Multivariate homogeneity testing in a northern case study in the province of Quebec, Canada

Fateh Chebana*1; Taha B.M.J.Ouarda1; Laura Fagherazzi2; Pierre Bruneau3; Marc Barbet3; Salaheddine El Adlouni 1 and Marco Latraverse2

1Canada Research Chair on the Estimation of Hydrometeorological Variables, Hydro-Québec/NSERC Chair in statistical Hydrology, INRS-ETE, University of Quebec, 490 rue de la Couronne, Quebec, QC, Canada
2Hydro-Québec, Complexe Desjardin, Montreal, QC., Canada
3Hydro-Québec, 855 Ste-Catherine est, Montreal, QC, Canada

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

In regional frequency analysis, the examination of the regional homogeneity represents an important step of the procedure. Flood events possess multivariate characteristics which cannot be handled by classical univariate regional procedures. For instance, classical procedures do not allow to assess regional homogeneity while taking into consideration flood peak, volume and duration. Chebana and Ouarda (2007) proposed multivariate discordancy and homogeneity tests. They carried out a simulation study to evaluate the performance of these tests. In the present paper, practical aspects are investigated jointly on flood peak and flood volume of a data set from the Côte-Nord region in the province of Quebec, Canada. It is shown that, after removing the discordant sites, the remaining ones constitute a homogeneous region for the volumes and heterogeneous region for the peaks. However, if both variables are jointly considered, the obtained region is possibly homogeneous. Furthermore, the results demonstrate the usefulness of the bivariate test to take into account the dependence structure between the variables representing the event, and to take advantage of more information from the hydrograph.
1. Introduction

Most hydrological events are described by several correlated variables. Multivariate representations of hydrologic phenomena include, for instance, storm duration and intensity (Yue, 2001a; Salvadori and De Michele, 2004); flood peak, volume and duration (Ashkar, 1980; Yue et al., 1999; Ouarda et al., 2000; Yue, 2001b; Shiau, 2003; De Michele et al., 2005; Zhang and Singh 2006) and drought volume, duration and magnitude (Kim et al., 2003; Ashkar et al., 1998). It is essential to understand the multivariate characteristics of such events for several engineering planning, design and management activities. Snyder (1962) and Wong (1963) can be considered as the first authors to carry out multivariate analysis in hydrology.

The study of the joint probabilistic behaviour of two or more correlated random variables that characterize the event is necessary for a thorough understanding of multivariate hydrological events. Copulas have recently been shown to represent a useful statistical tool for hydrological applications bringing the dependence behaviour between variables (e.g. Salvadori and De Michele, 2004). To represent the joint probability distribution of flood peak and volume and the joint probability distribution of flood volume and duration, Yue et al. (1999) used the Gumbel mixed model with standard Gumbel marginal distributions. Yue (2001b) and Shiau (2003) used the Gumbel logistic model with standard Gumbel marginal distributions to model flood volume and peak for different basins. El Adlouni et al. (2004) presented several copulas to model flood peak and volume with respectively Gumbel and Gamma marginal distributions.

Generally, extreme events are rare and the records are short. Consequently, the at-site frequency estimation is difficult and/or not reliable. Regional frequency analysis (RFA) is proposed as a method to overcome this lack of data. Hence, RFA is commonly used for the estimation of extreme hydrological events at sites where little or no data is available. It is based on the transfer of data available from other stations in the same hydrologic region. The delineation of hydrological homogeneous regions and the regional estimation are the two main steps of a RFA. Several authors investigated this procedure with different approaches including

The literature on flood frequency analysis can be classified into four classes according to the local/regional and univariate/multivariate aspects. In the first two classes there are the local-univariate and regional-univariate studies where many references can be found in the literature. The third class contains local-multivariate flood frequency analysis (e.g., Ashkar, 1980; Yue et al., 1999; Ouarda et al., 2000; Yue, 2001b; Shiau, 2003; De Michele et al., 2005 and Zhang and Singh, 2006). However, very little attention has been given to the forth class which consists in regional-multivariate studies (Ouarda et al., 2000 and Chebana and Ouarda, 2007). Ouarda et al. (2000) carried out a canonical correlation analysis procedure for a joint regional study of flood peak and volume in the province of Quebec, Canada.

Chebana and Ouarda (2007) proposed discordancy and homogeneity tests based on multivariate L-moments and copulas. The proposed multivariate discordancy and homogeneity tests are extensions of those given by Hosking and Wallis (1993). Chebana and Ouarda (2007) evaluated the performances of these multivariate tests using simulations. They demonstrated, for a given extreme event, the importance of jointly and simultaneously treating all variables and identifying a single homogeneous region. In the present paper, practical aspects of these multivariate tests are studied based on a real-word regional data set. The data set corresponds to sites from the Côte-Nord region in the eastern part of the province of Quebec, Canada. The application is carried out on flood event and the focus is on the volume and peak variables.

The paper is organized as follows. Section 2 contains the theoretical background, including flood characteristics, bivariate modeling, multivariate L-moments and the discordancy and homogeneity tests. Section 3 is devoted to the description of the case study. The procedure
followed in this study is presented in Section 4 whereas Section 5 deals with the corresponding results. Concluding remarks are presented in Section 6.

2. Background

In this section, the background elements to apply the multivariate discordancy and homogeneity tests are presented. Flood characteristics, bivariate modeling including copulas and marginal distributions, and the discordancy and homogeneity tests are briefly described.

2.1 Flood characteristics

In Figure 1, a typical flood hydrograph is illustrated. A flood hydrograph is mainly characterized by its volume, duration and peak. Flood duration has to be determined first in order to compute flood volume. Flood duration can be determined whenever the start date \( s_i \) and end date \( e_i \) are identified for the \( i \)th series as \( D_i = e_i - s_i \). The annual flood volume series can be constructed using the following formula (see e.g., Yue et al., 1999):

\[
V_i = \sum_{j=s_i}^{e_i} q_{ij} - \frac{1}{2} (q_{i0} + q_{ie}), \quad i = 1,2,\ldots
\]

where \( q_{ij} \) represents the observed streamflow value at the \( j \)th day of the \( i \)th year, \( q_{i0} \) and \( q_{ie} \) are respectively the observed daily streamflow values on the start date and end date of flood runoff for the \( i \)th year. The annual flood peak series is given by

\[
Q_i = \max \{ q_{ij}, j = s_i, s_i + 1,\ldots, e_i \}
\]

2.2 Bivariate flood modeling

In bivariate modeling, one should obtain a joint bivariate distribution for the variables. However, one should also distinguish the dependence structure from the margins. To this end, one needs to specify three elements: a copula to describe the dependence structure between the two random variables, along with a marginal distribution for each variable.
2.2.1 Copulas

In the remainder of the paper we denote $F_1$ and $F_2$ the marginal distribution functions of given random variables $X^{(1)}$ and $X^{(2)}$, and $F_{1,2}$ is the joint distribution function of $(X^{(1)}, X^{(2)})$.

Independently of the marginal distributions, a copula is a description and a model of the dependence structure between the two random variables. To overcome the limitations of classical dependence measures, copulas have recently received increasing attention in various science fields (see for instance Nelsen, 1999). A copula is a function $C: I \times I \rightarrow I$ ($I = [0, 1]$) such that:

- for all $u, v \in I$: $C(u, 0) = 0$, $C(u, 1) = u$, $C(0, v) = 0$, and $C(1,v) = v$;
- for all $u_1, u_2, v_1, v_2 \in I$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$: $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$

The link between copulas and bivariate distributions is provided by Sklar’s (1959) result. It states that there exists a copula $C$ such that:

$$F_{1,2}(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \quad \text{for all } x_1 \text{ and } x_2$$

When $F_1$ and $F_2$ are continuous, the copula $C$ is unique.

Archimedean and extreme value (EV) copulas represent classes of particular interest. The class of EV copulas is given by the formula (Pickands, 1981):

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

where the dependence function $A$ is convex and defined on $[0, 1]$ with $\max \{t, 1-t\} \leq A(t) \leq 1$. A bivariate Archimedean copula is characterized by the expression:

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

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

As it is already shown in previous studies, e.g. Salvadori and De Michele (2004), Archimedean copulas represent convenient multivariate models for hydrological flood events.
When the multivariate context is involved, some practical questions can be raised regarding copulas, for instance:

- How a copula can be fitted to a given sample?
- How copula’s parameters can be estimated?
- And how a sample can be generated from a model defined through a given copula?

Partial answers to these questions are given for the Archimedean and extreme value copulas.

First, the fitting problem is resolved for Archimedean copulas. According to Genest and Rivest (1993), an Archimedean copula, with a generator function $\psi$, is characterized by the following function:

$$\psi(K(z)z) = -z$$

which can be estimated by:

$$\hat{K}(z) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{[w_i < 1]} \text{ where } w_i = \frac{1}{N-1} \sum_{i=1}^{N} \mathbb{I}_{[x_{i_1} < x_{i_2}, x_{i_3} < x_{i_4}], i = 1, \ldots, N}$$

for a given bivariate sample $(x_{11}, x_{12}), (x_{21}, x_{22}), \ldots, (x_{N1}, x_{N2})$. It is shown in Genest and Rivest (1993) that $\hat{K}$ is a consistent estimator of $K$ under weak regularity conditions.

It is shown in several studies (e.g. Yue, 2001b and Shiau, 2003) that an interesting copula to model flood characteristics is the Gumbel logistic copula given by:

$$C_m(x, y) = \exp \left\{ \left[ (-\log x)^m + (-\log y)^m \right]^{1/m} \right\}, m \geq 1, \ 0 \leq x, y \leq 1$$

which is an Archimedean copula with generator function $\psi(x) = (-\log x)^m$, and it is also an extreme value copula with dependence function $A(t) = \left( t^m + (1-t)^m \right)^{1/m}$. The corresponding function $K$ defined in (6) for the Gumbel logistic copula is given by $K_m(z) = z - \frac{z \log(z)}{m}$.

Second, the parameter estimation problem for Archimedean copulas is also resolved. In particular, the parameter $m$ of the copula $C_m$ is related to the correlation coefficient $\rho$ through the equation (Gumbel and Mustafi, 1967):
\[ m = \frac{1}{\sqrt{1 - \rho}}, \quad 0 \leq \rho < 1 \] (9)

Hence, it can be estimated by a plug-in of the empirical version of the correlation coefficient in equation (9). However, it can also be estimated by:

\[ \hat{m} = 1 + \frac{\hat{\tau}_{1,2}}{1 - \hat{\tau}_{1,2}} \] (10)

where \( \hat{\tau}_{1,2} \) is the empirical estimator of

\[ \tau_{1,2} = 4E \left[ F_{1,2}(X^{(1)}, X^{(2)}) \right] - 1 \] (11)

which is a version of the Kandall’s tau coefficient for the random vector \( (X^{(1)}, X^{(2)}) \). A simple estimator of the Kandall’s tau coefficient is given by \( \hat{\tau}_{1,2} = 4\bar{G} - 1 \) where \( \bar{G} \) is the mean of the “pseudo-sample” \( G_i = \frac{1}{n-1} \# \left\{ (X_j^{(1)}, X_j^{(2)}) : X_j^{(1)} < X_i^{(1)}, X_j^{(2)} < X_i^{(2)} \right\}, \quad i = 1, \ldots, n \) (see Genest and Rivest, 1993).

Regarding the last question, related to the generation of samples from the variables \( (X^{(1)}, X^{(2)}) \) according to the extreme value copula, an algorithm is developed by Ghoudi et al. (1998). The algorithm is summarized in the following. Let \( U_1, U_2 \) 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 \( G_Z(z) = z + z(1 - z)A'(z)/A(z), \quad 0 \leq z \leq 1 \). This algorithm consists of the following steps:

1. Simulate \( Z \);
2. Given \( Z \), take \( W = U_1 \) with probability \( p(Z) \) and \( W = U_1 U_2 \) with probability \( 1 - p(Z) \), where \( p(z) = z(1 - z)A''(z)/(A(z)g_Z(z)) \);
3. Set \( X^{(1)} = W^{Z/A(Z)} \) and \( X^{(2)} = W^{(1-Z)/A(Z)} \).
When using this algorithm in practice it is important to take into consideration the numerical nonparametric smoothing, since it depends on functions related to the first and second derivatives of the function \( A \). Despite the general validity of this procedure, extra information about the model, e.g. parametric form of \( A \), can be useful to increase the speed and accuracy of the generation algorithm.

### 2.2.2 Marginal modeling

The 2-parameter Gumbel distribution can be used to model the marginal flood variables (Yue, 2001b and Shiau, 2003). However, as it is indicated in Hosking and Wallis (1993) and Chebana and Ouarda (2007), it is preferable to employ a 4-parameter Kappa distribution for the homogeneity test. Its cumulative distribution function is given by:

\[
F(x) = \left[ 1 - h \left( 1 - \frac{x - u}{\alpha} \right)^{\frac{1}{h}} \right]^{1/\kappa}
\]  

(12)

with parameters \( u \) (position), \( \alpha \) (scale), \( \kappa \) and \( h \) (shape).

The parameters of the Kappa distribution can be estimated by the \( L \)-moment method (Hosking and Wallis, 1997). Indeed, if we denote respectively by \( \hat{\lambda}_k \) the \( L \)-moment and \( \hat{t}_k \) the \( L \)-moment coefficient of order \( k \), the first Kappa \( L \)-moments are given by:

\[
\begin{align*}
\hat{\lambda}_1 &= u + \alpha (1 - g_1) / \kappa \\
\hat{\lambda}_2 &= \alpha (g_1 - g_2) / \kappa \\
t_3 &= (-g_1 + 3g_2 - 2g_3) / (g_1 - g_2) \\
t_4 &= (-g_1 + 6g_2 - 10g_3 + 5g_4) / (g_1 - g_2),
\end{align*}
\]

(13)

where

\[
g_r = \begin{cases} 
\frac{r \Gamma(1 + \kappa) \Gamma(r/h)}{h^{\kappa} \Gamma(1 + \kappa + r/h)}, & h > 0 \\
\frac{r \Gamma(1 + \kappa) \Gamma(-\kappa - r/h)}{(-h)^{\kappa} \Gamma(1 - r/h)}, & h < 0
\end{cases}
\]

(14)
Therefore its parameters can be estimated by the use of equations (13) but there are no simple and direct expressions. Hosking (1996) developed a routine to find numerically the Kappa \( L \)-moment parameter estimators.

### 2.3 Multivariate \( L \)-moments

Instead of traditional moments, for statistical inference of hydrological variables, the \( L \)-moment approach offers strong advantages for modeling heavy-tailed distributions. For a review related to \( L \)-moments the reader can consult Hosking and Wallis (1997). Multivariate \( L \)-moments are principally developed by Serfling and Xiao (2007).

By analogy with the covariance representation of the \( L \)-moment of order \( k \), multivariate \( L \)-moments are matrices \( \Lambda_k \) with \( L \)-comoment elements defined by:

\[
\lambda_{k[i,j]} = \text{Cov}\left(X^{(i)}, P^*_k F_j(X^{(j)})\right), \quad i, j = 1, 2 \text{ and } k = 2, 3, \ldots \tag{15}
\]

where \( P^*_k \) is the so-called shifted Legendre polynomial. As it can be seen, the elements \( \lambda_{k[i,j]} \) and \( \lambda_{k[j,i]} \) are not necessarily equal. The first \( L \)-comoment elements are given by:

\[
\begin{align*}
\lambda_{2[12]} &= 2\text{Cov}\left(X^{(1)}, F_2(X^{(2)})\right) \\
\lambda_{4[12]} &= 6\text{Cov}\left(X^{(1)}, \left(F_2(X^{(2)}) - 1/2\right)^2\right) \\
\lambda_{4[12]} &= \text{Cov}\left(X^{(1)}, 20\left(F_2(X^{(2)}) - 1/2\right)^3 - 3\left(F_2(X^{(2)}) - 1/2\right) + 1\right)
\end{align*}
\tag{16}
\]

The \( L \)-comoment coefficients are given by:

\[
\tau_{2[12]} = \frac{\lambda_{2[12]}}{\lambda_{2[11]}} \quad \text{and} \quad \tau_{k[12]} = \frac{\lambda_{k[12]}}{\lambda_{2[12]}}, \quad \text{for } k = 3, 4, \ldots \tag{17}
\]

where \( \lambda_{k[j]} = \lambda_{k[j]} \) is the classical univariate \( k \)th \( L \)-moment of the variable \( X^{(j)} \). The matrix of the \( L \)-comoment coefficients is written as:

\[
\Lambda_k^* = (\tau_{k[j]}),_{i,j=1,2} = \begin{pmatrix}
\tau_{k[11]} & \tau_{k[12]} \\
\tau_{k[21]} & \tau_{k[22]}
\end{pmatrix}, \quad \text{for } k = 2, 3, \ldots \tag{18}
\]
and for $k = 1$, the first order bivariate $L$-moment corresponds to the mean vector $\hat{\lambda}_1 = E(X^{(1)}, X^{(2)})'$.

### 2.4 Discordancy and homogeneity tests

#### 2.4.1 Discordancy

A preliminary screening step, before proceeding with the homogeneity analysis, consists in identifying discordant sites among a set of $N$ sites. A multivariate extension of the Hosking and Wallis (1993) discordancy test is proposed by Chebana and Ouarda (2007). It is defined for each site $i$ using the matrix $U'_i = \begin{bmatrix} \Lambda^{(2)}_2 & \Lambda^{(3)}_3 & \Lambda^{(4)}_4 \end{bmatrix}$ which is composed by the three $L$-moment matrices $\Lambda^{(2)}_2$, $\Lambda^{(3)}_3$ and $\Lambda^{(4)}_4$ defined by (18). Hence, a site $i$ is discordant, with respect to the considered set of sites, if $\|D_i\|$ takes large values, where:

$$
D_i = \frac{1}{3} (U_i - \overline{U})' S^{-1} (U_i - \overline{U}) ,
$$

$$
S = \frac{1}{N - 1} \sum_{i=1}^{N} (U_i - \overline{U})(U_i - \overline{U})' ,
$$

$$
\overline{U} = \frac{1}{N} \sum_{i=1}^{N} U_i ,
$$

$\|A\|$ denotes the spectral norm of a matrix $A$ given by $\|A\| = \sqrt{\text{maximum eigenvalue of } A' A}$ and $A'$ is the transpose of a matrix or a vector $A$. Note that Chebana and Ouarda (2007) considered other matrix norms and indicated that no significant difference was observed in the results obtained with the other norms.

The constant $c = \chi_{1-0.05}(3)/3 = 2.6$ may be considered as a critical value for $\|D_i\|$ for large regions, where $\chi_{1-\alpha}(d)$ denotes the quantile of a chi-square distribution of order $\alpha$ with $d$ degrees of freedom. Chebana and Ouarda (2007) proposed the use of a bootstrap technique to determine a critical value for short values of $N$. Hosking and Wallis (1997) advised to examine the data for sites with the largest $\|D_i\|$ values, regardless of the magnitude of these values.
2.4.2 Homogeneity test

The following multivariate homogeneity test is proposed by Chebana and Ouarda (2007). It is an extension of the univariate test proposed by Hosking and Wallis (1993). It can be summarized in the followings. Let $V_{II}$ be the statistic defined as:

$$V_{II} = \frac{\sum_{i=1}^{N} n_i \left( \Lambda_2^{*}\right)^2}{\sum_{i=1}^{N} n_i}$$

(22)

where $\Lambda_2^{*} = \left( \sum_{i=1}^{N} n_i \right)^{-1} \sum_{i=1}^{N} n_i \Lambda_2^{*\left(i\right)}$ and $\Lambda_2^{*\left(i\right)}$ is the $L$-covariation coefficient matrix for site $i$, with record length $n_i, i = 1, ..., N$. In order to get interpretable results of the computed value of the statistic $V_{II}$ from the observations, it is convenient to standardize it by the use of a large number of simulated homogeneous regions. The simulated regions are homogeneous with sites having the same record lengths as their observed counterparts. Hence, the statistic that measures the heterogeneity of a set of sites is given by:

$$H_{II} = \frac{V_{II} - \mu_{V_{sim}}}{\sigma_{V_{sim}}}$$

(23)

where $\mu_{V_{sim}}$ and $\sigma_{V_{sim}}$ are respectively the mean and standard deviation of the $N_{sim}$ values of $V_{II}$ of simulated regions. The EV or Archimedean copulas with the marginal 4-parameter Kappa distributions are the bivariate distributions on which the simulations are carried out to compute $\mu_{V_{sim}}$ and $\sigma_{V_{sim}}$. A region of sites is declared to be homogeneous if $H_{II} < 1$, acceptably homogeneous if $1 < H_{II} < 2$ and definitely heterogeneous if $H_{II} > 2$. Note that in the univariate framework, the statistics $V_{II}$ and $H_{II}$ are equivalent to the classical statistics defined by Hosking and Wallis (1993). For more details concerning the multivariate homogeneity test, the reader is referred to Chebana and Ouarda (2007).
The test statistic $H_\parallel$ is standardized on the basis of the mean and standard-error of $N_{\text{sim}}$ simulated homogeneous regions. The value of $N_{\text{sim}} = 500$ is shown to be appropriate to allow the test to perform well. However, higher values of $N_{\text{sim}}$ allow to improve the estimation of $\mu_{\text{sim}}$ and $\sigma_{\text{sim}}$ and hence to make the right decisions when the values of $H_\parallel$ are close to the thresholds 1 and 2.

3. Case study

The application of the multivariate discordancy and homogeneity tests concerns a regional data set of interest for the Hydro-Québec company. The phenomenon to be studied is the flood, with bivariate characteristics, that is, volume $V$ and spring peak $Q$. The data is from sites of the Côte Nord in the north part of the province of Quebec, Canada. The data set counts $N = 26$ stations with record lengths $n_i$ from 14 to 48. Some information about the data are given in Table 1. The geographical location of the underlying sites is presented in Figure 2.

4. Study procedure

The procedure of the study is composed in the following three main steps:

1. Correlation: Assessment of the correlation coefficient between the variables $V$ and $Q$ for each site.

2. Discordancy:
   a. Evaluation of the discordancy of each site on the basis of each variable $V$ and $Q$ and on the joint variable $(V,Q)$.
   b. Identification of the corresponding discordant sites.

3. Homogeneity: To carry out the homogeneity test given in equation (23), the following steps are required:
   a. Removing the discordant sites identified in 2.b, and using the remaining sites in the following steps.
b. Modeling the joint variable \((V,Q)\): Employing the Archimedean copula characterization (7).

The marginal of both \(V\) and \(Q\) is taken to be the Kappa distribution (12).

c. Estimating the model parameters: For the marginal Kappa distribution the estimation is based on the \(L\)-moment ratios \(1, \bar{t}_2, \bar{t}_4, \bar{t}_4\) with \(\bar{t}_k = \frac{\sum n_i t_i^{(i)}}{\sum n_i}\). Their parameters are estimated using equations (13) and (14). The parameter \(m\) of the copula (8) can be estimated by:

\[
\hat{m} = 1 + \frac{\bar{\tau}_{1,2}}{1 - \bar{\tau}_{1,2}}, \quad \text{where} \quad \bar{\tau}_{1,2} = \frac{\sum_{i=1}^{N} n_i \hat{\tau}_{1,2}^{(i)}}{\sum_{i=1}^{N} n_i}
\]  

(24)

where \(\hat{\tau}_{1,2}\) is the empirical at-site estimator of \(\tau_{1,2}\) defined in (11).

d. Computation of the homogeneity tests:

i. Computation of observed \(V_{\|}\) statistics (equation 22) both on the margins and on the joint variables.

ii. Computation of the statistics \(V_{\text{sim}}\), based on \(N_{\text{sim}} = 500\) generated homogeneous regions. The bivariate regions are generated according to the model defined through the Gumbel logistic copula (8) and Kappa distribution (12) for the margins. For both \(V\) and \(Q\) univariate regions are generated according to the inversion formula of (12). The sampling procedure for the bivariate samples is based on the algorithm developed by Ghoudi et al. (1998) and the inversion formula of (12).

iii. Assessment of the mean \(\mu_{\text{sim}}\) and the standard deviation \(\sigma_{\text{sim}}\) on the basis of the \(N_{\text{sim}}\) values of \(V_{\text{sim}}\).

iv. Evaluation of homogeneity statistics \(H_{\|}\) respectively for the variables \(V, Q\) and \((V,Q)\), by combining \(V_{\|}, \mu_{\text{sim}}\) and \(\sigma_{\text{sim}}\) using the expression (23).
5. Results and discussion

Results of the correlation between the variables $V$ and $Q$ are reported in Table 1. Their values are positive and generally exceed 0.5. This means that the variables $V$ and $Q$ are generally highly correlated. Hence the use of bivariate analysis is of interest compared to the use of two univariate homogeneity tests. Table 2 presents the discordancy results. The sites that may be discordant are identified and their respective $L$-moment values that may cause the discordancy are also identified. Namely, sites 2 and 16 for $V$; site 2 or sites 2 and 3 for $Q$; and sites 2 and 21 for $(V, Q)$. These sites are eliminated to allow application of the respective homogeneity tests. Clearly, the discordant sites are not the same for $V, Q$ or $(V, Q)$.

The application of the Archimedean copula characterization (6) with the estimate (7) leads to fit the Gumbel logistic copula to the bivariate data of each site. The illustration of this fitting is presented in Figure 3 for each site of the data set. Regarding the marginal distributions, the fitting is based on the empirical cumulative distribution function. One of the most used expressions is given by Cunnane (1978):

$$P_k = \frac{k - 0.4}{N + 0.2}$$

for a sorted sample $x_{(1)} \leq ... \leq x_{(k)} \leq ... \leq x_{(N)}$. Figure 4 illustrates, as an arbitrary selected example, the fitting of the marginal distributions and the dependence structure for the data of station 02RF001 (the 7th station). Hence, for both variables $V$ and $Q$ the samples are fitted with the Gumbel distribution, for which the cumulative distribution function is given by:

$$F(x) = \exp\left\{-\exp\left(-\frac{x - \beta}{\alpha}\right)\right\}, \text{ x real, } \alpha > 0 \text{ and } \beta \text{ real}$$

The use of the Gumbel distribution in this context was discussed in several previous studies (e.g., Yue et al., 1999; Yue, 2001b and Shiau, 2003). Note that the Gumbel distribution is covered by the Kappa distribution, since the latter contains the generalized extreme value as a special case.

The detected discordant sites present some special characteristics in terms of their $L$-moments or their physiographical attributes. Site 2 is discordant for all variables. Site 2 has the
lowest mean $l_1$, the largest $L$-CV $t_2$ and the lowest $L$-skewness $t_3$ for the volume. However, for the peak, site 2 has the lowest mean $l_1$, the largest $L$-CV $t_2$ and the lowest $L$-kurtosis $t_4$. Furthermore, site 2 represents the smallest basin in the region. Also, 50% of its basin area is controlled by a reservoir. All this could explain its rejection. Site 3, which may be discordant for $Q$, has a very small value of the $L$-skewness $t_3$. Site 16 is particular because of its short record length ($n = 16$, Table 1) as well as the small values of the three characteristics $t_2$, $t_3$ and $t_4$ for the volume. Site 21 is detected to be discordant for the joint variables $(V, Q)$ because of the poor fitting of the Gumbel logistic copula as can be seen from Figure 3.

After removing the discordant sites, the weighted $L$-moments and corresponding model parameters are computed. These results are presented in Table 3. The estimated parameters are those of the distribution of the homogeneous generated regions.

Homogeneity results are reported in Table 4. The decision about the homogeneity of the remained sites is taken according to the values of the test statistics $H_{||}$. In the bivariate framework, the region is possibly homogeneous since the value of $H_{||}$ is in the range $[1, 2]$. When considering only the volume, the value of the statistic $H_{||}$ is less than 1. Hence, the region is homogeneous. As for the peak, the region is declared to be heterogeneous since the statistic $H_{||}$ has a value greater than 2. Note that the decision concerning the regional homogeneity in relation to the peak variable remains the same whether site 3 is removed or not. Indeed, from Table 3, we observe that removing site 3 has no effect on the values of the weighted $L$-moments and has no significant effect on the values of Kappa parameter estimates.

The advantage of choosing a Kappa distribution is that it allows to include several distributions used in hydrology. This in turn allows to fit the same distribution to all sites avoiding hence the subjective choice of a different distribution for each site. However, a disadvantage of this choice is the high number of parameters to be estimated, especially in the
multivariate context. Consequently, the estimation of these parameters increases the uncertainty of the model. In order to reduce this uncertainty, based on the parsimony principle, one looks for a model with the smallest number of parameters.

6. Conclusions and recommendations

The multivariate discordancy and homogeneity tests based on $L$-moments are applied to a set of sites from the Côte-Nord region in the eastern part of the province of Quebec, Canada. These tests are proposed by Chebana and Ouarda (2007) where a simulation study was carried out to evaluate their performance. The main conclusion highlights the importance of considering jointly and simultaneously all variables characterizing the extreme event and hence identifying a single homogeneous region. In the present paper, practical aspects of these multivariate tests are investigated jointly on flood peak and flood volume. Some of such aspects include the selection of the bivariate distribution using goodness-of-fit tests for the each marginal distribution as well as for the copula, the estimation of the corresponding parameters and a description of the discordant sites.

After removing the discordant sites, the remaining ones represent a homogeneous region for the volume, heterogeneous region for the peak and possibly homogeneous region if both variables are simultaneously considered. The results concerning the volume present a certain level of concordance with the bivariate results with respect to the values of the statistics $H_{II}$. Physically, this can be explained by the fact that the volume contains more information than the peak concerning the whole hydrograph. This shows the interest of the bivariate test to take into account the dependence structure between the variables, and to take advantage of more information from the hydrograph. The implementation and the use of the multivariate tests are simple similarly to the univariate ones. The corresponding Matlab® programs are available from the authors on request.
Acknowledgment

The authors wish to thank Hydro-Québec and the Natural Sciences and Engineering Research Council (NSERC) of Canada for the financial support of this study. The authors wish also to thank the Editor-In-Chief and three anonymous reviewers whose comments helped improve the quality of the paper.

References


List of tables

Table 1. Information concerning the data set.
Table 2. Values of $L$-moments and discordancy statistics for each site in the region.
Table 3. Weighted $L$-moments and the corresponding model parameters.
Table 4. Homogeneity results.

List of figures

Figure 1. Typical flood hydrograph.
Figure 2. Geographical chart of the location of the sites.
Figure 3. Fitting Gumbel logistic copula to each site of the data set.
Figure 4. Distribution of volume (a), peak (b) and joint volume and peak (c) for the data of station 02RF001 (the 7th station)
<table>
<thead>
<tr>
<th></th>
<th>Station number</th>
<th>Station name</th>
<th>Area (km²)</th>
<th></th>
<th>(V,Q) correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02RH049</td>
<td>Petit Saguenay</td>
<td>729</td>
<td>24</td>
<td>0.50</td>
</tr>
<tr>
<td>2</td>
<td>02RH048</td>
<td>Des Ha Ha</td>
<td>564</td>
<td>19</td>
<td>0.73</td>
</tr>
<tr>
<td>3</td>
<td>02RH034/35</td>
<td>Aux Écorces</td>
<td>1120</td>
<td>34</td>
<td>0.50</td>
</tr>
<tr>
<td>4</td>
<td>02RH027</td>
<td>Pikauba</td>
<td>489</td>
<td>34</td>
<td>0.34</td>
</tr>
<tr>
<td>5</td>
<td>02RG005</td>
<td>Métabetchouane</td>
<td>2270</td>
<td>30</td>
<td>0.54</td>
</tr>
<tr>
<td>6</td>
<td>02RC011</td>
<td>Petite Péribonka</td>
<td>1090</td>
<td>31</td>
<td>0.62</td>
</tr>
<tr>
<td>7</td>
<td>02RF001</td>
<td>Chamouchouane (Ashuapmushuan)</td>
<td>15300</td>
<td>43</td>
<td>0.70</td>
</tr>
<tr>
<td>8</td>
<td>02RD002</td>
<td>Mistassibi</td>
<td>8690</td>
<td>39</td>
<td>0.52</td>
</tr>
<tr>
<td>9</td>
<td>02RD003</td>
<td>Mistassini</td>
<td>9620</td>
<td>43</td>
<td>0.52</td>
</tr>
<tr>
<td>10</td>
<td>02RD004</td>
<td>Manouane</td>
<td>3720</td>
<td>23</td>
<td>0.39</td>
</tr>
<tr>
<td>11</td>
<td>02RH045</td>
<td>Valin</td>
<td>740</td>
<td>31</td>
<td>0.42</td>
</tr>
<tr>
<td>12</td>
<td>02RH046</td>
<td>Ste-Marguerite</td>
<td>1100</td>
<td>21</td>
<td>0.48</td>
</tr>
<tr>
<td>13</td>
<td>02SC001</td>
<td>DesEscoumins</td>
<td>779</td>
<td>19</td>
<td>0.49</td>
</tr>
<tr>
<td>14</td>
<td>02SC002</td>
<td>Portneuf</td>
<td>2580</td>
<td>20</td>
<td>0.80</td>
</tr>
<tr>
<td>15</td>
<td>02UA003</td>
<td>Godbout</td>
<td>1570</td>
<td>30</td>
<td>0.75</td>
</tr>
<tr>
<td>16</td>
<td>02UC003</td>
<td>Aux-Pékans</td>
<td>3390</td>
<td>16</td>
<td>0.54</td>
</tr>
<tr>
<td>17</td>
<td>02VA001/3</td>
<td>Tonerre</td>
<td>674</td>
<td>40</td>
<td>0.64</td>
</tr>
<tr>
<td>18</td>
<td>02VB004</td>
<td>Magpie</td>
<td>7200</td>
<td>27</td>
<td>0.66</td>
</tr>
<tr>
<td>19</td>
<td>02VC001</td>
<td>Romaine</td>
<td>13000</td>
<td>48</td>
<td>0.68</td>
</tr>
<tr>
<td>20</td>
<td>02WA001</td>
<td>Nabisipi</td>
<td>2060</td>
<td>25</td>
<td>0.78</td>
</tr>
<tr>
<td>21</td>
<td>02WA002</td>
<td>Aguanus</td>
<td>5590</td>
<td>19</td>
<td>0.60</td>
</tr>
<tr>
<td>22</td>
<td>02WB002</td>
<td>Natashquan</td>
<td>15600</td>
<td>39</td>
<td>0.75</td>
</tr>
<tr>
<td>23</td>
<td>02WC001</td>
<td>Etamamiou</td>
<td>2950</td>
<td>19</td>
<td>0.82</td>
</tr>
<tr>
<td>24</td>
<td>02XB001</td>
<td>St Augustin</td>
<td>5750</td>
<td>14</td>
<td>0.73</td>
</tr>
<tr>
<td>25</td>
<td>02XC001</td>
<td>St Paul</td>
<td>6630</td>
<td>25</td>
<td>0.73</td>
</tr>
<tr>
<td>26</td>
<td>02UC002</td>
<td>Moisie</td>
<td>19000</td>
<td>39</td>
<td>0.65</td>
</tr>
</tbody>
</table>
Table 2. Values of $L$-moments and discordancy statistics for each site in the region.

<table>
<thead>
<tr>
<th># site</th>
<th>Volume and Peak</th>
<th>Volume</th>
<th>Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$l_1$</td>
<td>$t_2$</td>
<td>$t_3$</td>
</tr>
<tr>
<td>1</td>
<td>234.97</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td>2</td>
<td><strong>89.65</strong></td>
<td><strong>0.27</strong></td>
<td><strong>-0.12</strong></td>
</tr>
<tr>
<td>3</td>
<td>293.35</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td>4</td>
<td>133.63</td>
<td>0.20</td>
<td>-0.03</td>
</tr>
<tr>
<td>5</td>
<td>573.30</td>
<td>0.19</td>
<td>-0.02</td>
</tr>
<tr>
<td>6</td>
<td>235.76</td>
<td>0.19</td>
<td>0.08</td>
</tr>
<tr>
<td>7</td>
<td>3622.78</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td>8</td>
<td>2459.44</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>9</td>
<td>2310.05</td>
<td>0.17</td>
<td>0.11</td>
</tr>
<tr>
<td>10</td>
<td>1040.77</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>11</td>
<td>275.02</td>
<td>0.16</td>
<td>0.04</td>
</tr>
<tr>
<td>12</td>
<td>428.91</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td>13</td>
<td>244.40</td>
<td>0.22</td>
<td>0.13</td>
</tr>
<tr>
<td>14</td>
<td>904.78</td>
<td>0.22</td>
<td>0.09</td>
</tr>
<tr>
<td>15</td>
<td>511.32</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>16</td>
<td>984.44</td>
<td><strong>0.11</strong></td>
<td><strong>0.01</strong></td>
</tr>
<tr>
<td>17</td>
<td>273.02</td>
<td>0.17</td>
<td>0.13</td>
</tr>
<tr>
<td>18</td>
<td>2427.84</td>
<td>0.20</td>
<td>0.06</td>
</tr>
<tr>
<td>19</td>
<td>4195.56</td>
<td>0.16</td>
<td>-0.09</td>
</tr>
<tr>
<td>20</td>
<td>833.33</td>
<td>0.16</td>
<td>-0.04</td>
</tr>
<tr>
<td>21</td>
<td>2237.82</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>22</td>
<td>5014.35</td>
<td>0.18</td>
<td>0.06</td>
</tr>
<tr>
<td>23</td>
<td>1176.24</td>
<td>0.21</td>
<td>0.14</td>
</tr>
<tr>
<td>24</td>
<td>2398.37</td>
<td>0.17</td>
<td>-0.03</td>
</tr>
<tr>
<td>25</td>
<td>2222.21</td>
<td>0.17</td>
<td>0.08</td>
</tr>
<tr>
<td>26</td>
<td>5534.18</td>
<td>0.16</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Numbers written in bold and italic character indicate the discordant sites; those written in bold character represent the particular values of $L$-moments that possibly caused the discordancy.
Table 3. Weighted $L$-moments and corresponding model parameters

<table>
<thead>
<tr>
<th>Removed discordant sites</th>
<th>$\overline{L}_1$</th>
<th>$\overline{L}_2$</th>
<th>$\overline{L}_3$</th>
<th>$\overline{L}_4$</th>
<th>$h$</th>
<th>$\kappa$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$m$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Univariate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>2 and 16</td>
<td>1.00</td>
<td>0.18</td>
<td>0.06</td>
<td>0.11</td>
<td>0.1053</td>
<td>0.2258</td>
<td>0.3196</td>
<td>0.8569</td>
</tr>
<tr>
<td>Peak</td>
<td>2</td>
<td>1.00</td>
<td>0.17</td>
<td>0.12</td>
<td>0.15</td>
<td>-0.2887</td>
<td>0.0024</td>
<td>0.2150</td>
<td>0.9090</td>
</tr>
<tr>
<td>Peak</td>
<td>2 and 3</td>
<td>1.00</td>
<td>0.17</td>
<td>0.12</td>
<td>0.15</td>
<td>-0.2370</td>
<td>0.0061</td>
<td>0.2197</td>
<td>0.9016</td>
</tr>
<tr>
<td><strong>Bivariate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>2 and 21</td>
<td>1.00</td>
<td>0.18</td>
<td>0.05</td>
<td>0.10</td>
<td>0.1200</td>
<td>0.2411</td>
<td>0.3250</td>
<td>0.8552</td>
</tr>
<tr>
<td>Peak</td>
<td>2 and 21</td>
<td>1.00</td>
<td>0.17</td>
<td>0.12</td>
<td>0.15</td>
<td>-0.3089</td>
<td>-0.0061</td>
<td>0.2138</td>
<td>0.9098</td>
</tr>
</tbody>
</table>

Table 4. Homogeneity results

| Discordant sites         | $V_{||}^{obs}$ | $\mu_{\text{sim}}$ | $\sigma_{\text{sim}}$ | $H_{||}$ | Decision          |
|--------------------------|----------------|---------------------|------------------------|---------|-------------------|
| Bivariate (V, Q)         | 2 and 21       | 0.0296              | 0.0251                 | 0.0029  | 1.5962            | possibly homogeneous |
| Univariate V             | 2 and 16       | 0.0231              | 0.0218                 | 0.0031  | 0.4279            | homogeneous          |
| Univariate Q             | 2 and 3        | 0.0329              | 0.0228                 | 0.0035  | 2.5664            | heterogeneous        |
Figure 1. Typical flood hydrograph

Figure 2. Geographical chart of the location of the sites
Figure 3. Fitting Gumbel logistic copula to each site of the data
Figure 4. Distribution of volume (a), peak (b) and joint volume and peak (c) for the data of station 02RF001 (the 7th station)