.14	1.81	1.27
14	Portneuf	2580	20	0.80	0.99	0.32	1.06
15	Godbout	1570	30	0.75	0.91	0.89	1.29
16	Aux-Pékans	3390	16	0.54	3.19	0.38	2.25
17	Tonerre	674	48	0.64	0.51	1.65	2.25
18	Magpie	7200	27	0.66	0.12	1.23	1.11
19	Romaine	13 000	48	0.68	1.62	0.48	0.57
20	Nabisipi	2060	25	0.78	1.12	0.64	0.54
21	Aguanus	5590	19	0.60	1.53	0.84	3.07
22	Natashquan	15 600	39	0.75	0.28	0.39	1.02
23	Etamamiou	2950	19	0.82	1.06	1.33	1.32
24	St Augustin	5750	14	0.73	0.62	0.67	0.92
25	St Paul	6630	25	0.73	0.31	1.35	1.11
26	Moisie	19000	39	0.65	1.16	0.32	0.54

Table 2 : Homogeneity after exclusion of the discordant sites

	V	***Q***	**(*V*,*Q*)**
H	0.7052	2.4081	1.5234

Table 3 : Results of *Sn* Goodness-of-fit test and AIC criterion for considered copulas. Gray color indicates that Frank copula is accepted by *Sn* goodness-of-fit test (p-value column) and has the smallest AIC (AIC column) for the corresponding site.

Site	Gumbel		Frank		Clayton		Galambos		Husler-Reiss		Placket	
	P-value	AIC	P-value	AIC	P-value	AIC	P-value	AIC	P-value	AIC	P-value	AIC
1	0.190	-75.0	0.078	-97.5	0.012	-55.2	0.237	-74.5	0.318	-74.1	0.034	-66
3	0.086	-106.2	0.081	-151.7	0.175	-72.2	0.095	-105.4	0.117	-104.8	0.086	-86
4	0.173	-98.6	0.154	-200.9	0.460	-70.8	0.183	-97.4	0.194	-96.6	0.191	-84
5	0.150	-114.0	0.137	-177.7	0.083	-82.3	0.169	-113.2	0.187	-112.4	0.068	-102
6	0.128	-146.8	0.133	-103.9	0.064	-102.3	0.030	-146.5	0.028	-145.9	0.044	-131
7	0.152	-210.9	0.054	-202.4	0.003	-140.0	0.159	-210.2	0.202	-209.1	0.020	-182
8	0.135	-142.9	0.207	-181.7	0.120	-95.0	0.148	-142.1	0.175	-141.3	0.135	-116
9	0.148	-175.2	0.404	-231.1	0.041	-115.7	0.155	-174.2	0.197	-173.2	0.214	-145
10	0.459	-48.0	0.231	-96.7	0.104	-36.6	0.480	-47.4	0.522	-47.1	0.218	-42
11	0.002	-101.2	0.017	-172.8	0.242	-71.4	0.002	-100.3	0.002	-99.5	0.017	-86
12	0.120	-66.7	0.113	-59.9	0.200	-48.7	0.114	-66.4	0.112	-66.2	0.180	-58
13	0.016	-71.4	0.037	-101.0	0.079	-56.2	0.016	-70.9	0.016	-70.4	0.036	-69
14	0.027	-101.0	0.058	24.1	0.019	-78.1	0.026	-101.1	0.029	-101.0	0.020	-98
15	0.120	-160.8	0.041	-164.2	0.048	-118.1	0.118	-160.3	0.130	-159.2	0.066	-153
16	0.243	-43.0	0.124	-56.0	0.227	-33.1	0.253	-42.7	0.249	-42.5	0.199	-39
17	0.048	-164.2	0.003	-230.3	0.000	-113.1	0.059	-163.2	0.069	-162.1	0.003	-143
18	0.092	-123.2	0.112	-105.3	0.255	-88.9	0.081	-122.7	0.069	-122.1	0.135	-113
19	0.214	-236.0	0.177	-180.7	0.150	-149.6	0.208	-235.3	0.222	-234.4	0.192	-193
20	0.352	-122.2	0.326	28.1	0.059	-90.4	0.366	-122.1	0.345	-121.9	0.321	-116
22	0.241	-193.4	0.215	-199.3	0.006	-131.4	0.254	-192.7	0.302	-191.6	0.022	-171
23	0.002	-100.3	0.011	-184.9	0.040	-85.4	0.002	-100.5	0.007	-101.2	0.002	-104
24	0.173	-64.5	0.091	29.3	0.002	-55.0	0.179	-64.5	0.206	-64.6	0.052	-64
25	0.138	-107.0	0.310	-111.4	0.031	-78.5	0.140	-106.6	0.140	-106.0	0.035	-99
26	0.168	-182.9	0.280	-192.1	0.041	-123.6	0.158	-182.2	0.193	-181.2	0.127	-159
Pooled data	0.0005	-329.35	0.045	-365.65	0.0005	-318.94	0.0005	-327.34	0.0005	-300.11	0.04785	-364.8

Table 4: Regional parameters of marginal distributions and copula

Marginal distribution					Copule
Peak (Gumbel)		Volume (GEV)			Frank
μ_r	σ_r	k_r	σ_r	μ_r	γ_r
1.16	0.33	0.16	0.28	0.88	2.06

Table 5 : Performance criteria of multiregression index flood model

	R^{2*}	$MRB^*(\%)$	$RRMSE^*$ (%)
Q	0.94	1.24	16.75
V	0.97	0.70	11.68

Table 6 : Univariate regional growth curve values

Marginal distribution		*p*				
		0.9	**0.95**	**0.99**	**0.995**	**0.999**
Volume	**(GEV)**	1.40	1.53	1.77	1.85	2.02
Peak	**(Gumbel)**	1.90	2.13	2.67	2.90	3.43

Table 7 : Performance of the univariate and bivariate quantiles corresponding to the nonexceedance probabilities 0.9, 0.95, 0.99, 0.995 and 0.999.

Risk	Criterion	($V.Q$)	V	Q
p=0.9	RB	1.99	-1.25	-0.94
	ARB	9.60	3.44	11.87
	RRMSE	15.25	7.68	13.88
p=0.95	RB	3.27	-1.34	-1.05
	ARB	11.25	4.32	13.56
	RRMSE	17.34	9.37	15.83
p=0.99	RB	2.49	-0.78	-1.26
	ARB	12.03	5.95	15.99
	RRMSE	17.93	13.56	18.79
p=0.995	RB	3.41	-0.20	-1.73
	ARB	12.87	7.25	16.93
	RRMSE	19.06	14.90	19.77
p=0.999	RB	3.23	0.56	-1.51
	ARB	12.21	7.89	18.95
	RRMSE	18.21	15.79	21,78

Figures

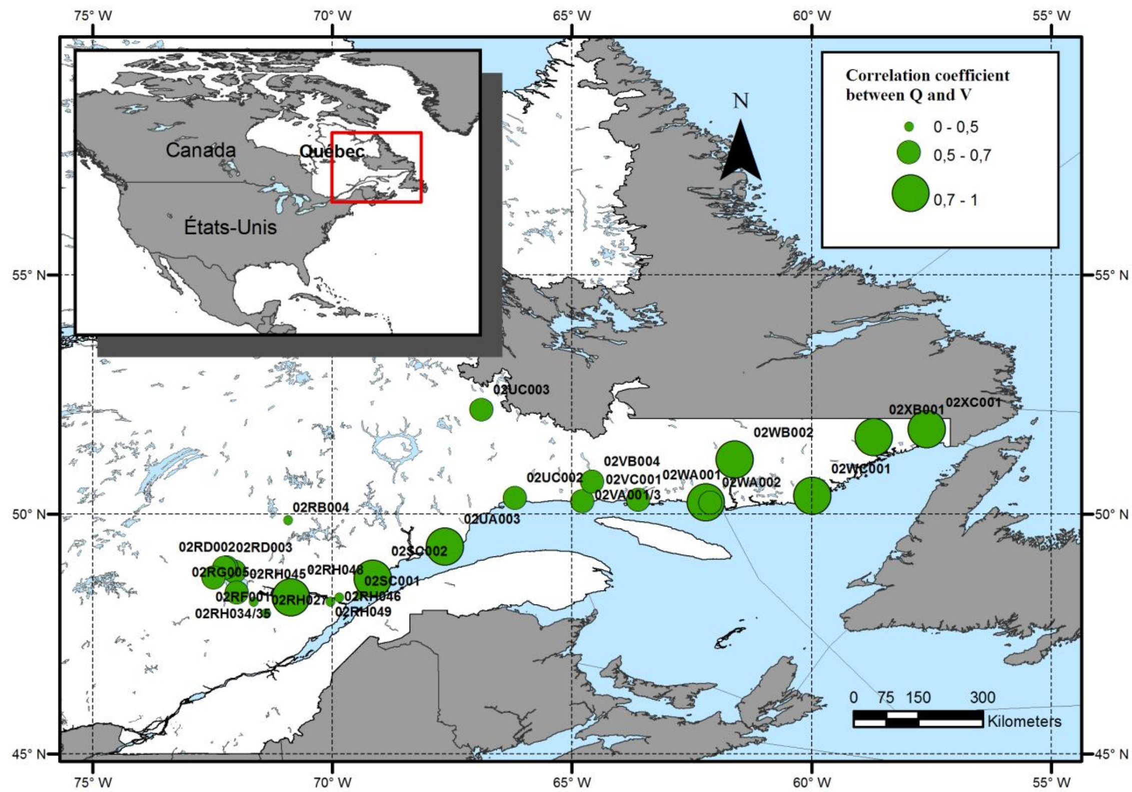
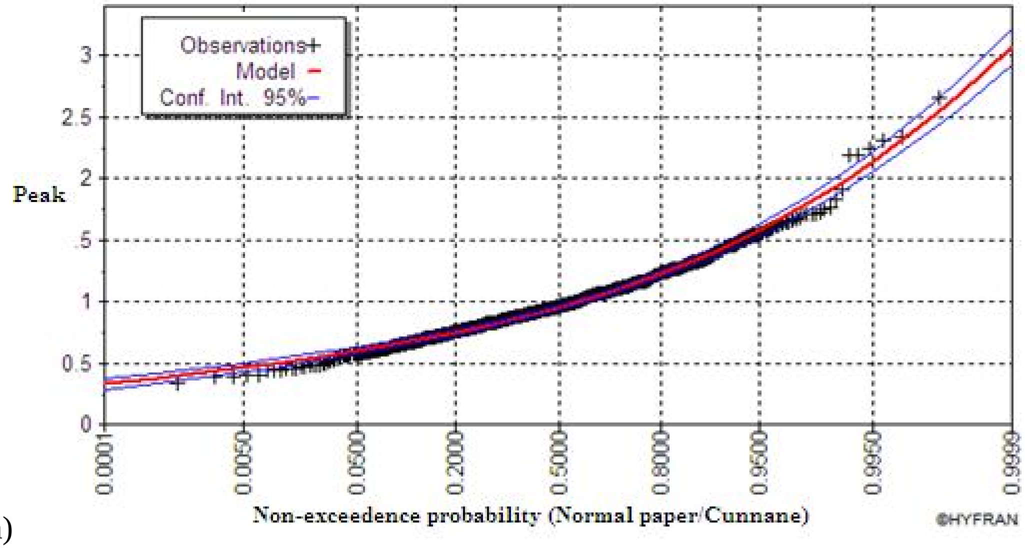
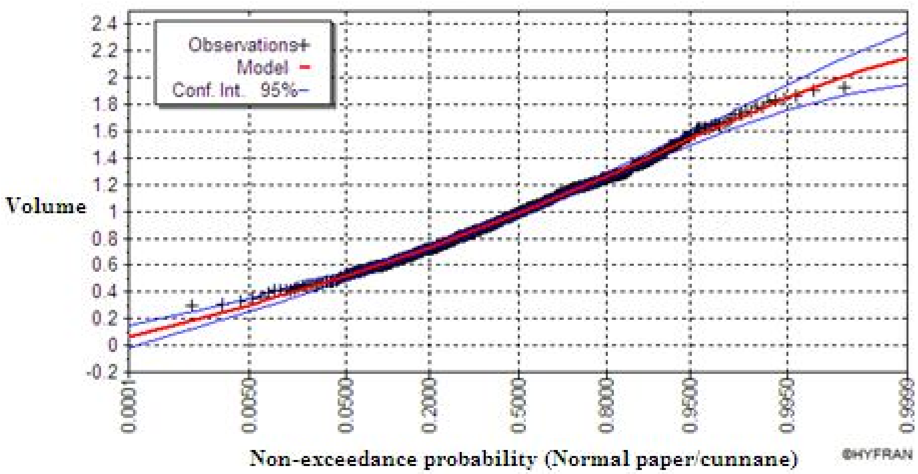


Figure 1 : Geographical chart of the location of the sites



a)



b)

Figure 2 : Fitting of the marginal distribution of a) *Q* and b) *V*.

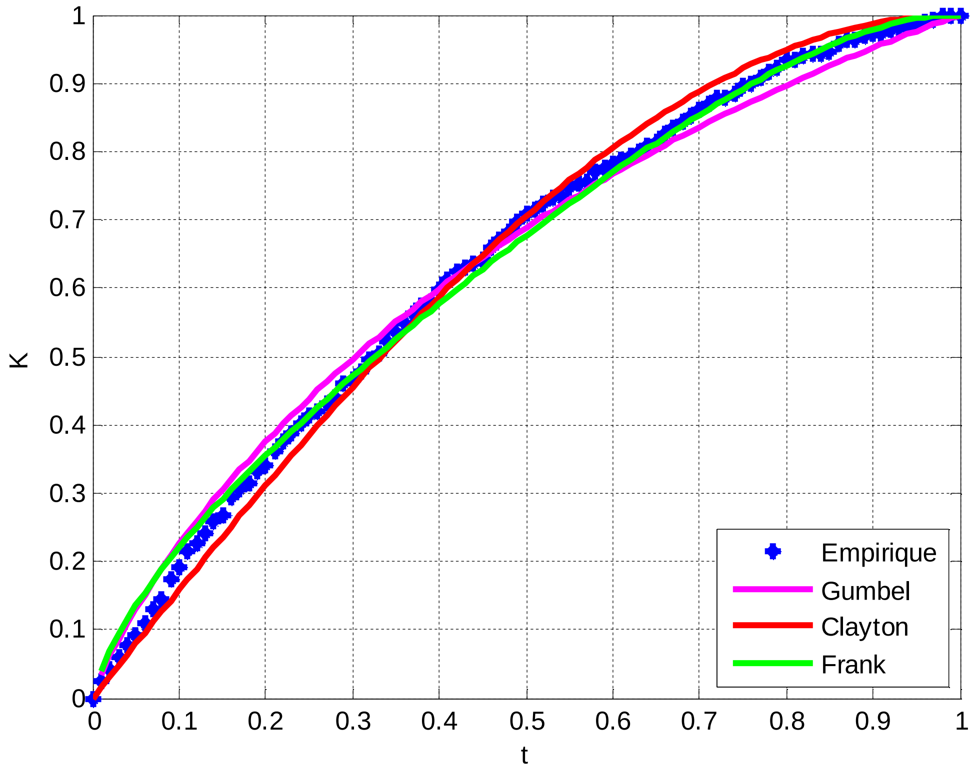


Figure 3 : Copula fitting using K-function

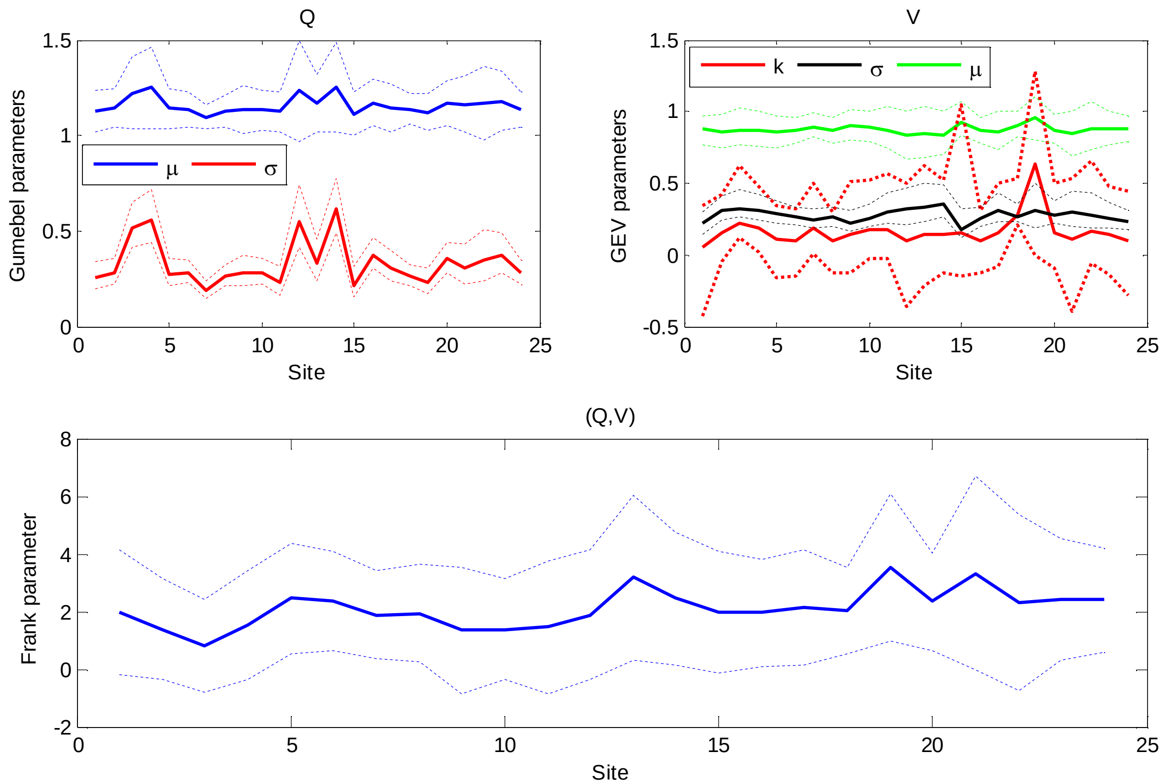


Figure 4: Parameters of marginal distributions and copula. Dashed lines indicate the confidence interval corresponding to each parameter

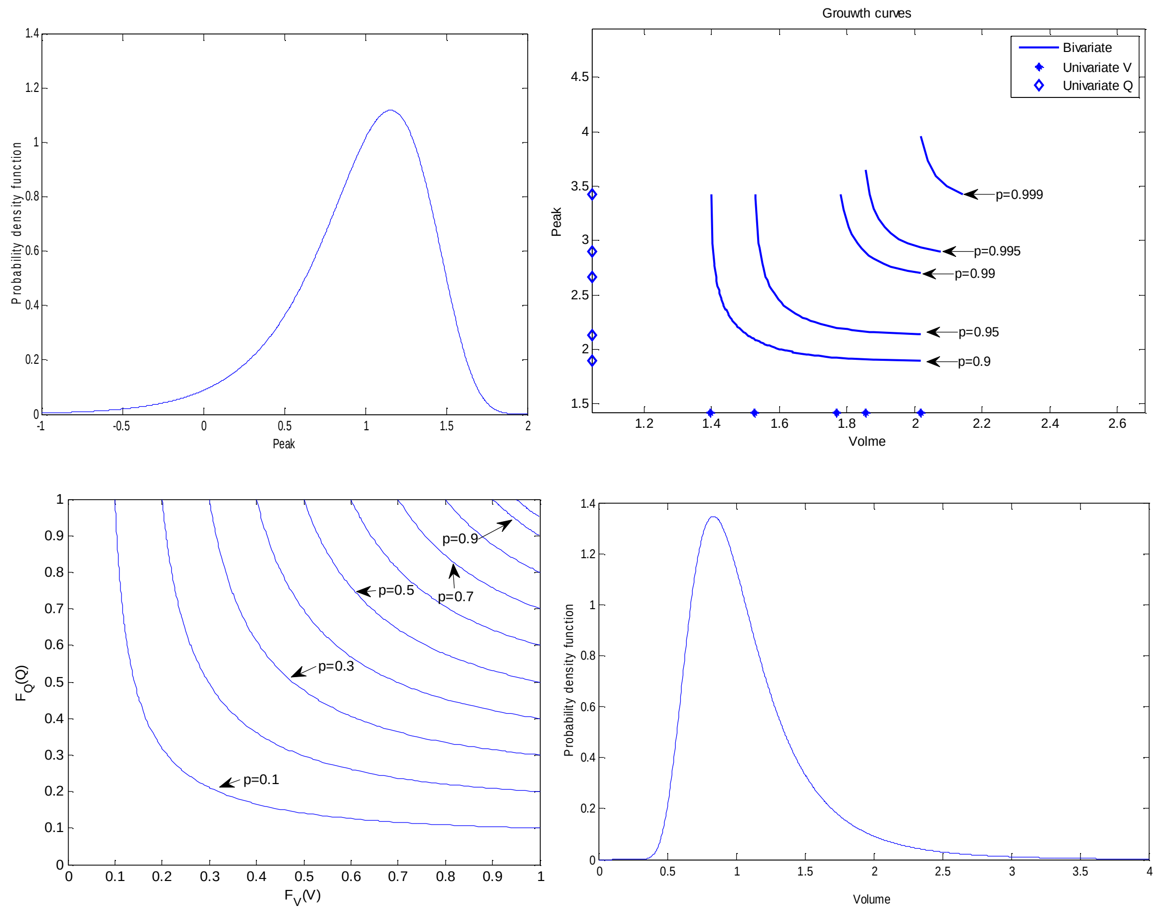


Figure 5 : Estimated regional bivariate and univariate growth curves, quantile curve in the unit square and the marginal distributions for *Q* and *V*

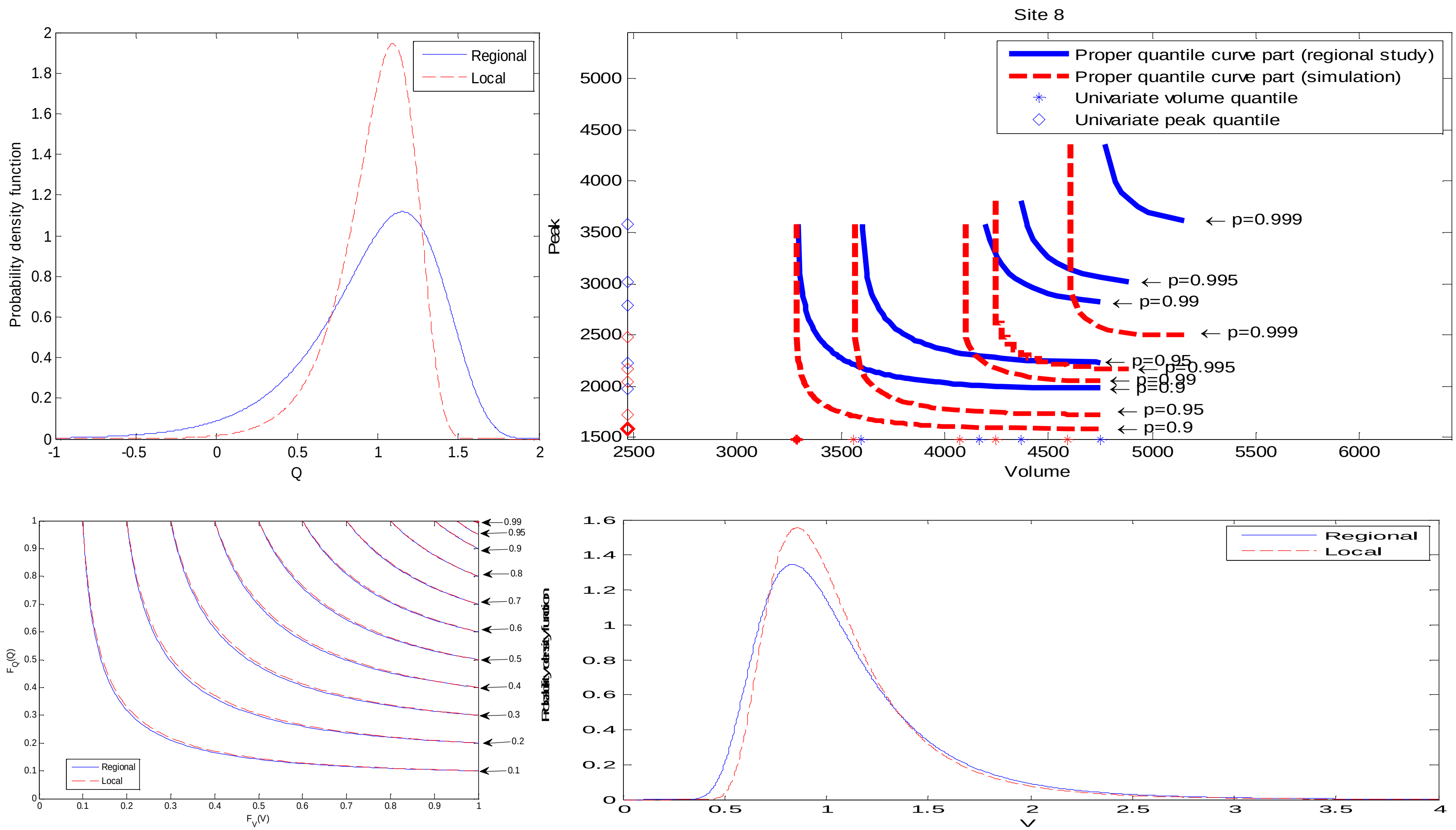


Figure 6a : Univariate and bivariate quantiles corresponding to a nonexceedance probability *p*=0.9, 0.95, 0.99, 0.995 and 0.999 in Mistassibi, quantile curve in the unit square and side panels showing the marginal distributions (local and regional) of *Q* and *V*

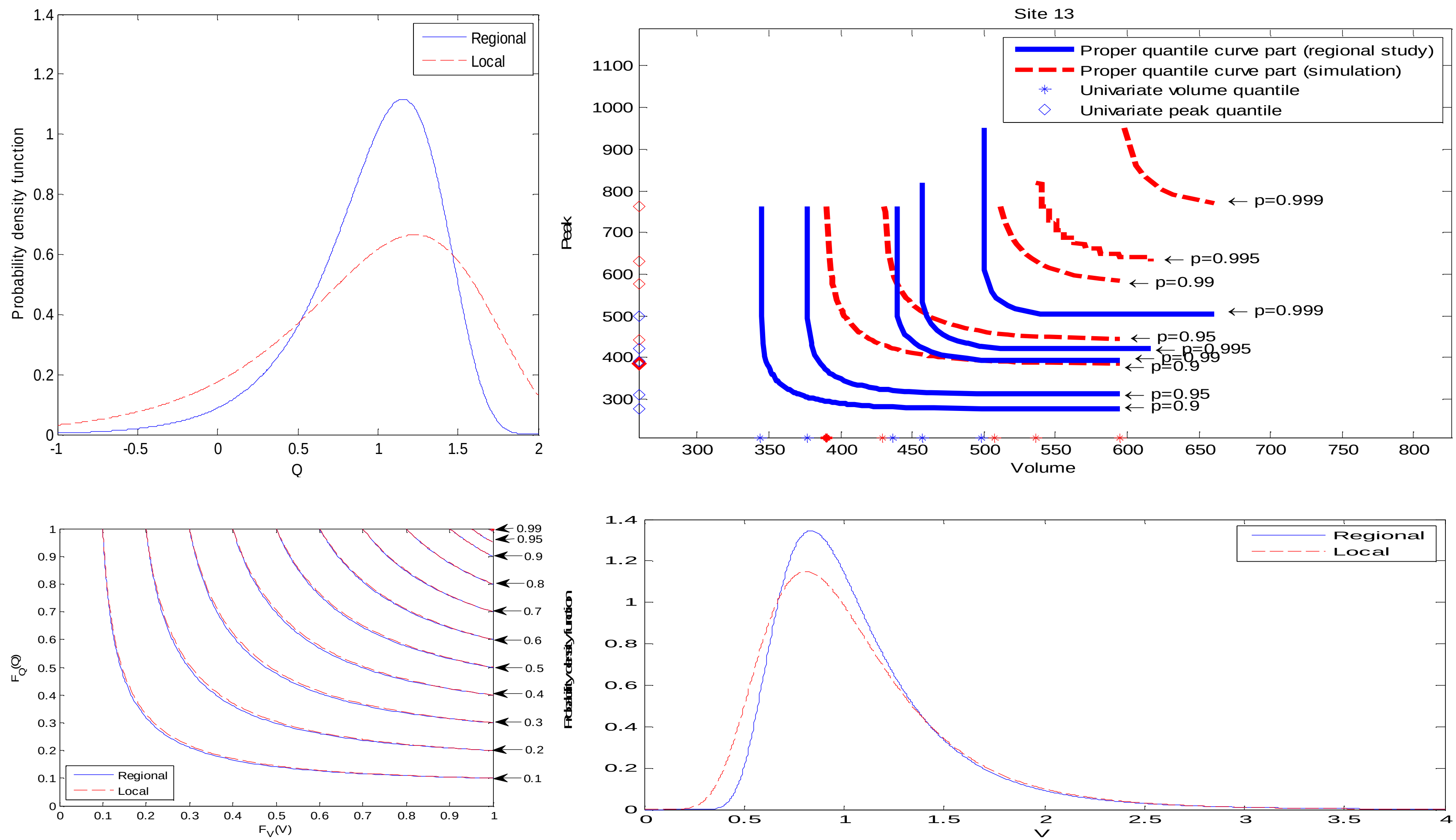


Figure 6b: Univariate and bivariate quantiles corresponding to a nonexceedance probability *p*=0.9, 0.95, 0.99, 0.995 and 0.999 in Des Escoumins , quantile curve in the unit square and side panels showing the marginal distributions (local and regional) of *Q* and *V*.

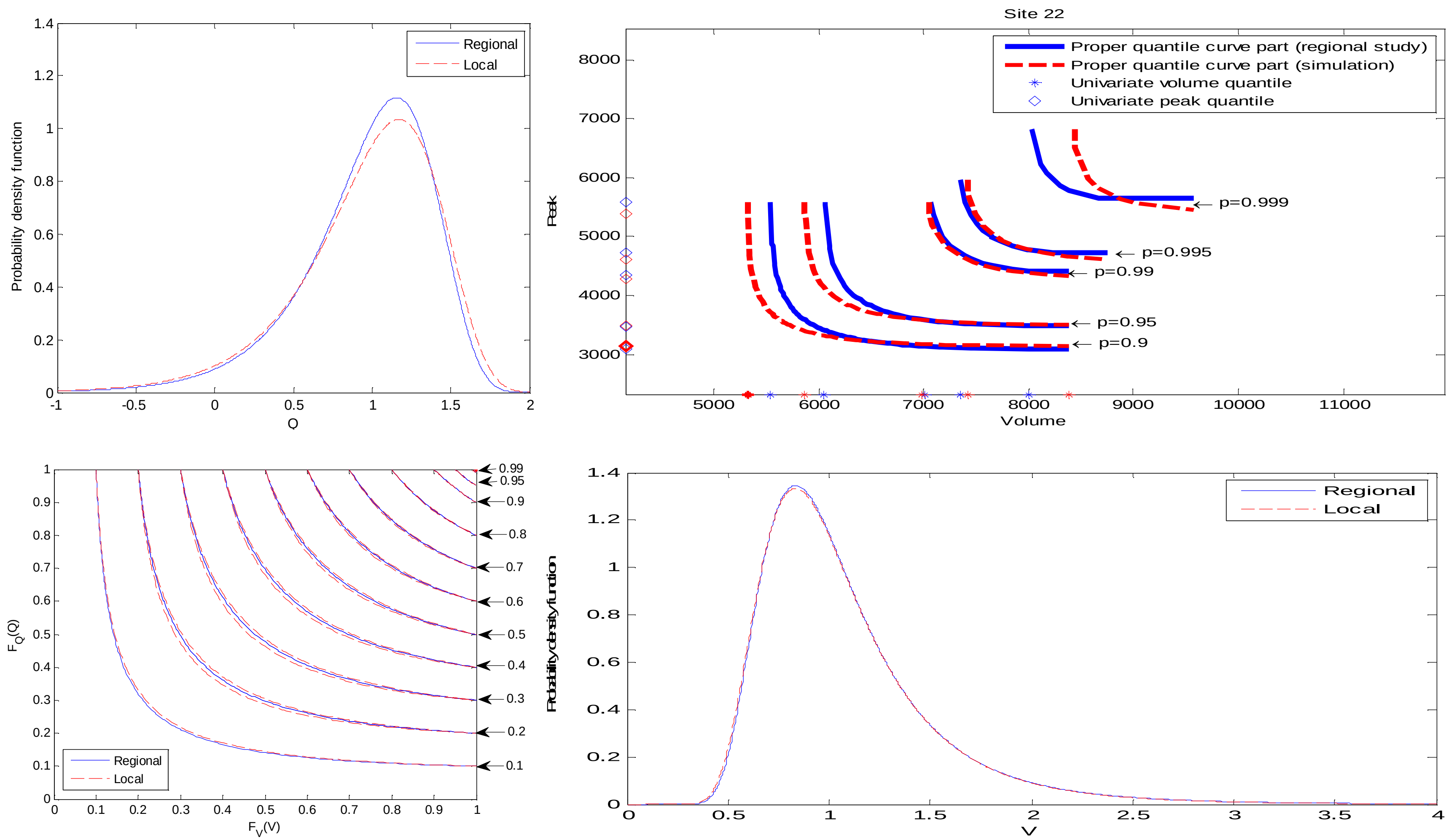


Figure 6c: Univariate and bivariate quantiles corresponding to a nonexceedance probability *p*=0.9, 0.95, 0.99, 0.995 and 0.999 in Natashquan , quantile curve in the unit square and side panels showing the marginal distributions (local and regional) of *Q* and *V*.

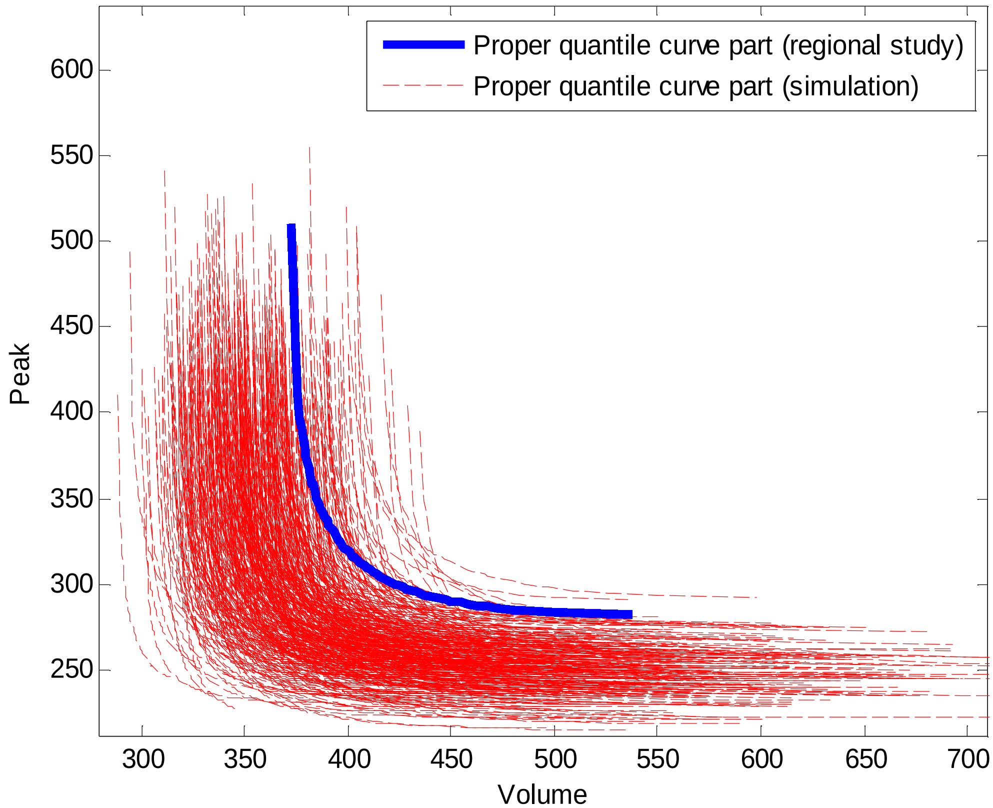


Figure 7 : Bivariate quantiles (Regional and the 500 simulation) corresponding to a nonexceedance probability *p*=0.9 in the Petit Saguenay station.

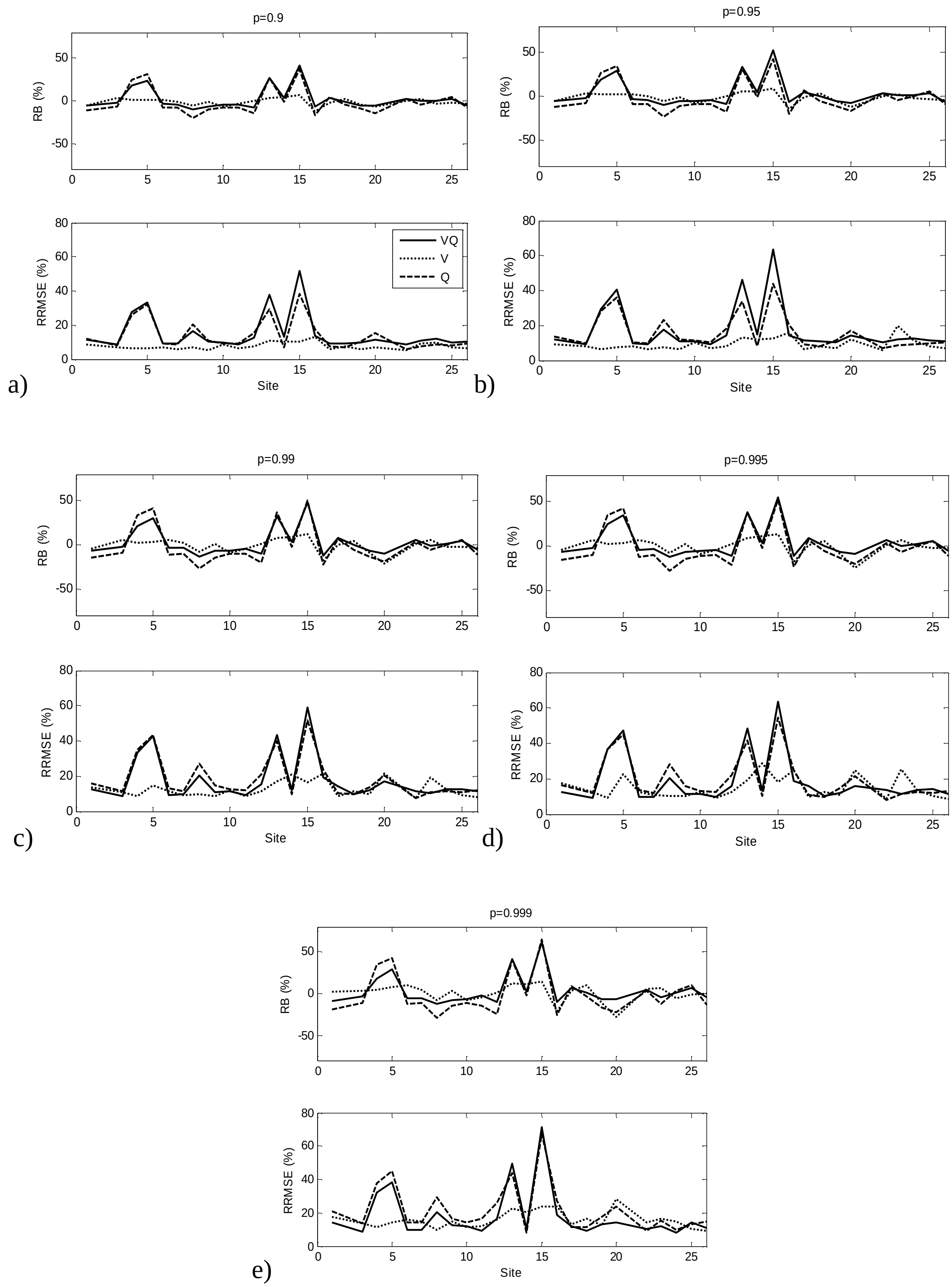
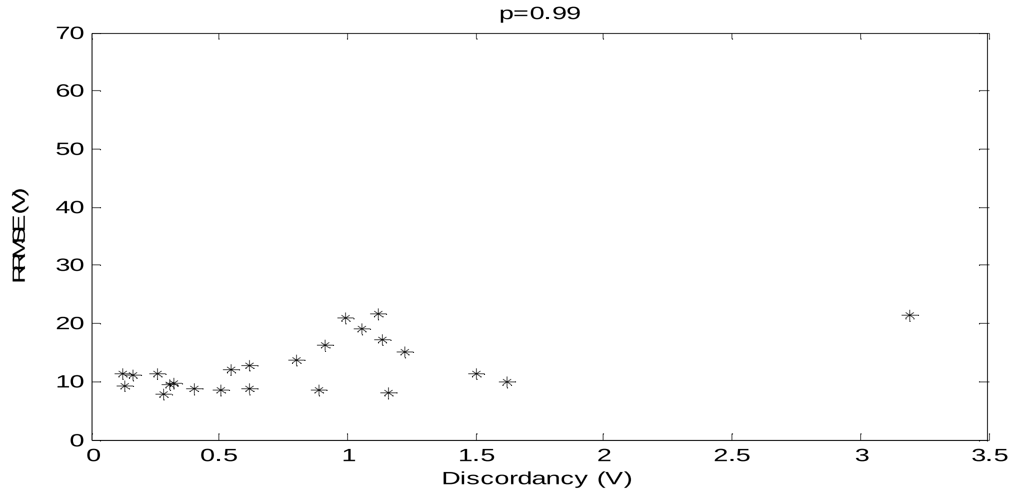
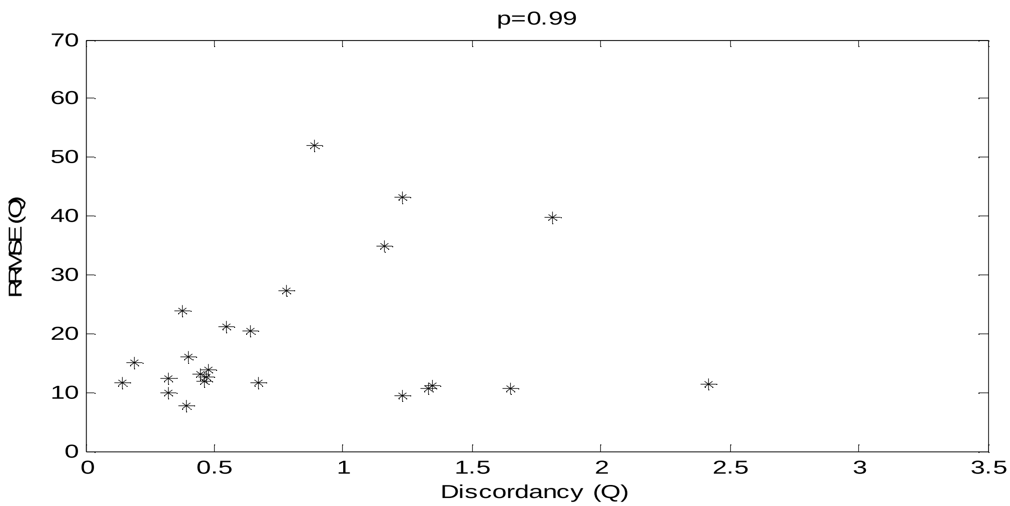


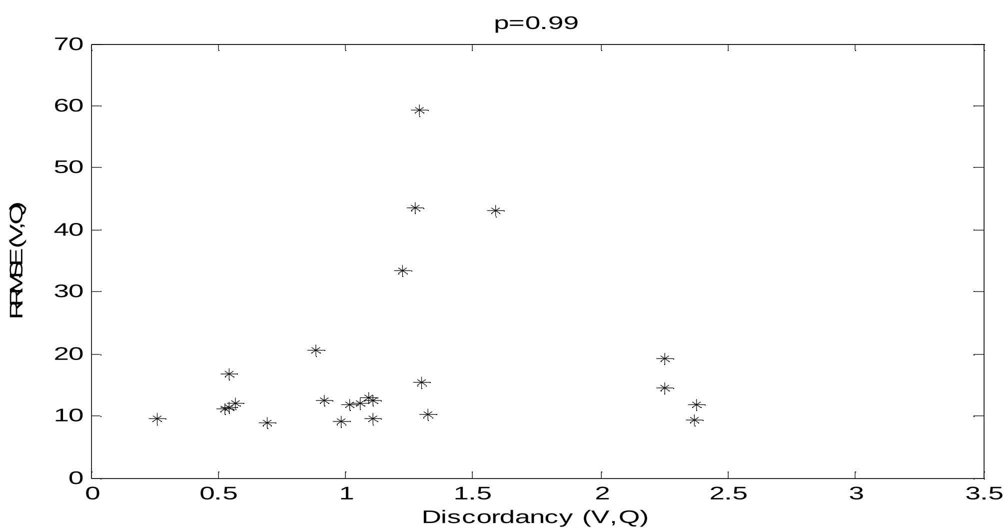
Figure 8 : Performance of the univariate and bivariate quantiles for each site with a) p=0.9, b) p=0.95, c) p=0.99, d) p=0.995 and d) p=0.999. Continuous line: *VQ*; dotted line: *V* and dashed line: Q.



a)

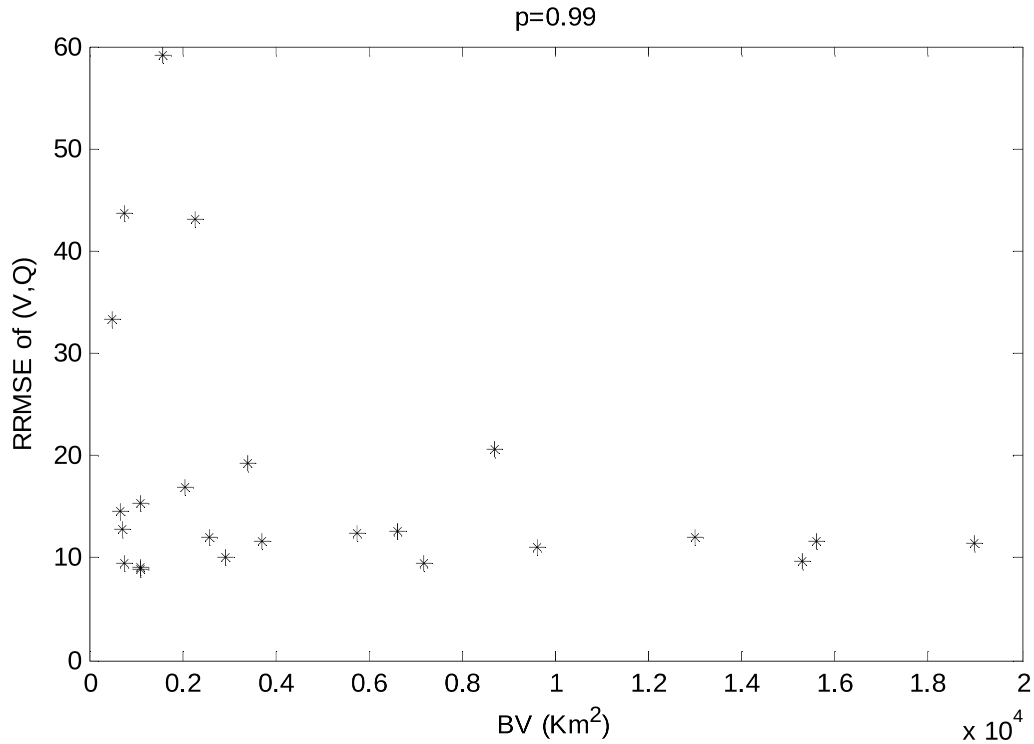


b)

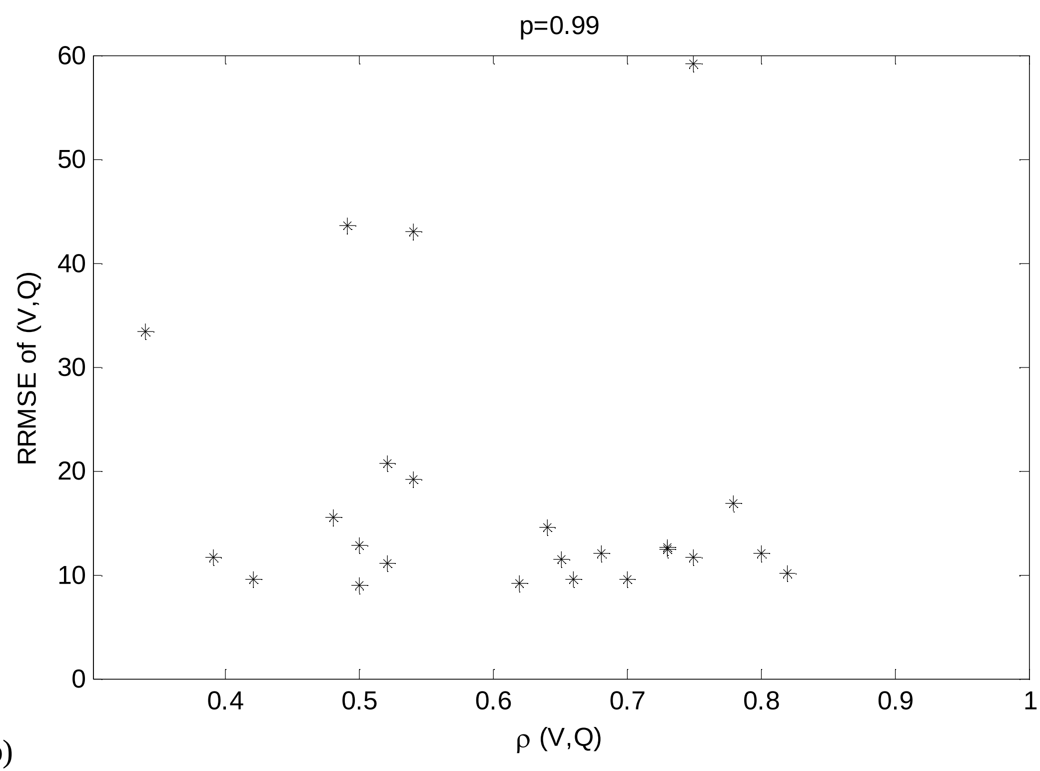


c)

Figure 9 : RRMSE (%) of the three models with respect to the corresponding discordance values for *p*=0.99: a) margin for V, b) margin for Q, and c) bivariate



a)



b)

Figure 10 : RRMSE of bivariate quantile for *p*=0.99 with respect to a) watershed area (BV) and b) correlation between *V* and *Q*.

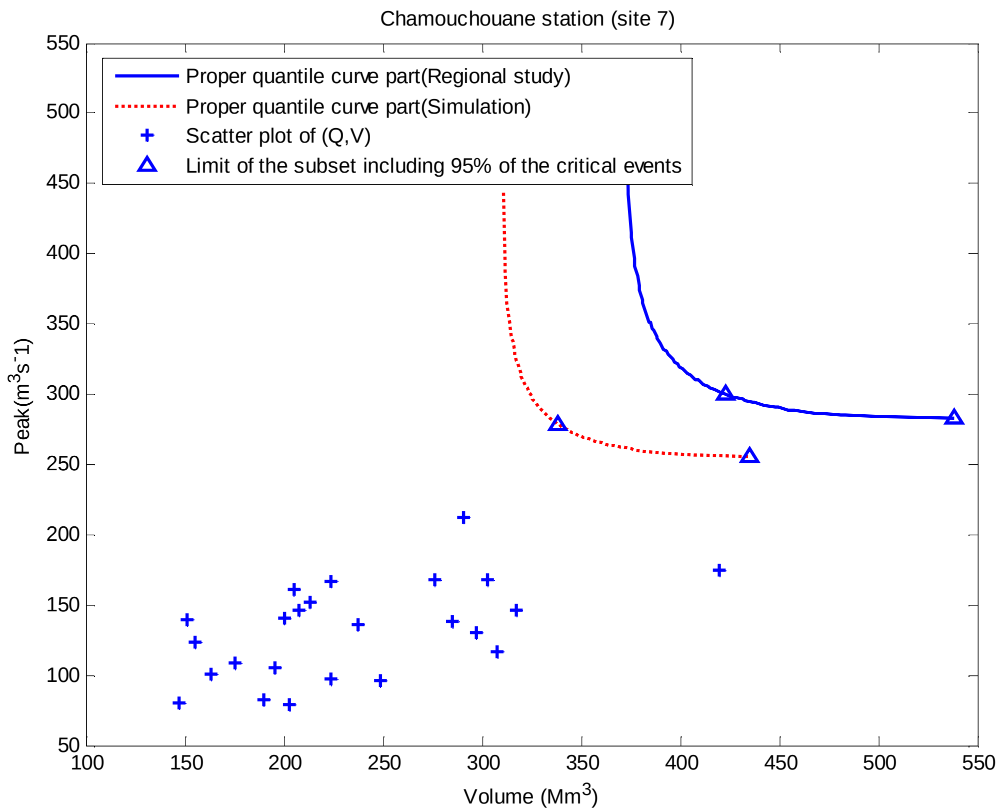


Figure 11 : Bivariate quantiles of Chamouchouane station corresponding to a nonexceedance probability p=0.9 with scatter plot of (Q,V) and the limit of subset that includes the critical events with probability (1-α)=0.95. Simulation in dotted line and sample data in solid line