Multivariate shift testing for hydrological variables, review, comparison and application

F. Chebana[1]*

M.-A. Ben Aissia[1]

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

[1] Statistical Hydroclimatology Research Group, 490, Rue de la Couronne, Quebec, Qc, Canada, G1K 9A9.

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

Masdar Institute of Science and Technology,

PO Box 54224, Abu Dhabi, UAE.

* Corresponding author: fateh.chebana@ete.inrs.ca

February 2017

Abstract

Hydrological frequency analysis (HFA) is commonly used for the assessment of the risk associated to hydrological events. HFA is generally based on the assumptions of homogeneity, independence and stationarity of the hydrological data. Hydrological events are often described through a number of dependent characteristics, such as peak, volume and duration for floods. Unfortunately, in this multivariate setting, the verification of the above assumptions is often neglected. When a shift occurs in a data series, it can affect the stationarity and the homogeneity of the data. The objective of this paper is to study tests for shift detection in multivariate hydrological data. The considered shift tests are mainly based on the notion of depth function, except for one test that is considered for comparison purposes. A simulation study is performed to evaluate and compare the power of all these tests with hydrological constraints. A flood analysis application is also carried out to show the practical aspects of the considered tests. The power of the considered tests is influenced by a number of factors, including the sample size, the shift amplitude, the magnitude of the series and the location of the shift in the series.

Keywords: shift, hypothesis testing, multivariate, stationarity, homogeneity, flood, depth.

1. Introduction

In general, in order to perform the statistical analysis of hydrological data a number of fundamental assumptions are required. More precisely, preliminary testing for stationarity, homogeneity and independence is a necessary step in any hydrologic frequency analysis (HFA) study [e.g. *Rao and Hamed*, 2000]. One or more of these assumptions can fail because of a number of reasons. For instance, the assumption of stationarity may not be verified because of a regime shift that can be due to an abrupt change in the watershed characteristics caused by natural or anthropogenic actions on the physical environment, such as deforestation or the construction of a hydraulic structure [e.g. *Bobée and Ashkar*, 1991; *Burn and Hag Elnur, 2002*, *Ouarda and El-Adlouni*, *2011*]. Because of the growing evidence concerning climate change, the common assumption of stationarity of hydrologic phenomena may no longer hold. The presence of shifts in data series is highlighted in several hydrometeorological studies, such as floods [*Seidou and Ouarda*, 2007], precipitation [*Beaulieu et al.*, 2008, 2010; Ouarda et al., 2014; *Chen et al.*, 2016], low-flows [Ehsanzadeh et al., 2011], wind speed [Naizghi and Ouarda, 2016], and temperature data [*Jandhyala et al.*, 2014].

The analysis of multivariate events is of particular interest in several applied fields, including hydrology. Indeed, complex hydrological events, such as floods, droughts and storms are multivariate events characterized by a number of correlated variables. For instance, volume (V), peak (Q) and duration (D) describe floods [*Ouarda et al.*, 2000; *Shiau*, 2003; *Yue et al.*, 1999]. The use of univariate HFA can lead to inaccurate estimation of the risk associated to a given event. Recently, several studies adopted the multivariate framework to treat extreme hydrological events, see e.g. [*Chebana*, 2013] for a summary and recent references.

HFA is composed of four main steps: i) descriptive and explanatory analysis, ii) verification of the basic assumptions including stationarity, homogeneity and independence, iii) modeling and estimation, and iv) risk evaluation and analysis. In the univariate setting, these steps are extensively treated [e.g. *Rao and Hamed*, 2000]. In the multivariate context, the first two steps (i and ii) attracted considerably less attention than the two others. For an overview of step i) in the multivariate framework, the reader is referred to Chebana and Ouarda [2011]. Checking the basic assumptions (step ii) is generally ignored in the hydrological literature in the multivariate setting. For instance, it is not treated in Kao and Govindaraju [2007], Song and Singh [2009] and Vandenberghe et al. [2010]. This step has a significant impact on steps iii) and iv). Therefore, ignoring step ii) may lead to inaccurate models and hence to wrong results and inappropriate decisions regarding resource management and infrastructure design. In order to avoid the loss of human lives and property associated with design event underestimation, or the increase in construction cost associated with overestimation, it is necessary to treat step ii) for a sound and complete multivariate HFA.

Non-stationarity is a very wide notion and includes in particular the presence of one or several shifts in the data. Recently, *Chebana et al.* [2013] provided a review and application of multivariate nonparametric tests for monotonic trends and presented approaches that can be considered as a preliminary step in a complete multivariate HFA. *Chebana et al.* [2013] indicated that, for multivariate hydrological data, various types of non-stationarities can be found for which appropriate tests should by reviewed, compared and applied.

The available literature on shift detection in the hydrological context is focused on the univariate setting. Nevertheless, statistical literature exists for the general multivariate setting. Hence, existing comparisons and evaluations of the proposed tests are based on scenarios and hypotheses that are not adapted to the hydrological context (e.g. sample size, scale, and distributions). In addition, these comparative studies are not exhaustive and are often not based on quantifiable performance criteria.

Consequently, there is a need for comparative studies that consider all available tests and are representative of hydrological reality, scale and constraints.

Several multivariate shift tests are based on the concept of depth function. The latter is a statistical notion to measure the *depth* (or its opposite, the *outlyingness*) of a given point with respect to a multivariate data cloud or its underlying distribution. Depth functions were developed in the seventies and have been receiving increasing interest [e.g. *Tukey*, 1975; *Liu*, 1990; *Zuo and Serfling*, 2000; *Mizera and Müller,* 2004; *Zuo and Cui,* 2005; *Lin and Chen,* 2006; *Liu and Singh*, 2006; *Chebana and Ouarda* 2011; *Singh and Bárdossy*, 2012; *Lee et al., 2014; Wazneh et al., 2013; 2015*]. Depth functions provide a scale-standardized measure of the position of any data point relative to the center of the distribution due to its affine-invariant property [*Li and Liu*, 2004]. For the location shift, this property allows us to view the depth-based test statistics as scale-standardized measures. Therefore, depth-based tests can be performed without the difficulty of estimating the variance of the null sampling distributions. Instead, the decision rule is derived by obtaining p-values using the idea of permutation.

The objectives of the present paper are: 1) to show the importance of the testing step in a multivariate HFA, in particular shift testing, 2) to review shift tests that are available in the statistical literature and which are applicable to hydrological variables within the multivariate HFA context, and 3) to perform an overall evaluation and comparison of these tests under hydrological constraints (such as short sample size, specific distributions).

This paper is organized as follows. Section 2 introduces the definitions and notations related to the shift concept. The considered tests are described in Section 3. The simulation study to evaluate the performance of these tests is presented in Section 4. Section 5 illustrates an application of the reviewed tests on hydrological data. The conclusions of the study and a number of perspectives are reported in Section 6.

2. Shift concept

A shift can be defined by the date at which at least one feature of a statistical model (e.g., location, scale, intercept and trend) undergoes an abrupt change [*Seidou et al.*, 2007]. A large number of techniques can be found in the literature to identify the date of a potential shift and to check its significance. Most of the methodologies use statistical hypothesis testing to detect shifts in the slope or intercept of linear regression models [*Easterling and Peterson*, 1995; *Vincent*, 1998; *Lund and Reeves*, 2002]. For instance, *Solow* [1987], *Easterling and Peterson* [1995], *Vincent* [1998], *Lund and Reeves* [2002] and *Wang* [2003] used the Fisher test to compare a model with and without a shift. The Student and Wilcoxon tests can also be applied sequentially to detect shifts in data series [*Beaulieu et al.*, 2007, 2008].

Note that not all shift approaches are based on hypothesis testing. For instance, *Wong et al.* [2006] used the grey relational method [*Moore*, 1979; *Deng*, 1989] for single shift detection in stream flow data series. In some rare cases, curve fitting methods were used [e.g. *Sagarin and Micheli*, 2001; *Bowman et al.*, 2006]. Extensive reviews of shift detection and correction methodologies in hydrology and climate sciences can be found in *Peterson et al.* [1998] and *Beaulieu et al.* [2009].

To define a shift, let $\left(x_i\right)_{i=1,\dots,n}$ be a given *d*-variate dataset and $1 < s < n$ be a possible shift. If such *s* exists, the series is divided into two subsamples with sizes *s* and *m* = *n*-*s* such that:

$$\begin{aligned} \left(y_1,\dots,y_s\right) &= \left(x_1,\dots,x_s\right) \\ \left(z_1,\dots,z_m\right) &= \left(x_{s+1},\dots,x_n\right) \end{aligned} \qquad (1)$$

Denote by G_1 and G_2 respectively the cumulative distribution functions of these two subsamples. The two distributions G_1 and G_2 have the same form, except for the location, i.e. $G_1(x) = G_2(x + \boldsymbol{\delta})$

for all $x \in R^d$ where $\boldsymbol{\delta} \in R^d$ is a constant vector. Consequently, when testing the presence of a shift at a position *s* of the series $\left(x_i\right)_{i=1,\ldots,n}$, the null and alternative hypotheses are respectively:

$$H_0: \ \boldsymbol{\delta} = 0 \text{ i.e. there is no location shift} \tag{2}$$

$$H_1: \ \boldsymbol{\delta} \neq 0 \text{ i.e. there are two different subsamples at least in one component of } \boldsymbol{\delta}. \tag{3}$$

3. The considered tests

In the present paper, several tests to detect a shift in the location of multivariate series are considered. Except for the C-test, all the presented tests are based on depth functions. The C-tests is considered for comparison purposes. More details are given below regarding p-value evaluation. Table 1 presents a summary of the tests considered in this study.

3.1. **Depth functions**

The absence of a natural order for multivariate data led to the introduction of depth functions [Tukey, 1975]. They are developed and used in a number of research fields, e.g. in statistics by Mizera and Müller, [2004] and Ghosh and Chaudhuri [2005], in economics and social sciences by Caplin and Nalebuff [1991a; b], in industrial quality control by Liu and Singh [1993] and in water sciences by Chebana and Ouarda [2008]. A detailed description and review of depth functions can be found in Zuo and Serfling [2000]. In the following we present a very brief overview of the main concepts. For a given cumulative distribution function F on $\Re^d$ $(d \geq 1)$, a depth function can be defined. It is any non-negative bounded function which possesses a number of suitable properties, i.e. *Affine invariance*, *Maximality at center*, *Monotonicity relative to the deepest point*, *Vanishing at infinity*.

A number of depth functions have been developed and studied [Zuo and Serfling, 2000]. In the following, we present some of the key ones which are considered in this study:

1. *Tukey (or Halfspace) depth* : for $x \in R^d$ with respect to a probability P on R^d , it is defined as:

$$TD(x;P) = \inf \{P(H): H \text{ a closed halfspace that contains } x\} \qquad (4)$$

Chebana and Ouarda [2011] presented a simple illustration of the computation of this depth function.

2. *Mahalanobis depth:* for a given distribution F on R^d with μ and A any corresponding location and covariance measures, respectively, it is given by:

$$MD(x;F) = \frac{1}{1 + d_A^2(x,\mu)} \qquad (5)$$

where $d_A^2(x,y) = (x-y)' A^{-1}(x-y)$ is the Mahalanobis distance between points $x, y \in R^d$ given a positive definite matrix A.

3. *Simplicial depth:* it is expressed as:

$$SD(x;P) = P\{x \in S[X_1,...,X_{d+1}]\} \qquad (6)$$

where $S[X_1,...,X_{d+1}]$ is the random d-dimensional simplex with vertices $X_1,...,X_{d+1}$ which is a random sample from the distribution P.

By replacing F with a suitable empirical function $\hat{F}_n$, a corresponding sample version of a statistical depth function $D(x; F)$ may be defined and denoted by $D_n(x) = D(x;\hat{F}_n)$. Its asymptotic properties have been studied, for instance, in Liu [1990], Massé [2002; 2004] and Lin and Chen [2006]. The computation of some depth functions is complex, especially for high dimensions, and requires approximations and specific algorithms, see for instance, Miller et al. [2003] and Massé and Plante [2009].

In principle, each depth-based test can be defined using any available depth function. However, some of these tests were originally defined and their properties are studied on the basis of a specific depth function. Even though the problem and the tests can be defined in any dimension, the simulation study is based on the bivariate case. The obtained results and conclusions cannot be directly extended and generalized.

3.2. Description of tests

In this section, the considered multivariate shift detection tests are described as well as the method to evaluate their p-values. Performance comparison of these tests in the literature is also presented.

The C-test (Cramér test)

The Cramér test is a two-sample test proposed by *Baringhaus and Franz* [2004]. It is a generalisation of the univariate test proposed by *Cramér* [1928]. However, it is more appropriate to detect shifts in location. This test is based on the difference of Euclidian distances between the observations of the two different subsamples and the half sum of all Euclidian distances of observations of the same subsample. The corresponding test statistic is given by:

$$C = \frac{sm}{s+m}\left[\frac{1}{sm}\sum_{i=1}^{s}\sum_{j=1}^{m}\left\|y_i - z_i\right\| - \frac{1}{2s^2}\sum_{i,j=1}^{s}\left\|y_i - y_j\right\| - \frac{1}{2m^2}\sum_{i,j=1}^{m}\left\|z_i - z_j\right\|\right] \quad (7)$$

where $\left\|y_i - z_j\right\|$ is the Euclidian distance between the i^{th} observation of the first subsample and the j^{th} observation of the second subsample. Recall that s is the location of the shift (and hence the size of the first subsample) and $m = n\text{-}s$ is the size of the second subsample.

The null hypothesis H_0 is rejected for large values of C. A large value of C means that the distance between the observations of the two subsamples is large and consequently, the two subsamples are different. To calculate the p-value, the bootstrapping method is used.

The M-test (Monitoring the Maximum Depth Points)

According to *Li and Liu* [2004], the deepest point of a distribution is a location parameter. Consequently, if G_1 and G_2 are identical distributions, they would have the same deepest point, that is, the deepest points θ_{G_1} and θ_{G_2} should be the same. In addition, for a given depth function *D*, we have $D_{G_2}(\theta_{G_1}) = D_{G_1}(\theta_{G_2})$. If there is an important change in location, θ_{G_1} and θ_{G_2} would be different and θ_{G_2} would be located far away from the subsample from G_1 for which the depth value $D_{G_1}(\theta_{G_2})$ with respect to G_1, is smaller, and vice-versa. Based on this idea, *Li and Liu* [2004] proposed the statistic:

$$M = \min\left\{D_{G_2}(\theta_{G_1}), D_{G_1}(\theta_{G_2})\right\} \tag{8}$$

Li and Liu [2004] used the simplicial depth function SD (6), but other depth functions can be used. Indeed, *Li and Liu* [2004] suggested the Mahalanobis depth function MD (5) for the elliptical distribution. They specified that the SD and TD depth functions can be used with any distribution. The null hypothesis H_0 is rejected for small values of *M*. To approximate the corresponding p-value, *Li and Liu* [2004] proposed Fisher's permutation test [*Snedecor and Cochran*, 1967].

The T-test (Monitoring Shrinking Cusp Point)

Li and Liu [2004] described a graphical approach called DD-plot (for depth-depth) to compare the location of two subsamples. In the context of the T-test, a DD-plot consists in plotting (D) $\left(D_{G_1}(x), D_{G_2}(x)\right)$ with *x* being from either subsample. When the two subsamples follow exactly the same distribution, the DD-plot is a diagonal line that passes through the origin as illustrated in Figure 1a. However, if there is a location change, the graph has a form of leaf with its tip pointing toward the origin (Figure 1b). The more important the location change is; the closer the tip will be to the origin (Figure 1c). The T-test is based on an approximation of the distance between the tip and the origin of the DD-plot. We define the set of points:

$$\Omega = \left\{ x_i \mid i \in \{1,...,n\}, \text{there is no } x_j : D_{G_1}\left(x_j\right) \geq D_{G_1}\left(x_i\right) \text{ and } D_{G_2}\left(x_j\right) \geq D_{G_2}\left(x_i\right) \right\} \quad (9)$$

Then we find the point $x_{\min}$ of Ω such that:

$$\left| D_{G_1}\left(x_{\min}\right) - D_{G_2}\left(x_{\min}\right) \right| = \min_{x \in \Omega} \left| D_{G_1}\left(x\right) - D_{G_2}\left(x\right) \right| \quad (10)$$

If there are several points $x_{\min}$, we take the mean of the corresponding coordinates. The point identified by (10) is an approximation of the leaf-tip point of the DD-plot. The test statistic is then given by:

$$T = \left(D_{G_1}\left(x_{\min}\right) + D_{G_2}\left(x_{\min}\right) \right) \big/ 2 \quad (11)$$

Even though, the distance of the leaf-tip to the origin is approximately $\sqrt{2}T$, the use of the statistic T is equivalent. Similarly to the M-test, *Li and Liu* [2004] used the SD function (6) for the T-test. However, MD (5) and TD (4) depths can also be used. The p-value is obtained using the Fisher's permutation test.

The W-test (Wilcox test)

The W-test was developed by *Wilcox* [2005]. Similarly to the M-test, the W-test is based on the idea that under the null hypothesis, the medians of the two subsamples must be similar. To define the W-test statistic, first the difference of each component is calculated $d_{ij}^{(u)} = z_i^{(u)} - y_j^{(u)}, u = 1,...,d;\ i = 1,...,s;\ j = 1,...,m$ to constitute the vector $d_{ij} = \left(d_{ij}^{(1)}, \cdots, d_{ij}^{(d)}\right)$. *Wilcox* [2005] defined the test statistic by:

$$W = D_F(\mathbf{0}) \Big/ \max_{i=1,...,s;\, j=1,...,m} D_F(d_{ij}) \quad (12)$$

where F is the distribution of the set of vectors d_{ij} and D is the TD depth function (4). Under the null hypothesis, we have $W = 1$, whereas under the alternative hypothesis, we have $W < 1$. The asymptotic distribution of W is unknown. However, *Wilcox* [2005] proposed some critical values C_α for significance levels $\alpha = 0.01;\ 0.025;\ 0.05;\ 0.10$. The values of C_α are derived empirically from simulations using a least squares regression method, and under the assumption of normality. The null hypothesis is rejected when W is lower than C_α.

The QIA- and QIB-tests (quality index tests)

Liu and Singh [1993] developed a Wilcoxon-type rank test based on data depth. This test can detect a location shift and/or a positive scale shift. The statistic of this test is given by:

$$Q_a = \frac{1}{n}\sum_{i=1}^{m} \#\left\{y \in \{y_1, \cdots, y_s\} : D_G(y) \le D_G(z_i)\right\} \qquad (13)$$

Under the null hypothesis, $Q_a = 0.5$ whereas if there is a shift in location, then $Q_a < 0.5$. *Liu and Singh* [1993] used MD (5). *Zuo and He* [2006] found that under some regularity conditions, the asymptotic distribution of Q_a calculated with MD (5), TD (4) or projection depth is normal $N\left(\mu, \sigma^2\right)$ with mean $\mu = 0.5$ and variance $\sigma^2 = \left(s^{-1} + m^{-1}\right)/12$. In the present study, the asymptotic (QIA-test) and bootstrap (QIB-test) methods are used to evaluate the *p*-values.

The Z-test (Zhang test)

Zhang et al. [2009] developed a new test based on the statistic Q_a (13) where the statistic of the Z-test is given by:

$$Z = \frac{6}{n} s \times m \left(Q_a - 0.5\right)^2 \qquad (14)$$

To define Z, *Zhang et al.* [2009] used MD (5). To find the asymptotic distribution of Z, we define the matrix A:

$$A = \begin{bmatrix} 1-p_1 & \sqrt{p_1 p_2} \\ \sqrt{p_1 p_2} & 1-p_2 \end{bmatrix} \quad (15)$$

where $p_i = \frac{n_i}{n}$, i = 1 or 2 and n_i is the number of observations in the i^{th} subsample. Let r be the rank of A, and the nonzero eigenvalues of A are denoted by $\lambda_1,..,\lambda_r$. Under H_0, Z follows asymptotically a sum of independent chi-square distributions:

$$Z \approx \lambda_1 \chi^2(1) + \lambda_2 \chi^2(1) + \ldots + \lambda_r \chi^2(1) \quad (16)$$

This relation is also valid for the half-space and projection depth functions. The asymptotic method is used to evaluate the corresponding *p*-value.

3.3. The p-value computation

The *p*-value of a given test is a simple criterion commonly used by practitioners to decide for the acceptance or rejection of a target null hypothesis. The p-value is based on the distribution of the statistics of the underlying test. For some of the considered tests in the present study, the asymptotic or the exact distribution of the test statistic is unknown or difficult to obtain. Consequently, approximations of the distribution of test statistics, under the null hypothesis, are required. To this end, resampling methods are used. In the present paper, a permutation method [*Snedecor and Cochran*, 1967] and a bootstrap method are used. They are briefly described below. More details can be found, for instance, in *Good* [2005].

To apply the permutation method, the observations should be exchangeable, i.e. the observations should be independent and identically distributed [see e.g. *Efron and Tibshirani*, 1994]. This method consists in permuting n_p times the sample $(x_i)_{i=1,\ldots,n}$ *without replacement* where n_p is a large number. For each permuted sample, the s first elements constitute the first subsample and the remaining ones constitute the second subsample. The test statistic, generically denoted by S, is calculated for each permutation $\left(S^*_{i\ ,i=1,\ldots,n_p}\right)$. The null hypothesis should be rejected for small values of the statistic. The p-value is the proportion of $\left(S^*_{i\ ,i=1,\ldots,n_p}\right)$ smaller or equal to the value S_{obs} obtained from the original observed sample.

The bootstrap method is similar to the permutation method, except that the sample $\left(x_i\right)_{i=1,\ldots n}$ is resampled *with replacement* and the independence assumption is necessary [see e.g. *Efron and Tibshirani*, 1994].

3.4. **Review of comparative studies**

Some performance comparisons of the above tests are presented in the literature. The M- and T-tests, given respectively in (8) and (11), were compared to the Hotelling [1947] T^2 test by *Li and Liu* [2004]. The Hotelling's T^2 test is the most frequently used parametric test to detect location shift [e.g. *Ye et al.*, 2002]. For normally distributed samples with unit variances, the powers of these three tests were found to be comparable, whereas for samples with Cauchy distribution with the same parameter, the M- and T- tests were shown to be more powerful than the Hotelling's test. Moreover, in this case, the M-test outperformed the T-test. Note that both considered distributions (normal and Cauchy) are symmetric. In order to evaluate the performance of these tests for skewed distributions, Dovoedo and Chakraborti [2015] considered ten distributions belonging to five well-known families of multivariate skewed distributions.

Liu and Singh [2006] compared also the quality index test (13) to Hotelling's test. For normal samples, the performances of the two tests were similar, while for Cauchy and Exponential samples the quality index test outperformed the Hotelling's test. *Baringhaus and Franz* [2004] found that the C-test (7) performs almost as well as Hotelling's test for normal and non-normal samples.

These comparisons and evaluations are not appropriate for hydrological applications, since the considered samples are not representative of the hydrological conditions where sample sizes are generally short, and the variables mainly follow extreme distributions such as the Gumbel and the Generalized Extreme Value (GEV) [e.g. El-Adlouni et al., 2010]. The Normal, Cauchy and t distributions are not commonly used in multivariate HFA. In addition, in the literature, only partial comparisons of the above tests were carried out and no overall comparison has been performed dealing with all of them (to the best knowledge of the authors, the only references performing such comparisons are those given in this section).

4. Simulation study

The objective of this simulation study is to evaluate and compare the performances of all the previously presented tests in the hydrological context, such as in the case of flood series based on flood peak *Q* and volume *V*. We also adopt samples with small sizes such as commonly encountered in hydrology.

4.1. Adaptation to floods

The previously presented tests can be applied to hydrological events such as floods, rain storms and droughts. In this paper, we focus on floods. Floods can be described by their peak *Q*, volume *V* and duration *D*, which can be correlated. Indeed, according for instance to *Yue* [2001] there is generally a strong correlation between *Q* and *V*, between *V* and *D* and a moderate correlation between *Q* and *D*. In the present paper, the above considered tests are used to detect location shifts in *Q* and *V*. These two variables are the most studied in hydrology for both the univariate and the bivariate cases (see e.g. Chebana, 2013).

According to *Sklar* [1959], a bivariate distribution can be composed of marginal distributions and a copula. Some previous studies showed that the *Q* and *V* series can be marginally fitted by a

Gumbel distribution [*Chebana and Ouarda*, 2007; *Shiau*, 2003; *Yue*, 2001; *Yue et al.*, 1999]. The cumulative Gumbel distribution is given by:

$$F(x) = \exp\left\{-\exp\left(-\frac{x-\beta}{\sigma}\right)\right\},\ x \text{ and } \beta \text{ real},\ \sigma > 0 \tag{17}$$

where *x* plays the role of each of the variables *Q* and *V*. The dependence between *Q* and *V* can be represented by the Gumbel logistic model [e.g. *Aissia et al.*, 2012; *Chebana et al.*, 2009; *Shiau*, 2003; *Yue et al.*, 1999], expressed according to the following copula:

$$C_b(u,v) = \exp\left\{-\left[\left(-\log(u)\right)^b + \left(-\log v\right)^b\right]^{1/b}\right\},\ b \ge 1 \text{ and } 0 \le u, v \le 1 \tag{18}$$

Note that $b = 1/\sqrt{1-\rho}$ where ρ is the usual correlation coefficient [see e.g. *Genest and Rivest*, 1993; *Gumbel and Mustafi*, 1967].

The presented tests may be affected by several factors. In the simulation study, we examine the impact of the record length *n* (sample size) as well as the degree of change (shift amplitude) in each component of the multivariate series.

For the simulation study, we generate samples (*Q*, *V*) according to models (17) and (18). We consider the Gumbel distribution as marginal for both *Q* and *V*. The corresponding parameters are denoted by:

- σ_{Q1} and β_{Q1} for respectively the scale and location parameters for *Q* of the first *s* observations (before the shift); and
- σ_{Q2} and β_{Q2} for respectively the scale and location parameters for *Q* after the shift.

We define similarly the parameters of *V* $\left(\sigma_V, \beta_V\right)$ and the parameter *b* of the logistic Gumbel copula.

For the G distribution before the shift, we selected the parameters of the Skootamatta basin in Ontario (Canada) which are also employed for simulation studies by *Chebana and Ouar*da [2007; 2009]. Consequently, σ_{Q1} = 15.85, β_{Q1} = 51.85, σ_{V1} = 300.22 β_{V1} = 1239.8 and b = 1.414. Due to space limitations, the reader is referred to the above references for more details regarding the Skootamatta basin.

We study the effect of the following two factors on the performance of the tests: the record length (*n*: sample size) and the amplitude of shifts in the location parameters β, since the tests are mainly designed to detect shifts in the location. Usually, the dependence parameter appears in the copula whereas the location and scale parameters are present in the marginal distributions [*Hobæk Haff et al.*, 2010]. For location shift, we denote $G_1(x) = G_2(x+\delta)$ where $\delta = (\delta_Q, \delta_V)$ is the vector of the shifts in the location of Q and the location of V respectively. In addition, the dependence level between the two variables Q and V is considered with three dependence levels corresponding to ρ = 0.25 (low), ρ = 0.50 (moderate) and ρ = 0.75 (high) where the associated copula dependence parameter is respectively b = 1.155, 1.414 and 2.0.

Even though the considered tests and the simulations are presented in the bivariate setting, they can also be defined when more than two variables are involved to characterize the phenomenon. In theory, the concepts of these tests can be extended to higher dimensions. However, some technical difficulties could arise. First, the computation of some depth functions (which is the basis of a number of the above tests) is complex and requires approximations and specific algorithms for higher dimensions (e.g. for the simplicial depth). Second, a number of issues that are related to models (especially for copulas) such as uncertainty increase, effectiveness of goodness-of-fit testing, model formula complexity and questionable representativety of some models, need to be addressed. Third, the number of the shift possibilities increases rapidly with the dimension, for instance, with 3 variables we have 8 possibilities where the shift occurs without accounting for the different shift amplitudes (for each variable) as well as the different types of dependence between the variables (3 pairwise and 1 overall). Hence, the simulation results obtained in this paper cannot be generalised directly to higher dimensions, and additional work will be required for this purpose.

4.2. Simulation design

The conducted simulation study consists of two steps. In the first one, we generate a large number *N* of samples to evaluate the effects of different factors on the performance of the tests. Three sample sizes are considered n = 30, 50 and 80 corresponding to s=5, 10; 5, 10, 20 and 5, 10, 20, 30 respectively. For each sample size, several amplitudes of location shift are considered: δ= 10, 20, -20, 40 and 70%. We generate the samples as follows:

I. *No change in all parameters*: All the parameters of the distribution are the same before and after the shift. This allows to obtain samples under the null hypothesis (no shift) and therefore, for each record length n, we calculate the probability of type one error (α);

II. *Change in location parameters:* The distribution before the shift (G_1) is the same as after the shift (G_2), except for the location parameters β in the marginal. We consider 3 cases:

   a. Change only in location of Q: $\delta_Q = 10, 20, 40 \text{ and } 70\%$;

   b. Change only in location of V: $\delta_V = 10, 20, 40 \text{ and } 70\%$;

   c. Change in the location of Q and V simultaneously: $\left(\delta_Q,\ \delta_V\right)$ = (10,10), (20,20), (20,-20), (40,40), and (70,70)%.

For the evaluation of p-values, based on the permutation and the bootstrap methods, we use n_p = 500 permutations or bootstrap samples. This value of n_p is proposed by *Li and Liu* [2004] for the M- and T-tests and is superior to the value 200 proposed by *Baringhaus and Franz* [2004] for the C-test.

In the second step of the simulation study, we evaluate the performance of each test on the basis of the estimate $\hat{\alpha}$ of the type one error α and the power of the considered tests. In the present study, we fix α = 5%. Consequently, we reject H_0 if the p-value is less than 5%. We consider a number of replications N=3000 which higher than the number of replications used by *Li and Liu* [2004], *Wilcox* [2005] and *Zhang et al.* [2009].

Since the peak and the volume have very different scales, we also considered standardizing the generated samples (with the known standard deviation and its empirical estimate of the whole sample before and after the shift). Note that the standard deviation of a Gumbel distribution can be obtained directly from its scale parameter σ as $\pi\sigma/\sqrt{6}$.

4.3. Simulation results

In order to avoid repetition and for notation simplicity, the depth function will only be written in the test index when it is needed. For example, M_{TD}-test is the M-test with TD depth function.

I. Type one error estimation

The estimates $\hat{\alpha}$ of α for the considered tests are presented in Table 2 (with and without standardization). First, we observe that the results are almost the same with and without standardization for all situations and tests. Since the critical level is fixed at α = 5%, a performing test should have $\hat{\alpha}$ as close as possible to 5%. From Table 2, we see that $\hat{\alpha}$ generally approaches 5% when n increases. Values of $\hat{\alpha}$ for the M-test are close to 5% except for M_{TD} and M_{SD} in the case (n,s)=(30,10). The T- and C-tests have $\hat{\alpha}$ around 5% whatever the sample size. The W-test underestimates α while the QIB-, QIA- and Z-tests overestimate it. However, the QIB_{SD}-, QIA_{TD}- and Z_{TD}- tests have $\hat{\alpha}$ higher than 20% when (n,s)=(30,10) which means that they reject H_0 more frequently when it is true.

II. Power evaluation

Table 3 summarises the simulation results for shift detection tests for several shift amplitudes in *Q*, *V* and (*Q*,*V*). In general, these results show good behaviour for the tests in terms of power. The power increases with the shift amplitude δ and with the sample size *n*. In the present paper, a test power is considered high when it exceeds 95%.

For *n*=30, Table 3 (part a) shows that high powers are generally recorded for large shift amplitudes i.e. $(\delta_Q,\ \delta_V)$ = (70,0) or $(\delta_Q,\ \delta_V)$ = (70,70). For the M- and T-tests, best powers are recorded with the MD depth function. The TD depth function gives best powers for the W-, QIA- and Z-tests while for the QIB-test, the best power is reached with the SD depth function. However, as seen before, the QIB_{SD}-, QIA_{TD}- and Z_{TD}-tests are problematic when estimating α. Note that the depth function that provides the best test power is not necessarily the one with which the test was originally defined, e.g. M- and T-tests. For the C-test, the power depends on the variable in which the shift has occurred. Indeed, a shift only in *Q* leads to low power for the C-test, while the opposite is true when the shift is either in *V* or in (*Q*,*V*). This is due to the difference in the first term in (7) which can be affected by the scale of the series. In the case of floods, *Q* and *V* series have very different scales. Consequently, a change in *Q* does not have a great effect on the test statistic while the opposite is true for *V* (and hence for (Q,V)). We can conclude that the C-test is more sensitive to a change in *V* than a change in *Q*. This result was not shown in previous studies since the simulations were based on variables of the same nature and scale. This can be explained by the fact that the statistic C is based on the Euclidian distance which is not affine invariant whereas the depth-based tests are not affected by the scale since depth functions are usually affine invariant [*Zuo and Serfling*, 2000].

For n = 50, from Table 3 (part b), we can see that high powers are obtained starting form $\left(\delta_Q,\ \delta_V\right)$ = (0,40). For each test, the depth functions that lead to the best power when n = 30 are generally the same when n=50. The powers when n=50 are generally higher than the power corresponding to n=30 with a few exceptions: for QIB-, QIA-, and Z-tests with $\left(\delta_Q,\ \delta_V\right)$=(0,10), (10,0), (10,10), (0,20), (20,0) or (20,20).

Table 3 (part c) summarizes the simulation results of the presented tests when n=80. Results show that high powers are observed starting from $\left(\delta_Q,\ \delta_V\right)$= (20,-20) for the M-, T- and W_{TD}-tests. For the M-test, results are similar for the three considered depth functions for each shift amplitude whereas for the other tests, depth functions leading to the highest powers for n=80 are also the same as for n=30 or 50. Generally, the performances of the tests increase when the shifts of V and Q have different signs. For instance, the powers for $\left(\delta_Q,\ \delta_V\right)$= (20,-20) are higher than those corresponding to $\left(\delta_Q,\ \delta_V\right)$= (20, 20) for all tests. Note that the C-test power increases with n except when the shift is located only in Q.

From these results one can conclude that, generally, best results are obtained by the M-, T- and W-tests (with power higher or equal to that of the rest of the tests). For low sample sizes, high powers are observed for large shift amplitudes (70%), while for large sample sizes, high powers are observed starting from $\left(\delta_Q,\ \delta_V\right)$=(20,-20)%. For low shift amplitudes (10%), low powers are recorded for all the considered tests. Figure 2 illustrates the applicability (where power is reasonable or high) of considered tests for the combinations of the studied sample sizes and shift amplitudes.

As shown in Table 3, the powers of the tests, in particular the C-test, are affected by the different scales in the variables V and Q. Table 4 presents results corresponding to the case when the generated series are standardized using the corresponding estimated standard deviation. We observe that the standardized C-test provides better results especially when the change is symmetrical in V or in Q, such as the case $\left(\delta_Q,\ \delta_V\right)$ = (0,20) or (20,0)%. However, it is still affected in the sense that the power is not the same when the variables are affected symmetrically. The other tests remain almost the same after standardization even though the power is reduced for some tests (e.g. QIB_SD, QIA, *n*=50).

In Table 5, we consider standardizing with the estimated or known standard deviation. We observe from Table 5 that the power is close to being symmetric regarding the change in V or Q when the standard deviation is estimated, and the power becomes almost symmetric when the standard deviation is known. The improvement is increasing with the sample size where, for instance, the power is almost identical when a change affects either V or Q with the same shift magnitude. Note that by construction, the depth-based tests should not be affected by the scale since the depth functions are affine-invariant (see Li and Liu, 2004).

Table 6 presents evaluations of the power of the previous tests (with standardized samples) with different possibilities of the location of the shift through different values of s. We observe that for a given *n*, the power generally increases with *s*, with some exceptions such as for QIA and QIB for which the power decreases with *s*. We observe also that small values of *s* (mainly s = 5 in the present study) affect the depth computations of some tests like the M and QIB tests which presented unexpected behaviors (always 0% for M or 100% for QIB).

Variations of the type one error (α) estimations and the power with respect to the dependence level are presented in Table 7. Regarding α estimation, for a given test, the estimation is practically unaffected for all three dependence levels. Regarding the power, in general for all depth-based tests, the power is increasing with some exceptions related to the values of δ_Q and δ_V, such as (0,10) and (10,10). The C-test seems to be almost unaffected by the dependence level.

During the simulation, a problem related to the set Ω occurred with the T-test. Indeed, the set Ω given in (9) can be empty. It was observed that Ω is rarely empty in general with the SD and MD depths, but it is often empty with the TD depth. This issue was not mentioned or considered in *Li and Liu* [2004]. These cases are excluded from the present computations.

From the present simulation study, the following general observations can be made (also illustrated in Figure 2):

- The C-test is more sensitive to a change in V than a change in Q;
- For a small sample size (n=30), high power is observed only for high shift amplitudes;
- For a large sample size (n=80), best powers are observed for the M-, T- and W-tests;
- The QIB-, QIA- and Z-tests can be problematic especially for low shift amplitudes;
- For type one error estimation, QIB_{SD}-, the QIA- and Z_{TD}-tests are problematic, especially when n=30. Good performances are observed for the M-, T-, W- and C-tests with all depth functions;
- For low shift amplitudes $\left(\delta_Q,\ \delta_V\right)$ = (0, 10), (10, 0) or (10, 10), powers are low. This means that a 10% change in one or both location parameters is not detected by the considered tests;
- The C-test is severely affected by the scale and samples should be standardized to reduce this effect. However, the depth-based tests are less affected by the variable scale;
- Generally, the power increases with the location shift s. However, some tests provided inconsistent results when s is very close to the beginning (or the end) of the series;
- Generally the power of the depth-based tests increases with the dependence level whereas the C-test is almost unaffected by this factor.

5. Application

In this section, the previously considered tests are applied to the data series of three stations (Moisie, Magpie and Romaine) with natural flow regimes. Moisie and Romaine are among a number of stations selected in Canada to be part of the Reference Hydrometric Basin Network (RHBN) used for the study of the impacts of climate change on hydrologic regimes in the country [*Ouarda et al.* 1999]. The three considered stations are located in the Cote Nord Region of the province of Quebec, Canada. The *Moisie* station (reference number 072301) is located on Moisie River at 1.5 km upstream of the Québec North Shore Labrador Railway (QNSLR) bridge with a drainage basin area of 19 012 km^2. Data series are available from 1968 to 1998. The *Magpie* station (reference number 073503) is located at the outlet of Magpie Lake. Its drainage basin has an area of 7 201 km^2 and observations are available from 1979 to 2004. The *Romaine* station (reference number 073801) is located at 16.4 km from the Chemin-de-fer bridge on Romaine River, with a drainage basin area of 12 922 km^2 and available data from 1961 to 2006. Figure 3 and Table 7 present respectively the geographical location and general information about the considered stations.

Spring flood characteristics Q and V are extracted from daily streamflow series for each station. The peak Q is defined as the maximum annual of daily streamflow series whereas the volume V is the cumulative streamflow over the flood event, see e.g. Aissia et al. [2012] for formal definitions of flood variables. Note that the variables Q and V correspond to the same flood event each year. In particular, they correspond to the annual spring flood event which is generally the important flood event in the year and is caused mainly by snow melting [Aissia et al., 2012].

Figure 4 shows the time series of Q and V for the three stations. Since these stations are geographically close to each other (Figure 3), it is expected that any eventual shift would be observed in all three stations. From Figure 4 we can see that a shift can be located in *Q* and *V* around 1984 for all three stations. Therefore, the previously presented tests (with and without standardizing the samples) are applied for each station in 1984. Statistics and p-values of the considered tests are summarized in Table 8. Note that, instead of the p-value, for the W-test the conclusion is presented as: 1 if there is a shift, 0 if not, since this test is based on critical thresholds [*Wilcox*, 2005].

First, we observe that the standardization does not affect the values of the test statistics of the depth-based tests whereas the C-test statistics are completely different. However, the p-values are almost the same and the standardization generally does not change the conclusions. Results show that all considered tests are in agreement with the existence of a shift in the *Moisie* station data. For instance, the p-values of the T-, QIB-, QIA-, Z- and C-tests are less than 1%. For *Magpie* station, the M-test is the only test which does not detect the presence of a shift for all depth functions whereas the T-test indicates a shift with all depth functions. This can be explained by the fact that for small sample sizes (Table 3a) the power of the M-test is lower than the power of the T-test. Considering *Romaine* station, only the T_{SD}-, QIB_{TD}-, QIB_{MD}- and Z_{TD}-tests cannot confirm the existence of a shift in the year 1984.

From the results of the three stations, one can conclude that, the year 1984 is detected as a shift for the *Moisie* station by all tests (and depth functions) and for *Romaine* station by all tests (not all depth functions). However, for the *Magpie* station, 3 out of 6 tests detect the shift. Indeed, from Figure 4b one can see that a shift in 1984 is not very clear in *Magpie* station and the short sample data before the shift can have an impact on the power of considered tests. Since these stations are geographically close (Figure 3), one can say that 1984 represents probably a shift for all these stations.

6. Conclusions

The aim of this paper is to study shift detection in the multivariate hydrological setting by comparing the power of several tests and by adapting these tests for hydrological practice. Shift detection is required to insure the validity of HFA assumptions (homogeneity and stationarity) and has hence a strong impact on the selection of the appropriate multivariate distribution. All considered tests are based on data depth, except for the C-test, which is considered for comparison purposes. An overall simulation study that considers all the considered tests and which takes into account the hydrological context, is performed to evaluate and compare the power of the considered tests to detect shifts in the location parameter of Q, V and (Q,V). These tests are also applied to a real-world flood case study consisting of three stations from the province of Québec, Canada.

In general, the powers of these tests increase with the shift amplitude and with the sample size. However, the QIA-, QIB- and Z-tests may be problematic for small sample sizes and they overestimate the type one error α. The scale of the tested variables has an effect on the performance of the considered tests. Especially, the C-test is severely affected and requires a standardizing of the samples. In general, the tests are more powerful when the shift occurs far from the end or the beginning of the series. For low shift amplitudes, the considered tests do not perform well for all sample sizes. On the basis of the above comparison, and considering the nature of hydrological data, it can be recommended to use the M-, T- and W-tests. More precisely, for small sample sizes, the MD depth function is preferred for the M- and T-tests while the TD depth function is preferred for the W-test whereas TD and SD are not recommended when testing a shift far from the middle section of the series.

The application of the considered tests to observed hydrological data shows their ability to detect multivariate shifts. It is also observed that the performance of the tests is affected by the length of the sub-series before or after the shift. The current literature review and hydrologic simulations and application focused on the bivariate cases. It is recommended to examine the performance of these tests for higher dimensions in future research efforts.

ACKNOWLEDGEMENTS

The authors are grateful to the Editor, the Associate Editor and the reviewers for their comments and suggestions which helped improve the quality of the paper. The authors thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for the financial support and Marjolaine Dubé for her assistance.

References

Aissia, M. A. B., 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.

Baringhaus, L., and C. Franz (2004), On a new multivariate two-sample test, *Journal of Multivariate Analysis*, *88*(1), 190-206.

Beaulieu, C., T. B. M. J. Ouarda, and O. Seidou (2007), A review of homogenization techniques for climate data and their applicability to precipitation series, *Hydrological Sciences Journal*, *52*(1), 18-37.

Beaulieu, C., O. Seidou, T. B. M. J. Ouarda, X. Zhang, G. Boulet, and A. Yagouti (2008), Intercomparison of homogenization techniques for precipitation data, *Water Resources Research*, *44*(2), W02425.

Beaulieu, C., Seidou, O., Ouarda, T.B.M.J., and X. Zhang (2009). Intercomparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change point models, *Water Resources Research*, 45, W08410, doi:10.1029/2008WR007501.

Beaulieu, C., Ouarda, T.B.M.J., and O. Seidou. (2010). A Bayesian Normal Homogeneity Test for the detection of artificial discontinuities in climatic series, *International Journal of Climatology*, DOI: 10.1002/joc.2056.

Bobée, B., and F. Ashkar (1991), The gamma family and derived distributions applied in hydrology, *Water Resources Publication. Littleton, Colorado, USA*.

Bowman, A. W., A. Pope, and B. Ismail (2006), Detecting discontinuities in nonparametric regression curves and surfaces, *Statistics and Computing*, *16*(4), 377-390.

Burn, D. H., and M. A. Hag Elnur (2002), Detection of hydrologic trends and variability, *Journal of Hydrology*, 255, 107–122.

Chebana, F. (2013), Multivariate Analysis of Hydrological Variables, in *Encyclopedia of Environmetrics*, edited, John Wiley & Sons, Ltd.

Chebana, F., and T. B. M. J. Ouarda (2007), Multivariate L-moment homogeneity test, *Water Resources Research*, *43*(8).

Chebana, F., and T. B. M. J. Ouarda (2008), Depth and homogeneity in regional flood frequency analysis, *Water Resources Research*, *44*(11).

Chebana, F., and T. B. M. J. Ouarda (2009), Index flood–based multivariate regional frequency analysis, *Water Resources Research*, *45*(10), W10435.

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, and T. C. Duong (2013), Testing for multivariate trends in hydrologic frequency analysis, *Journal of Hydrology*, *486*(0), 519-530.

Chebana, F., T. B. M. J. Ouarda, P. Bruneau, M. Barbet, S. El 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.

Chen, S., Li, Y., Kim, J. and Kim, S. W. (2016), Bayesian change point analysis for extreme daily precipitation. Int. J. Climatol.. doi:10.1002/joc.4904

Cramér, H. (1928), On the composition of elementary errorsé: II, Statistical applications, *Skandinavisk Aktuarietidskrift*, *11*, 141-180.

Deng, J. L. (1989), Introduction to Grey System Theory, *J. Grey Syst.*, *1*(1), 1-24.

Dovoedo, Y.H. and Chakraborti, S. (2015), Power of depth-based nonparametric tests for multivariate locations, *Journal of Statistical Computation and Simulation*, 85:10, 1987-2006

Easterling, D. R., and T. C. Peterson (1995), A new method for detecting undocumented discontinuities in climatological time series, *International Journal of Climatology*, *15*(4), 369-377.

Efron, B., and R. J. Tibshirani (1994), *An Introduction to the Bootstrap*, Taylor & Francis.

El Adlouni, S., Chebana, F., and Bobée, B. (2010), Generalized Extreme Value versus Halphen System: Exploratory Study. *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000152, 79-89.

Ehsanzadeh, E., Ouarda, T. B. M. J. and Saley, H. M. (2011), A simultaneous analysis of gradual and abrupt changes in Canadian low streamflows. Hydrol. Process., 25: 727–739. doi:10.1002/hyp.7861

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

Good, P. (2005), *Permutation, Parametric and Bootstrap Tests of Hypotheses*, 315 pp., Springer New York.

Gumbel, E. J., and C. K. Mustafi (1967), Some Analytical Properties of Bivariate Extremal Distributions, *Journal of the American Statistical Association*, *62*(318), 569-588.

Hobæk Haff, I., K. Aas, and A. Frigessi (2010), On the simplified pair-copula construction — Simply useful or too simplistic?, *Journal of Multivariate Analysis*, *101*(5), 1296-1310.

Hotelling, H. (1947), Multivariate quality control: Illustrated by the air testing of sample bomb sight., in *In Selected Techniques of Statistical Analysis for Scientific and Industrial Research and Production and Management Engineering*, edited by McGraw-Hil, pp. 111-184, New York.

Jandhyala, V., P. Liu, S. Fotopoulos, and I. MacNeill, (2014) Change-Point Analysis of Polar Zone Radiosonde Temperature Data. J. Appl. Meteor. Climatol., 53, 694–714,doi: 10.1175/JAMC-D-13-084.1.

Kao, S.-C., and R. S. Govindaraju (2007), A bivariate frequency analysis of extreme rainfall with implications for design, *J. Geophys. Res.*, *112*(D13), D13119.

Lee, T.-S., Ouarda, T.B.M.J., Chebana, F., and D. Park (2014), Evaluation of a Depth-Based Multivariate -Nearest Neighbor Resampling Method with Stormwater Quality Data, *Mathematical Problems in Engineering*, 2014 (404198), doi:10.1155/2014/404198.

Li, J., and R. Y. Liu (2004), New Nonparametric Tests of Multivariate Locations and Scales Using Data Depth, *Statistical Science*, *19*(4), 686-696.

Lin, L., and M. Chen (2006), Robust estimating equation based on statistical depth, *Statistical Papers*, *47*(2), 263-278.

Liu, R. Y. (1990), On a Notion of Data Depth Based on Random Simplices, *The Annals of Statistics*, *18*(1), 405-414.

Liu, R. Y., and K. Singh (1993), A Quality Index Based on Data Depth and Multivariate Rank Tests, *Journal of the American Statistical Association*, *88*(421), 252-260.

Liu, R. Y., and K. Singh (2006), Rank tests for multivariate scale difference based on data depth, in *Data Depth: robust multivariate analysis, computational geometry, and applications*, edited by R. Y. S. Liu, K. and Souvaine, D.L., pp. 17-35, Àmerican Mathematical Society.

Lund, R., and J. Reeves (2002), Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model, *Journal of Climate*, *15*(17), 2547-2554.

Mizera, I., and C. H. Müller (2004), Location-scale depth, *Journal of the American Statistical Association*, *99*(468), 949-966.

Moore, R. E. (1979), *Methods and Applications of Interval Analysis*, SIAM, Philadelphia, USA.

Naizghi, M. S., and T. B. M. J. Ouarda (2016). Teleconnections and analysis of long-term wind speed variability in the UAE, International Journal of Climatology, DOI: 10.1002/joc.4700.

Ouarda, T. B. M. J. and S. El-Adlouni (2011). Bayesian nonstationary frequency analysis of hydrological variables. *Journal of the American Water Resources Association* (JAWRA), 1-10, 47(3): 496-505. DOI: 10.1111/j.1752-1688.

Ouarda, T., M. Haché, P. Bruneau, and B. Bobée (2000), Regional Flood Peak and Volume Estimation in Northern Canadian Basin, *Journal of Cold Regions Engineering*, *14*(4), 176-191.

Ouarda, T.B.M.J., Rasmussen, P.F., Cantin, J.F., Bobée, B., Laurence, R., Hoang, V.D. and G. Barbé (1999). Identification of a hydrometric data network for the study of climate change over the province of Quebec. Revue des Sciences de l'eau, 12(2): 425-448.

Ouarda, T.B.M.J., Charron, C., Niranjan Kumar, K., Marpu, P., Ghedira, H., Molini, A.L., Khayal (2014). Evolution of rainfall regime in the UAE, Journal of Hydrology, 514 (June): 258–270, DOI:10.1016/j.jhydrol.2014.04.032.

Peterson, T. C., et al. (1998), Homogeneity adjustments of in situ atmospheric climate data: A review, *International Journal of Climatology*, *18*(13), 1493-1517.

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

Sagarin, R., and F. Micheli (2001), Climate change in nontraditional data sets, *Science*, *294*(5543), 811.

Seidou, O., and T.B.M.J., Ouarda (2007). Recursion-based multiple changepoint detection in multivariate linear regression and application to river streamflows, *Water Resources Research*. 43, W07404, doi:10.1029/2006WR005021, 1-18.

Seidou, O., J. J. Asselin, and T. B. M. J. Ouarda (2007), Bayesian multivariate linear regression with application to change point models in hydrometeorological variables, *Water Resources Research*, *43*(8),.

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

Singh, S. K., and A. Bárdossy (2012), Calibration of hydrological models on hydrologically unusual events, *Advances in Water Resources*, *38*(0), 81-91.

Sklar, A. (1959), *Fonctions de répartition à n dimensions et leurs marges*.

Snedecor, G. W., and W. G. Cochran (1967), *Statistical Methods*, Iowa State University Press.

Solow, A. R. (1987), Testing for Climate Change: An Application of the Two-Phase Regression Model, *Journal of Climate and Applied Meteorology*, *26*(10), 1401-1405.

Song, S., and V. P. Singh (2009), Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data, *Stochastic Environmental Research and Risk Assessment*, 1-20.

Tukey, J. W. (1975), Mathematics and the picturing of data, paper presented at International Congress of Mathematicians, Canad. Math. Congress,, Vancouver, B. C., 1974.

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 Resour. Res.*, *46*(1), W01512.

Vincent, L. A. (1998), A Technique for the Identification of Inhomogeneities in Canadian Temperature Series, *Journal of Climate*, *11*(5), 1094-1104.

Wang, X. L. (2003), Comments on "Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model", *Journal of Climate*, *16*(20), 3383-3385.

Wazneh, H., Chebana, F., Ouarda, T.B.M.J. (2015), Delineation of homogeneous regions for regional frequency analysis using statistical depth function, *Journal of Hydrology*, Vol. 521: 232-244.

Wazneh H., Chebana F. and Ouarda T.B.M.J. (2013), Optimal depth-based regional frequency analysis. *Hydrol. Earth Syst. Sci.*, 17 : 2281-2296. DOI : 10.5194/hess-17-2281-2013

Wilcox, R. R. (2005), Depth and a multivariate generalization of the Wilcoxon-Mann-Whitney test, *American Journal of Mathematical and Management Sciences*, *25*(3-4), 343-363.

Wong, H., B. Q. Hu, W. C. Ip, and J. Xia (2006), Change-point analysis of hydrological time series using grey relational method, *Journal of Hydrology*, *324*(1–4), 323-338.

Ye, N., S. M. Emran, Q. Chen, and S. Vilbert (2002), Multivariate statistical analysis of audit trails for host-based intrusion detection, *IEEE Transactions on Computers*, *51*(7), 810-820.

Yue, S. (2001), A bivariate gamma distribution for use in multivariate flood frequency analysis, *Hydrological Processes*, *15*(6), 1033-1045.

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, C., Z. Lin, and J. Wu (2009), Nonparametric tests for the general multivariate multi-sample problem, *Journal of Nonparametric Statistics*, *21*(7), 877-888.

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

Zuo, Y., and H. Cui (2005), Depth weighted scatter estimators, *Ann. Statistics*, *33*(1), 381-413.

Zuo, Y., and X. He (2006), On the limiting distributions of multivariate depth-based rank sum statistics and related tests, *Annals of Statistics*, *34*(6), 2879-2896.

Tables

Table 1: Summary of the presented tests

	Reference	Designed to detect	p-value evaluation	Used depth functions	Comparison from the literature: For normal samples	Comparison from the literature: For non-normal samples
C-test Eq. (7)	Baringhaus and Franz (2004)	Location and/or scale shift	Bootstrap	NA	The C-test performs almost as well as Hotelling test	
M-test Eq. (8)	Li and Liu (2004)	Location shift	Permutation	- Simplicial* - Mahalanobis - Half-space	The powers of M-test, T-test and Hotteling tests are comparable	The M-test outperformed the T-test and both are more powerful than the Hotelling test
T- test Eq. (11)	Li and Liu (2004)	Location shift	Permutation	- Simplicial* - Mahalanobis		
W-test Eq. (12)	Wilcox (2005)	Location shift	Critical thresholds given in Wilcox[2005]	- Half-space - Simplicial* - Mahalanobis	NA	
Q-test Eq. (13)	Liu and Singh (1993)	Location and/or positive scale shift	Bootstrap or asymptotic	- If p-value found asymptotically: Mahalanobis* or Half-space - If bootstrap is the p-value evaluation: Half-space or Simplicial	The performances of the Q- and Hotelling tests are similar	The Q-test outperformed the Hotelling one
Z-test Eq. (14)	Zhang et al. (2009)	Multiple location and/or scale shift	Asymptotic	- Half-space - Mahalanobis	NA	

*with which the test was originally developed

Table 2 : Values of $\hat{\alpha}$ (estimate of α) for the considered tests and for each sample size.

n	*s*	M			T			W			QIB			QIA		Z		C
		TD	MD	SD*	TD*	MD	SD	TD*	MD	SD	TD	MD*	SD	TD	MD*	TD	MD*	
30	10	1.3	5.2	0.1	3.9	5.5	5.0	2.7	0.2	1.5	10.1	6.7	86.5	46.2	19.3	22.0	7.6	5.1
50	20	3.8	5.8	5.4	4.6	5.5	5.3	3.9	0.1	0.4	8.4	6.6	48.0	29.9	12.4	12.1	5.8	5.4
80	30	4.1	5.1	5.0	5.0	5.2	5.1	2.6	0.0	0.2	6.8	6.0	27.1	22.8	9.9	8.2	4.6	6.0
Standardized versions																		
30	10	0.5	4.9	0.1	4.1	5.9	5.9	2.9	0.3	1.5	10.5	6.8	86.7	46.5	19.3	21.6	7.6	5.3
50	20	2.1	5.4	4.5	3.7	5.5	4.9	2.8	0.0	0.4	8.3	6.9	47.3	29.2	12.3	11.7	5.4	4.9
80	30	4.2	4.9	4.7	4.9	5.4	5.4	2.7	0.0	0.2	6.6	5.3	27.9	23.0	9.9	8.1	4.3	4.8

with *n*: sample size, *s*: shift, *: the depth function with which the test is originally defined. Gray color indicates that $\hat{\alpha}$ is close to 5% (between 3% and 7%).

Table 3 : Power comparison for the considered tests to detect shifts in *Q*, *V* or (*Q*,*V*).

δ_Q	δ_V	M			T			W			QIB			QIA		Z		C
		TD	MD	SD*	TD	MD	SD*	TD*	MD	SD	TD	MD*	SD	TD	MD*	TD	MD*	
a)									$(n,s) = (30,10)$									
0	10	2.5	**10.5**	0.1	7.8	**11.6**	7.8	**7.0**	0.6	4.0	10.0	8.0	**84.5**	**42.4**	19.9	**27.1**	9.6	**13.9**
10	0	0.8	**8.4**	0.1	5.6	**9.0**	7.7	**5.1**	0.4	2.7	9.5	6.6	**85.4**	**43.1**	19.8	**24.1**	7.7	4.0
10	10	2.0	**10.6**	0.2	7.4	**11.8**	8.7	**7.2**	0.5	4.1	8.4	7.0	**83.3**	**39.3**	19.6	**26.8**	9.3	**14.2**
0	20	3.0	**27.1**	0.3	23.6	**32.6**	23.7	**27.1**	5.3	18.6	16.2	14.9	**89.6**	**53.0**	32.4	**48.7**	20.2	**40.5**
20	0	2.1	**17.8**	0.2	15.5	**22.2**	14.8	**16.6**	2.1	10.4	13.2	10.1	**87.7**	**48.9**	25.6	**37.5**	13.6	5.1
20	20	5.7	**26.3**	0.3	21.6	**29.7**	20.1	**25.1**	4.9	17.0	11.4	11.8	**83.7**	**41.6**	24.3	**47.9**	19.5	**38.9**
20	-20	13.1	**54.6**	0.4	49.4	**65.0**	41.5	**60.1**	21.6	47.9	45.6	38.2	**95.9**	**84.1**	61.6	**75.7**	46.6	**40.6**
0	40	17.5	**77.3**	0.9	71.1	**86.5**	65.1	**84.6**	51.3	76.7	54.1	53.6	**97.0**	**82.8**	72.7	**91.9**	74.9	**91.7**
40	0	14.1	**60.5**	0.8	53.1	**67.5**	46.2	**65.0**	26.9	54.0	36.8	35.4	**94.0**	**72.1**	55.2	**80.3**	51.6	5.5
40	40	17.1	**74.3**	0.7	65.7	**80.6**	63.8	**80.6**	43.2	71.1	39.8	43.3	**94.7**	**69.8**	60.2	**91.8**	72.0	**92.3**
70	0	22.6	**96.6**	1.0	87.4	**98.4**	84.2	**98.6**	86.2	96.6	81.0	83.4	**99.7**	**95.4**	92.7	**99.4**	96.1	6.3
70	70	23.5	**98.8**	1.4	86.9	**99.2**	90.8	**99.2**	93.5	97.5	83.7	88.6	**99.9**	94.7	**94.8**	**99.9**	99.4	**99.9**
b)									$(n,s) = (50,20)$									
0	10	11.2	**17.4**	15.8	15.5	**19.0**	15.3	**16.2**	1.6	4.5	9.2	8.9	**46.3**	**29.0**	15.1	**20.4**	7.5	**21.3**
10	0	7.7	**12.7**	11.3	10.8	**13.4**	10.4	**10.4**	0.7	2.5	7.4	7.1	**44.2**	**25.7**	12.7	**17.5**	7.2	5.4
10	10	11.1	15.1	**15.2**	13.6	**17.5**	13.7	**14.9**	1.0	3.9	5.5	6.4	**39.0**	**21.3**	11.6	**20.8**	8.5	**22.3**
0	20	38.6	**48.5**	48.4	46.9	**55.5**	42.7	**55.8**	14.1	27.9	15.2	17.2	**57.9**	**40.3**	26.9	**52.0**	23.6	**63.5**
20	0	24.7	33.0	**33.5**	31.0	**39.2**	28.9	**37.4**	6.3	14.4	11.4	13.3	**51.6**	**33.6**	21.1	**37.9**	14.4	5.1
20	20	37.9	45.4	**47.2**	42.8	**49.7**	39.7	**52.8**	11.1	25.6	8.4	11.6	**43.5**	**26.0**	18.7	**55.4**	24.3	**65.1**
20	-20	79.9	**86.7**	84.4	85.8	**91.2**	81.2	**92.9**	56.7	76.9	63.4	58.8	**89.5**	**87.1**	70.8	**88.1**	65.5	**64.5**
0	40	95.9	**97.5**	97.4	97.0	**98.4**	94.8	**99.2**	88.1	95.8	65.1	73.9	**93.4**	**84.8**	81.3	**98.8**	92.2	**99.6**
40	0	82.3	**87.5**	86.0	86.3	**91.5**	81.2	**93.0**	59.8	78.8	41.3	48.0	**80.8**	**68.4**	59.7	**90.5**	69.8	6.1
40	40	**96.8**	96.3	**96.8**	94.5	**97.5**	92.6	**98.8**	79.3	93.4	45.9	57.0	**84.0**	**67.7**	66.6	**98.9**	90.2	**99.7**
70	0	99.8	**99.9**	**99.9**	99.3	**100.0**	99.5	**100.0**	99.5	99.9	92.5	96.5	**99.5**	97.8	**98.2**	**100.0**	99.9	6.8
70	70	**100.0**	**100.0**	**100.0**	98.3	**100.0**	**100.0**	**100.0**	99.8	99.9	93.9	98.0	**99.8**	98.0	**98.7**	**100.0**	**100.0**	**100.0**
c)									$(n,s) = (80,30)$									
0	10	21.3	22.8	**23.5**	22.3	**26.3**	20.9	**22.2**	1.4	3.3	7.4	8.3	**25.0**	**22.2**	12.7	**17.2**	7.9	**30.0**
10	0	12.4	15.3	**16.8**	15.5	**17.9**	13.4	**13.0**	0.6	1.5	5.3	5.7	**23.5**	**19.2**	10.0	**11.7**	5.9	4.7
10	10	20.0	22.5	**24.3**	20.6	**23.1**	19.3	**21.2**	1.1	2.9	3.6	4.9	**17.3**	**12.9**	8.0	**20.2**	8.2	**31.3**
0	20	69.0	**70.9**	70.4	69.4	**75.5**	64.0	**76.4**	21.9	38.1	17.0	21.0	**40.7**	**37.3**	28.8	**59.7**	33.0	**82.5**
20	0	48.0	**51.6**	50.5	49.0	**56.2**	43.1	**54.1**	8.8	17.5	11.9	14.3	**33.1**	**29.7**	20.7	**36.6**	17.1	5.3
20	20	66.7	64.9	**67.4**	65.1	**69.0**	59.1	**73.4**	16.6	33.0	7.6	12.9	**23.1**	**19.6**	18.0	**66.2**	31.5	**84.0**
20	-20	97.3	**97.8**	97.0	97.7	**98.6**	95.6	**99.3**	78.9	90.8	78.3	74.9	**89.2**	**93.0**	82.7	**94.9**	82.0	**85.2**
0	40	**99.9**	**99.9**	99.8	99.9	**100.0**	99.6	**100.0**	97.4	99.5	79.0	87.3	**94.7**	**91.1**	90.8	**99.8**	98.8	**100.0**
40	0	**98.5**	98.2	98.2	98.3	**99.2**	96.2	**99.5**	81.2	92.8	51.7	60.5	**78.5**	**73.6**	69.6	79.0	**86.4**	5.9
40	40	**99.9**	99.7	99.8	**99.7**	**99.7**	99.1	**100.0**	92.6	98.9	50.1	66.5	**81.1**	69.5	**73.2**	**100.0**	98.8	**100.0**
70	0	**100.0**	**100.0**	**100.0**	99.9	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	98.3	99.5	**99.9**	99.6	**99.7**	**100.0**	**100.0**	6.7
70	70	**100.0**	**100.0**	**100.0**	99.7	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	97.8	99.8	**99.9**	99.5	**99.9**	**100.0**	**100.0**	**100.0**

with n: sample size, s: shift location, δ_Q: shift amplitude in Q, δ_V: shift amplitude in V and *: the depth function with which the test is originally defined. Gray color indicates a test power higher than 95%. Numbers written in bold and underlined indicate the best power of each test for the corresponding $(\delta_Q,\ \delta_V)$.

Table 4 : **Power comparison for the considered tests to detect shifts in *Q, V* or (*Q,V*) with standardized samples (with estimated standard deviation).**

δ_Q	δ_V	M			T			W			QIB			QIA		Z		C
		TD	MD	SD*	TD*	MD	SD	TD*	MD	SD	TD	MD*	SD	TD	MD*	TD	MD*	
a)									**(n,s) = (30,10)**									
0	10	1.1	**10.8**	0.0	8.4	**12.3**	9.1	**7.7**	0.7	4.2	10.2	7.6	**85.1**	**43.6**	21.1	**27.5**	9.3	**9.2**
10	0	0.8	**8.4**	0.1	6.2	**8.8**	6.6	**4.9**	0.3	2.6	9.9	6.9	**85.6**	**42.2**	19.5	**24.5**	7.8	**7.5**
10	10	1.0	**11.3**	0.1	7.9	**10.7**	8.5	**6.3**	0.2	3.5	7.2	6.3	**83.6**	**36.5**	17.3	**26.3**	8.8	**12.6**
0	20	3.5	**28.8**	0.1	24.1	**32.5**	23.5	**28.5**	4.8	19.4	18.0	15.0	**88.6**	**53.8**	32.4	**48.9**	20.9	**28.3**
20	0	2.2	**18.4**	0.1	14.7	**22.8**	14.9	**16.9**	2.4	10.6	12.9	10.6	**87.3**	**49.3**	26.1	**37.0**	13.5	**17.6**
20	20	3.4	**25.4**	0.1	21.9	**29.0**	21.0	**25.2**	3.5	16.9	11.6	11.8	**85.6**	**40.8**	25.3	**49.2**	20.0	**40.4**
20	-20	6.3	**53.8**	0.2	48.0	**64.6**	39.9	**60.9**	21.6	48.5	46.5	38.7	**96.2**	**84.8**	64.0	**78.2**	48.1	**52.8**
0	40	10.8	**79.0**	0.6	71.1	**86.2**	64.3	**85.5**	50.2	76.8	53.0	54.6	**97.3**	**83.6**	73.2	**93.4**	75.7	**87.0**
40	0	7.7	**59.1**	0.3	53.3	**68.3**	47.1	**65.0**	25.2	54.0	36.1	34.6	**94.3**	**72.5**	56.1	**79.3**	50.3	**66.3**
40	40	9.2	**74.3**	0.6	66.2	**81.1**	63.6	**81.1**	42.7	70.8	38.2	42.2	**94.4**	**68.8**	60.0	**91.7**	71.2	**93.7**
70	0	13.8	**96.1**	0.6	87.9	**98.4**	84.4	**98.5**	86.9	96.1	81.2	83.5	**99.7**	**95.8**	92.8	**99.6**	96.5	**99.1**
70	70	16.2	**99.1**	1.3	86.0	**99.3**	91.0	**99.3**	94.1	98.1	85.2	89.7	**99.9**	**95.3**	95.2	**99.9**	99.3	**100.0**
b)									**(n,s) = (50,20)**									
0	10	10.7	14.7	**14.8**	13.6	18.6	14.1	**14.2**	0.9	3.6	6.8	6.9	**29.4**	**16.7**	10.4	**22.4**	7.7	**15.3**
10	0	6.7	**10.6**	10.5	9.5	11.3	9.7	**9.3**	0.5	1.8	6.5	6.7	**27.3**	**16.0**	9.7	**18.3**	7.3	**10.1**
10	10	9.0	12.4	**12.8**	11.4	13.6	11.9	**11.9**	0.7	2.8	3.6	4.9	**19.9**	**10.8**	7.6	**25.3**	8.2	**19.5**
0	20	40.7	47.0	**47.8**	46.5	55.0	43.0	**54.8**	12.7	26.9	14.9	17.6	**41.6**	**29.9**	22.9	**55.2**	24.1	**50.4**
20	0	23.9	30.7	**31.8**	29.4	38.3	27.6	**34.9**	5.0	13.2	10.3	11.7	**33.7**	**22.0**	15.5	**37.7**	13.2	**31.9**
20	20	37.5	42.2	**46.0**	42.8	50.0	38.2	**53.6**	9.6	25.3	6.8	11.3	**27.8**	**16.9**	15.2	**63.3**	26.3	**67.4**
20	-20	79.3	**88.2**	83.4	85.7	91.2	80.4	**92.3**	57.8	77.0	65.4	61.6	**82.1**	**81.6**	68.9	**87.8**	64.7	**87.0**
0	40	96.6	**98.4**	97.5	97.6	98.6	96.6	**99.4**	88.1	96.5	68.0	77.1	**89.7**	**82.4**	82.2	**99.1**	94.2	**99.4**
40	0	83.3	**88.4**	87.9	87.4	92.2	84.4	**94.3**	58.7	80.8	41.7	49.8	**71.8**	**60.4**	56.8	**92.7**	69.1	**93.6**
40	40	96.0	**96.9**	**96.9**	95.4	97.4	94.5	**99.4**	77.8	94.0	40.5	55.0	**74.3**	57.0	**61.1**	**99.3**	90.9	**99.9**
70	0	**100.0**	**100.0**	**100.0**	99.5	99.9	100.0	**100.0**	99.8	**100.0**	95.3	98.3	**99.5**	98.4	**98.9**	**100.0**	**100.0**	**100.0**
70	70	**100.0**	**100.0**	**100.0**	98.8	100.0	100.0	**100.0**	**100.0**	**100.0**	94.9	99.1	**99.8**	98.4	**99.4**	**100.0**	**100.0**	**100.0**
c)									**(n,s) = (80,30)**									
0	10	21.3	**23.3**	22.7	21.9	27.1	19.5	**21.8**	1.2	3.1	6.3	8.0	**26.7**	**22.8**	12.8	**15.0**	7.0	**22.9**
10	0	13.2	**15.3**	14.7	14.1	16.5	13.4	**11.8**	0.5	1.2	6.1	6.3	**23.4**	**20.3**	10.4	**13.0**	6.9	**14.5**
10	10	18.8	20.4	**21.6**	19.5	22.5	16.9	**19.5**	0.7	2.6	3.2	4.6	**15.9**	**12.3**	7.3	**19.6**	8.0	**31.6**
0	20	71.1	**73.6**	72.1	71.7	76.5	66.5	**78.9**	21.5	39.0	17.7	21.5	**42.3**	**39.0**	29.7	**61.4**	32.4	**76.7**
20	0	49.3	51.2	**51.7**	50.4	57.5	45.7	**56.3**	7.5	16.7	10.9	13.6	**32.7**	**29.1**	20.3	**38.1**	16.2	**55.7**
20	20	66.2	65.2	**68.3**	64.3	68.6	59.7	**74.4**	15.8	33.4	7.0	12.7	**24.1**	**19.6**	17.1	**66.3**	32.1	**86.8**
20	-20	97.2	**97.7**	96.6	97.5	98.4	95.1	**98.9**	78.2	89.8	78.9	76.2	**90.1**	**92.8**	83.6	**94.3**	81.9	**98.3**
0	40	**99.9**	**99.9**	**99.9**	99.9	99.9	99.6	**100.0**	97.6	99.3	79.2	86.7	**94.6**	91.1	**91.2**	**99.9**	98.7	**100.0**
40	0	98.1	**98.3**	97.2	98.3	98.7	95.9	**99.5**	79.4	91.9	51.3	61.1	**77.5**	**73.2**	69.6	**97.0**	86.2	**99.4**
40	40	**99.8**	99.7	99.6	99.5	99.7	99.0	**100.0**	92.1	98.2	48.9	66.0	**79.8**	69.1	**73.1**	**99.9**	98.4	**100.0**
70	0	**100.0**	**100.0**	**100.0**	99.9	100.0	100.0	**100.0**	**100.0**	**100.0**	98.2	99.6	**99.9**	99.5	**99.8**	**100.0**	**100.0**	**100.0**
70	70	**100.0**	**100.0**	**100.0**	99.7	100.0	100.0	**100.0**	**100.0**	**100.0**	97.6	99.5	**99.9**	99.1	**99.8**	**100.0**	**100.0**	**100.0**

with *n*: sample size, *s*: shift location, δ_Q: shift amplitude in *Q*, δ_V: shift amplitude in *V* and *: the depth function with which the test is originally defined. Gray color indicates a test power higher than 95%. Numbers written in bold and underlined indicate the best power of each test for the corresponding $(\delta_Q,\ \delta_V)$.

Table 5 : Power comparison for the considered tests to detect shifts in *Q*, *V* or (*Q*,*V*) with standardized samples (with estimated or known standard deviation).

δ_Q	δ_V	M SD*	T TD*	W TD*	QIB MD	QIA MD*	Z MD*	C	M SD*	T TD*	W TD*	QIB MD	QIA MD*	Z MD*	C
		***Estimated* standard deviation**							***Known* standard deviation**						
		(n,s) = (30,10)							**(n,s) = (30,10)**						
0	10	0.0	8.4	7.7	7.6	21.1	9.3	9.2	0.1	6.4	5.4	7.2	19.0	8.3	7.1
10	0	0.1	6.2	4.9	6.9	19.5	7.8	7.5	0.0	7.1	5.2	6.8	19.0	7.7	7.3
10	10	0.1	7.9	6.3	6.3	17.3	8.8	12.6	0.1	6.7	5.8	6.3	17.8	8.2	10.6
0	20	0.1	24.1	28.5	15.0	32.4	20.9	28.3	0.1	15.6	17.0	10.6	26.2	13.7	19.4
20	0	0.1	14.7	16.9	10.6	26.1	13.5	17.6	0.1	16.0	18.3	11.8	26.6	15.0	20.1
20	20	0.1	21.9	25.2	11.8	25.3	20.0	40.4	0.2	16.2	19.2	9.7	21.9	16.3	32.5
20	-20	0.2	48.0	60.9	38.7	64.0	48.1	52.8	0.5	39.9	48.6	29.5	53.4	36.4	40.7
0	40	0.6	71.1	85.5	54.6	73.2	75.7	87.0	0.5	52.5	65.0	34.0	55.7	52.1	67.6
40	0	0.3	53.3	65.0	34.6	56.1	50.3	66.3	0.4	52.6	63.4	33.3	53.9	50.1	68.8
40	40	0.6	66.2	81.1	42.2	60.0	71.2	93.7	0.7	54.8	68.4	31.3	49.6	56.9	86.5
70	0	0.6	87.9	98.5	83.5	92.8	96.5	99.1	0.7	86.7	98.3	83.2	92.0	95.6	99.5
70	70	1.3	86.0	99.3	89.7	95.2	99.3	100.0	1.3	84.8	98.1	78.8	88.1	96.5	99.8
		(n,s) = (50,20)							**(n,s) = (50,20)**						
0	10	14.8	13.6	14.2	6.9	10.4	7.7	15.3	11.2	10.6	9.7	5.7	11.8	6.4	10.6
10	0	10.5	9.5	9.3	6.7	9.7	7.3	10.1	12.5	10.4	10.3	6.9	12.9	6.9	10.8
10	10	12.8	11.4	11.9	4.9	7.6	8.2	19.5	11.9	11.2	11.0	5.3	10.0	6.7	17.7
0	20	47.8	46.5	54.8	17.6	22.9	24.1	50.4	33.2	30.3	35.0	12.4	19.2	14.1	36.5
20	0	31.8	29.4	34.9	11.7	15.5	13.2	31.9	33.1	30.3	37.2	11.9	20.0	14.0	34.5
20	20	46.0	42.8	53.6	11.3	15.2	26.3	67.4	36.8	33.1	39.2	9.8	15.2	17.7	57.3
20	-20	83.4	85.7	92.3	61.6	68.9	64.7	87.0	74.5	75.7	86.0	47.0	60.2	50.7	75.2
0	40	97.5	97.6	99.4	77.1	82.2	94.2	99.4	85.9	86.2	93.4	48.0	59.0	69.4	93.1
40	0	87.9	87.4	94.3	49.8	56.8	69.1	93.6	87.1	86.6	93.2	47.9	59.0	70.1	93.6
40	40	96.9	95.4	99.4	55.0	61.1	90.9	99.9	89.8	86.7	95.1	38.7	48.6	76.3	99.2
70	0	100.0	99.5	100.0	98.3	98.9	100.0	100.0	99.9	99.6	100.0	97.1	98.1	100.0	100.0
70	70	100.0	98.8	100.0	99.1	99.4	100.0	100.0	99.9	98.5	100.0	93.9	96.0	99.9	100.0
		(n,s) = (80,30)							**(n,s) = (80,30)**						
0	10	22.7	21.9	21.8	8.0	12.8	7.0	22.9	15.7	15.8	13.1	6.6	10.4	6.2	15.0
10	0	14.7	14.1	11.8	6.3	10.4	6.9	14.5	15.6	14.7	13.2	6.6	10.8	6.4	16.5
10	10	21.6	19.5	19.5	4.6	7.3	8.0	31.6	17.6	16.1	15.6	4.7	8.0	7.6	26.2
0	20	72.1	71.7	78.9	21.5	29.7	32.4	76.7	49.6	49.4	54.4	12.5	18.9	15.7	53.3
20	0	51.7	50.4	56.3	13.6	20.3	16.2	55.7	49.9	50.6	55.1	13.4	19.6	16.4	52.9
20	20	68.3	64.3	74.4	12.7	17.1	32.1	86.8	55.1	52.8	60.3	10.3	14.9	23.3	76.5
20	-20	96.6	97.5	98.9	76.2	83.6	81.9	98.3	91.9	94.2	96.9	58.5	68.0	63.3	94.8
0	40	99.9	99.9	100.0	86.7	91.2	98.7	100.0	98.1	98.1	99.5	62.8	70.9	86.8	99.4
40	0	97.2	98.3	99.5	61.1	69.6	86.2	99.4	98.1	98.1	99.4	61.7	70.0	86.1	99.4
40	40	99.6	99.5	100.0	66.0	73.1	98.4	100.0	98.4	98.1	99.6	48.8	56.2	92.3	99.9
70	0	100.0	99.9	100.0	99.6	99.8	100.0	100.0	100.0	100.0	100.0	99.7	99.8	100.0	100.0
70	70	100.0	99.7	100.0	99.5	99.8	100.0	100.0	100.0	100.0	100.0	98.3	98.8	100.0	100.0

Table 6 : Power evaluation of the considered tests with various combinations of n and s.

δ_Q	δ_V	M			T			W			QIB			QIA		Z		C
		TD	MD	SD*	TD*	MD	SD	TD*	MD	SD	TD	MD*	SD	TD	MD*	TD	MD*	
a)								**(n,s) = (30,5)**										
0	10	0,0	**8,0**	0,0	6,1	**9,9**	6,9	2,5	0,0	**2,7**	14,9	8,6	**100,0**	**68,4**	38,8	**38,2**	16,0	**7,3**
10	0	0,0	**7,3**	0,0	4,8	**7,7**	7,0	1,6	0,0	**1,7**	14,2	7,4	**100,0**	**69,7**	38,3	**36,7**	14,1	**6,3**
10	10	0,0	**10,5**	0,0	6,4	**10,4**	7,6	2,4	0,0	**2,8**	12,9	8,1	**100,0**	**65,8**	37,3	**38,3**	15,3	**10,4**
0	20	0,0	**14,7**	0,0	15,2	**23,5**	14,9	9,5	0,1	**10,2**	20,3	12,8	**100,0**	**72,6**	45,9	**50,9**	24,8	**18,6**
20	0	0,0	**13,0**	0,0	11,1	**16,3**	11,4	5,3	0,1	**5,6**	16,7	10,0	**100,0**	**71,9**	43,0	**45,7**	18,6	**12,8**
20	20	0,0	**17,5**	0,0	15,6	**22,6**	13,9	10,7	0,1	**11,0**	16,1	10,6	**100,0**	**67,2**	41,4	**49,2**	22,2	**27,6**
20	-20	0,0	**23,5**	0,0	26,0	**42,7**	27,9	**21,7**	1,0	21,2	36,7	24,9	**100,0**	**89,4**	66,9	**71,2**	42,3	**32,7**
0	40	0,0	**41,0**	0,0	44,5	**66,4**	44,8	47,5	3,7	**49,7**	50,3	39,5	**100,0**	**89,2**	74,7	**85,5**	63,9	**65,3**
40	0	0,0	**29,7**	0,0	31,1	**47,4**	30,4	28,0	1,1	**29,4**	35,7	25,9	**100,0**	**84,0**	63,6	**73,6**	45,7	**43,0**
40	40	0,0	**44,9**	0,0	41,1	**61,7**	42,2	47,0	3,4	**48,1**	42,6	36,0	**100,0**	**81,9**	66,4	**81,3**	59,5	**78,4**
70	0	0,0	**64,3**	0,0	60,3	**88,3**	68,4	76,6	19,4	**77,8**	74,1	65,9	**100,0**	**95,2**	88,5	**95,3**	87,0	**90,2**
70	70	0,0	**81,7**	0,0	62,5	**93,6**	81,0	86,3	39,9	**86,9**	82,0	77,8	**100,0**	**96,3**	92,0	**98,0**	93,2	**99,3**
b)								**(n,s) = (50,5)**										
0	10	0,0	**9,0**	0,0	8,0	**10,1**	7,1	**1,6**	0,0	**1,6**	11,1	7,5	**100,0**	**73,9**	41,6	**39,2**	15,2	**7,8**
10	0	0,0	**6,8**	0,0	6,1	**7,9**	5,9	**0,9**	0,0	**0,9**	10,2	5,9	**100,0**	**72,9**	38,7	**35,9**	14,5	**6,2**
10	10	0,0	**9,4**	0,0	8,6	**11,1**	7,2	1,7	0,0	**2,0**	8,9	6,6	**100,0**	**66,7**	37,5	**34,7**	14,4	**10,2**
0	20	0,0	**17,6**	0,0	20,7	**27,0**	15,9	8,0	0,0	**8,5**	17,1	12,9	**100,0**	**76,2**	47,6	**51,3**	25,4	**21,3**
20	0	0,0	**13,0**	0,0	12,4	**18,4**	11,1	4,2	0,0	**4,6**	14,1	9,2	**100,0**	**74,2**	44,2	**45,8**	20,1	**13,7**
20	20	0,0	**18,3**	0,0	20,0	**27,2**	14,8	9,5	0,0	**9,9**	13,2	10,5	**100,0**	**69,5**	43,5	**49,5**	23,1	**31,1**
20	-20	0,0	**25,0**	0,0	37,0	**49,6**	31,3	**19,2**	0,5	18,9	35,1	25,9	**100,0**	**92,0**	70,4	**74,9**	46,6	**36,6**
0	40	0,0	**43,7**	0,0	54,5	**71,4**	46,7	45,4	2,9	**46,9**	47,2	40,7	**100,0**	**91,0**	76,4	**86,6**	65,4	**71,0**
40	0	0,0	**32,2**	0,0	41,8	**55,7**	33,5	28,1	0,7	**29,3**	33,0	27,0	**100,0**	**85,4**	65,8	**75,2**	49,2	**51,8**
40	40	0,0	**45,2**	0,0	53,2	**67,4**	43,7	46,8	2,2	**47,5**	41,2	35,5	**100,0**	**83,0**	68,8	**83,1**	62,2	**82,2**
70	0	0,0	**65,6**	0,0	75,4	**92,1**	70,5	76,8	16,1	**77,9**	73,1	66,9	**100,0**	**96,5**	90,6	**96,6**	89,6	**93,9**
70	70	0,0	**82,5**	0,0	80,5	**95,9**	79,9	87,2	39,2	**87,5**	82,3	78,8	**100,0**	**96,8**	92,7	**98,4**	94,8	**99,6**
c)								**(n,s) = (50,10)**										
0	10	7,7	**12,6**	6,8	10,2	**13,6**	10,8	**5,1**	0,1	2,2	9,2	7,9	**80,7**	**47,3**	20,7	**25,3**	9,2	**10,5**
10	0	5,7	**9,7**	5,4	8,1	**11,0**	9,1	**3,5**	0,1	1,4	8,8	6,8	**82,1**	**49,0**	20,7	**23,4**	7,3	**8,1**
10	10	7,7	**13,6**	6,0	9,9	**13,9**	9,8	**4,9**	0,1	1,7	7,0	5,9	**79,1**	**40,8**	19,1	**24,8**	9,0	**14,6**
0	20	22,3	**31,6**	17,0	32,4	**39,5**	28,3	**24,5**	2,6	14,8	15,7	14,1	**86,6**	**57,4**	34,1	**48,4**	21,5	**35,3**
20	0	16,3	**24,0**	13,4	21,9	**27,7**	20,2	**14,9**	1,2	7,7	12,1	10,8	**84,6**	**53,9**	29,1	**37,4**	13,9	**23,6**
20	20	23,6	**32,3**	18,6	29,2	**37,2**	27,1	**24,3**	2,7	14,9	10,7	11,4	**81,3**	**44,1**	26,6	**48,6**	20,9	**50,0**
20	-20	46,1	**62,4**	32,6	64,8	**75,5**	51,1	**61,5**	17,6	45,5	49,5	42,4	**95,4**	**89,5**	69,0	**81,2**	52,9	**66,0**
0	40	62,5	**85,5**	44,6	83,0	**91,6**	69,9	**85,7**	48,8	76,3	56,0	57,7	**97,1**	**87,2**	77,0	**94,0**	80,9	**92,6**
40	0	49,2	**66,8**	36,9	67,0	**77,5**	54,8	**65,1**	23,7	52,6	37,8	38,9	**93,5**	**76,1**	60,5	**81,8**	58,3	**77,0**
40	40	61,6	**82,1**	46,1	76,9	**86,7**	69,0	**82,6**	41,7	70,2	41,6	46,4	**93,8**	**74,1**	64,8	**93,4**	77,2	**97,2**
70	0	70,4	**98,2**	52,8	95,4	**99,5**	87,5	**98,7**	88,6	97,0	85,4	88,6	**99,8**	**97,3**	95,4	**99,7**	98,4	**99,9**
0	70	69,7	**99,4**	53,5	93,1	**99,8**	91,7	**99,5**	94,0	98,2	85,8	90,8	**99,7**	**96,1**	95,5	**99,8**	99,5	**100,0**

δ_Q	δ_V	M			T			W			QIB			QIA		Z		C
		TD	MD	SD*	TD*	MD	SD	TD*	MD	SD	TD	MD*	SD	TD	MD*	TD	MD*	
d)									**(n,s) = (80,5)**									
0	10	0,0	**9,2**	0,0	9,1	**11,6**	8,4	**1,4**	0,0	**1,4**	9,5	6,6	**100,0**	**74,1**	41,3	**38,5**	16,0	**8,6**
10	0	0,0	**8,1**	0,0	6,9	**9,5**	7,8	**1,0**	0,0	0,9	8,3	6,0	**100,0**	**74,1**	41,0	**37,2**	15,3	**6,6**
10	10	0,0	**10,7**	0,0	9,8	**12,2**	8,5	**1,7**	0,0	**1,7**	8,7	6,2	**100,0**	**69,2**	39,2	**37,5**	15,1	**11,5**
0	20	0,0	**17,3**	0,0	22,7	**29,7**	19,7	8,1	0,0	**8,8**	16,7	13,2	**100,0**	**77,7**	51,4	**53,1**	26,6	**23,4**
20	0	0,0	**12,2**	0,0	15,0	**19,7**	12,0	3,2	0,0	**3,5**	11,0	8,9	**100,0**	**76,0**	44,5	**45,1**	19,5	**14,6**
20	20	0,0	**18,4**	0,0	20,7	**29,0**	17,4	7,6	0,0	**7,8**	11,5	9,8	**100,0**	**70,9**	43,8	**49,8**	22,7	**32,5**
20	-20	0,0	**25,3**	0,0	43,1	**53,3**	39,0	**18,7**	0,3	18,0	33,2	26,0	**100,0**	**93,5**	73,0	**75,9**	49,1	**40,8**
0	40	0,0	**43,8**	0,0	60,8	**74,8**	54,1	46,7	2,0	**47,3**	46,0	40,6	**100,0**	**91,5**	77,8	**87,0**	67,6	**75,3**
40	0	0,0	**30,9**	0,0	43,8	**56,1**	37,9	25,7	0,4	**26,9**	30,9	26,0	**100,0**	**86,2**	64,5	**74,2**	48,6	**51,6**
40	40	0,0	**43,5**	0,0	56,8	**70,5**	48,6	45,7	1,4	**46,2**	38,2	34,6	**100,0**	**83,3**	68,3	**83,2**	64,0	**83,8**
70	0	0,0	**66,1**	0,0	81,7	**93,8**	75,0	76,9	16,2	**78,3**	73,9	67,8	**100,0**	**97,5**	90,7	**97,2**	89,9	**96,0**
0	70	0,0	**81,0**	0,0	84,7	**95,6**	82,7	86,4	38,1	**87,0**	81,9	78,8	**100,0**	**96,6**	93,0	**98,7**	95,3	**99,7**
e)									**(n,s) = (80,10)**									
0	10	9,4	**13,3**	10,3	12,3	**15,7**	10,7	**4,0**	0,1	1,5	7,8	7,2	**74,9**	**49,4**	22,1	**23,9**	8,7	**12,1**
10	0	7,3	**10,1**	9,4	9,8	**11,7**	9,7	**2,9**	0,1	0,9	6,9	6,3	**75,4**	**50,3**	20,9	**22,5**	8,1	**9,7**
10	10	9,3	**14,1**	10,0	11,8	**15,7**	10,8	**3,9**	0,2	1,9	5,4	5,9	**69,6**	**43,5**	19,4	**22,3**	8,3	**16,5**
0	20	34,2	**37,1**	33,5	40,5	**46,4**	32,0	**24,6**	2,4	14,7	15,4	16,1	**79,9**	**59,7**	36,4	**48,6**	23,2	**41,1**
20	0	21,7	**25,3**	21,2	26,2	**32,4**	22,8	**12,9**	0,8	6,7	10,7	10,7	**78,4**	**56,3**	29,7	**36,0**	14,9	**26,3**
20	20	34,8	**36,7**	33,3	38,6	**44,0**	31,4	**24,7**	2,1	14,2	9,4	11,2	**72,7**	**47,7**	28,5	**48,3**	20,6	**56,8**
20	-20	64,9	**69,8**	59,8	74,7	**81,9**	56,0	**62,4**	15,7	44,9	49,1	44,3	**93,6**	**90,4**	70,5	**83,0**	57,2	**72,9**
0	40	80,9	**88,2**	76,1	88,7	**93,9**	71,6	**85,7**	47,5	75,1	57,0	59,5	**95,5**	**88,6**	79,0	**95,1**	82,0	**95,8**
40	0	65,4	**71,8**	61,5	74,2	**82,0**	57,6	**65,0**	20,7	50,1	35,4	37,4	**91,0**	**78,8**	62,0	**83,7**	58,6	**82,0**
40	40	79,5	**83,5**	74,9	82,9	**89,1**	70,7	**81,9**	39,3	69,3	40,5	46,3	**91,0**	**76,3**	65,3	**93,7**	78,0	**98,0**
70	0	87,6	**99,1**	83,6	97,2	**99,7**	86,5	**98,7**	86,2	95,9	84,2	87,4	**99,4**	**97,1**	94,5	**99,7**	98,4	**99,9**
0	70	89,6	**99,7**	85,3	95,8	**99,8**	91,7	**99,5**	94,5	98,3	87,4	92,0	**99,8**	**97,2**	96,6	**99,9**	99,6	**100,0**
f)									**(n,s) = (80,20)**									
0	10	14,6	18,8	**19,5**	18,3	**22,4**	17,7	**12,1**	0,3	2,1	7,4	7,7	**37,0**	**30,9**	15,7	**16,6**	6,4	**18,1**
10	0	10,4	13,3	**14,0**	12,1	**16,0**	13,5	**7,1**	0,1	1,0	6,2	5,7	**34,8**	**28,7**	13,1	**13,4**	5,5	**12,7**
10	10	14,1	19,5	**20,0**	17,2	**21,0**	16,9	**11,5**	0,6	2,3	4,3	5,6	**28,0**	**21,4**	11,5	**17,9**	7,8	**26,8**
0	20	56,0	60,9	**61,1**	58,2	**67,1**	54,3	**56,6**	11,1	24,9	17,0	19,4	**52,5**	**46,0**	31,0	**53,5**	27,2	**64,2**
20	0	37,8	41,2	**42,1**	39,8	**47,2**	36,5	**35,1**	3,3	10,1	10,9	11,9	**42,4**	**37,2**	22,3	**34,0**	14,9	**44,0**
20	20	52,8	53,6	**58,6**	53,8	**58,6**	49,1	**53,2**	8,4	22,6	8,1	11,9	**34,5**	**27,4**	20,9	**57,9**	28,7	**78,5**
20	-20	92,0	**93,4**	92,0	93,4	**96,0**	88,5	**94,4**	54,4	76,8	69,9	66,9	**90,0**	**92,6**	79,7	**92,1**	73,5	**94,5**
0	40	99,1	**99,2**	99,0	98,8	**99,5**	97,0	**99,6**	88,1	96,4	71,0	78,6	**93,3**	**89,6**	85,6	**99,6**	96,4	**99,9**
40	0	93,2	**94,3**	93,6	93,7	**96,1**	87,2	**95,1**	58,4	78,9	45,3	52,0	**79,0**	**72,8**	64,9	**93,4**	78,5	**97,3**
40	40	**98,7**	98,4	98,2	97,7	**98,5**	94,7	**99,1**	78,5	92,9	45,7	60,2	**82,1**	**71,2**	70,1	**99,2**	94,3	**100,0**
70	0	**100,0**	**100,0**	**100,0**	99,8	**100,0**	99,7	**100,0**	99,8	99,9	95,0	97,8	**99,8**	98,9	**99,2**	**100,0**	**100,0**	**100,0**
0	70	**100,0**	**100,0**	**100,0**	99,2	**100,0**	**100,0**	**100,0**	99,9	**100,0**	95,2	99,0	**99,9**	98,8	**99,5**	**100,0**	**100,0**	**100,0**

Table 7 : Type one error estimate and power evaluation with respect to the dependence level

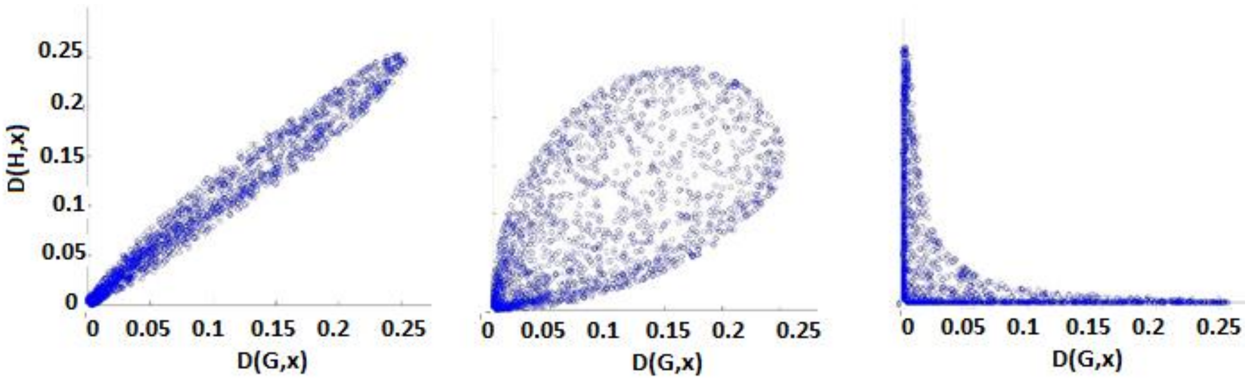
δ_Q	δ_V	M SD*	T TD*	W TD*	QIB MD	QIA MD*	Z MD*	C	M SD*	T TD*	W TD*	QIB MD	QIA MD*	Z MD*	C	M SD*	T TD*	W TD*	QIB MD	QIA MD*	Z MD*
		***Estimated* standard deviation**							***Estimated* standard deviation**							***Estimated* standard deviation**					
		(n,s) = (50,20)							**(n,s) = (50,20)**							**(n,s) = (50,20)**					
		b = 1.155 (with rho = 0.25)							**b = 1.414 (with rho = 0.50)**							**b = 2.000 (with rho = 0.75)**					
0	0	4.9	3.4	2.4	6.0	11.3	5.2	4.1	5.4	4.6	3.9	6.6	12.4	5.8	5.4	4.7	3.9	3.0	6.3	11.5	5.4
0	10	13.8	12.0	12.9	7.2	12.5	6.6	15.9	14.8	13.6	14.2	6.9	10.4	7.7	15.3	20.8	19.8	21.9	10.2	17.7	8.2
10	0	10.6	9.0	8.7	6.3	11.3	6.6	11.1	10.5	9.5	9.3	6.7	9.7	7.3	10.1	15.9	15.6	15.5	7.7	13.9	7.0
10	10	17.3	14.8	15.3	5.9	10.0	9.4	21.6	12.8	11.4	11.9	4.9	7.6	8.2	19.5	13.9	13.1	13.1	5.7	10.8	6.8
0	20	41.6	38.0	46.9	13.5	21.5	19.4	50.7	47.8	46.5	54.8	17.6	22.9	24.1	50.4	70.1	71.7	82.2	34.2	47.2	45.3
20	0	27.9	26.4	30.3	10.4	17.1	12.3	33.9	31.8	29.4	34.9	11.7	15.5	13.2	31.9	48.6	51.0	57.3	21.5	32.1	24.1
20	20	53.1	48.9	60.9	14.1	21.5	29.4	71.9	46.0	42.8	53.6	11.3	15.2	26.3	67.4	42.6	38.4	47.8	11.0	18.1	19.7
20	-20	68.2	68.5	79.6	41.2	55.0	42.5	80.1	83.4	85.7	92.3	61.6	68.9	64.7	87.0	97.3	98.6	99.7	89.7	94.5	95.3
0	40	94.2	92.5	97.8	58.7	69.5	84.1	99.0	97.5	97.6	99.4	77.1	82.2	94.2	99.4	99.9	99.5	100	95.8	97.9	99.8
40	0	80.7	77.8	88.8	36.2	47.3	56.9	90.9	87.9	87.4	94.3	49.8	56.8	69.1	93.6	97.3	97.9	99.6	79.5	86.2	93.6
40	40	97.7	96.7	99.2	62.6	71.3	95.0	100	96.9	95.4	99.4	55.0	61.1	90.9	99.9	94.7	91.8	97.5	52.6	62.1	84.8
70	0	99.8	99.3	100	91.2	94.4	99.2	100	100	99.5	100	98.3	98.9	100	100	100	97.7	100	99.9	100	100
70	70	100	97.6	100	99.0	99.7	100	100	100	98.8	100	99.1	99.4	100	100	100	98.8	100	97.6	98.7	100

Table 8 : General information about the *Moisie*, *Magpie* and *Romaine* stations.

Station name	Station number	Latitude	Longitude	Period of record (#years)	Area (Km^2)
Moisie	072301	50 21 09	-66 11 12	1968-1998 (31)	19 012
Magpie	073503	50 41 08	-64 34 43	1979-2004 (26)	7 201
Romaine	073801	50 18 28	-63 37 07	1961-2006 (46)	12 922

Table 9 : Test statistics and p-values of M-, T-, QIB-, QIA-, Z- and C-test and decision (1: shift, 0: no shift) of W-test.

Tests		M			T			W			QIB			QIA		Z		Cramer
		TD	MD	SD*	TD*	MD	SD	TD*	MD	SD	TM	MD*	SD	TD	MD*	TD	MD*	
a) Without standardizing the samples																		
Moisie	Stat	0.08	0.00	0.00	0.32	0.10	0.19	0.22	0.06	0.06	0.10	0.06	0.00	0.10	0.06	16.0	20.5	9353.21
	p-val	0.00	0.01	0.05	0.00	0.00	0.00	1	1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Magpie	Stat	0.30	0.00	0.00	0.60	0.20	0.30	0.5	0.3	0.3	0.3	0.19	0.00	0.30	0.19	0.99	2.92	1132.50
	p-val	0.16	0.54	0.53	0.08	0.09	0.07	0	0	0	0.47	0.32	0.02	0.11	0.02	0.32	0.09	0.03
Romaine	Stat	0.50	0.10	0.10	0.70	0.30	0.30	0.63	0.50	0.60	0.40	0.31	0.10	0.40	0.31	2.50	4.08	2478.40
	p-val	0.03	0.07	0.07	0.03	0.09	0.19	0	1	1	0.10	0.11	0.04	0.05	0.01	0.11	0.04	0.01
b) With standardizing the samples																		
Moisie	Stat	0.07	0.00	0.00	0.31	0.10	0.20	0.21	0.05	0.04	0.09	0.04	0.00	0.09	0.04	17.1	21.2	9.13
	p-val	0.00	0.03	0.03	0.00	0.00	0.00	1	1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Magpie	Stat	0.26	0.00	0.00	0.54	0.18	0.22	0.45	0.24	0.21	0.24	0.13	0.00	0.24	0.13	2.47	5.74	3.19
	p-val	0.06	0.34	0.38	0.02	0.04	0.01	0	1	1	0.18	0.09	0.01	0.03	0.00	0.12	0.02	0.00
Romaine	Stat	0.47	0.08	0.06	0.68	0.24	0.24	0.58	0.44	0.52	0.35	0.31	0.13	0.35	0.31	3.37	5.64	4.53
	p-val	0.02	0.04	0.05	0.01	0.01	0.05	1	1	1	0.09	0.07	0.02	0.04	0.01	0.07	0.02	0.00



a) b) c)

Figure 1 : DD-plot for a) two identical subsamples, b) two different subsamples and c) two very different subsamples.

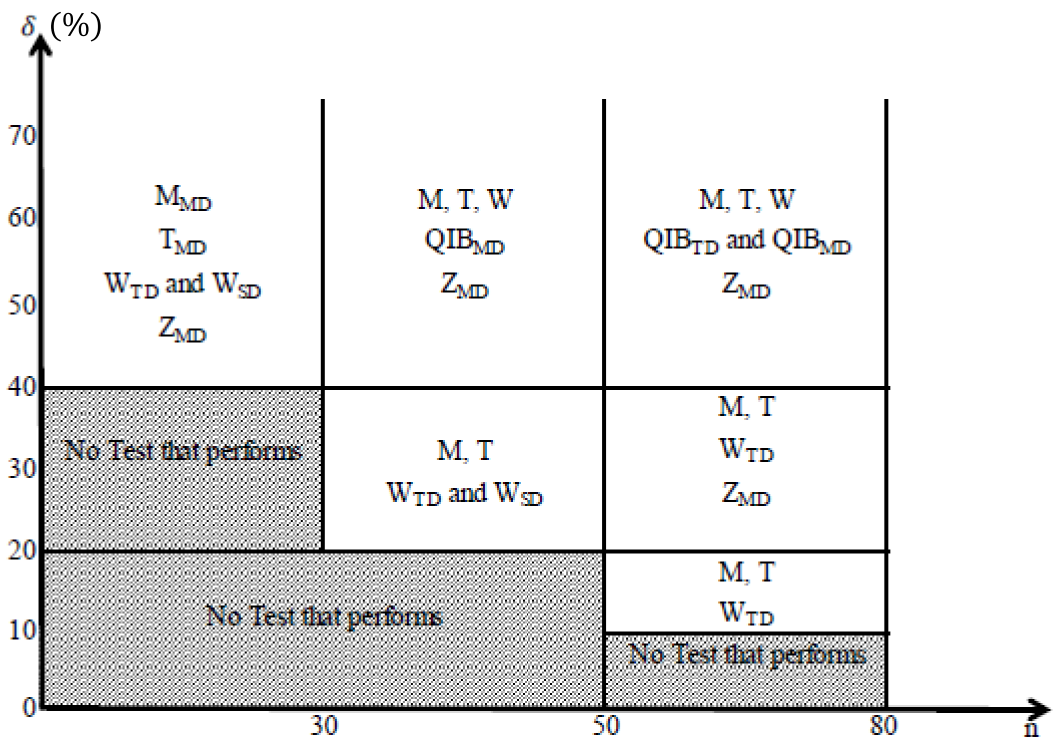


Figure 2 : Diagram of the applicability of considered tests for studied sample lengths (*n*) and shift amplitudes (δ).

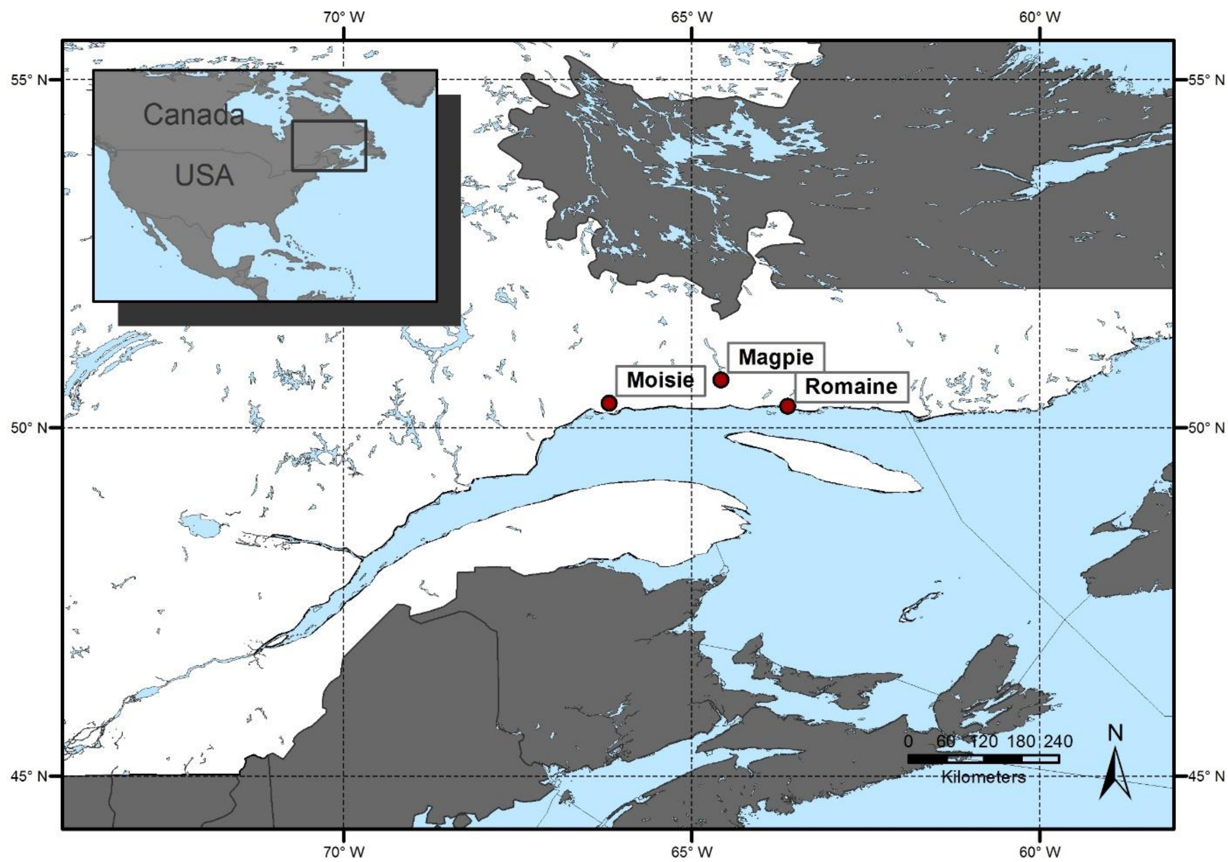
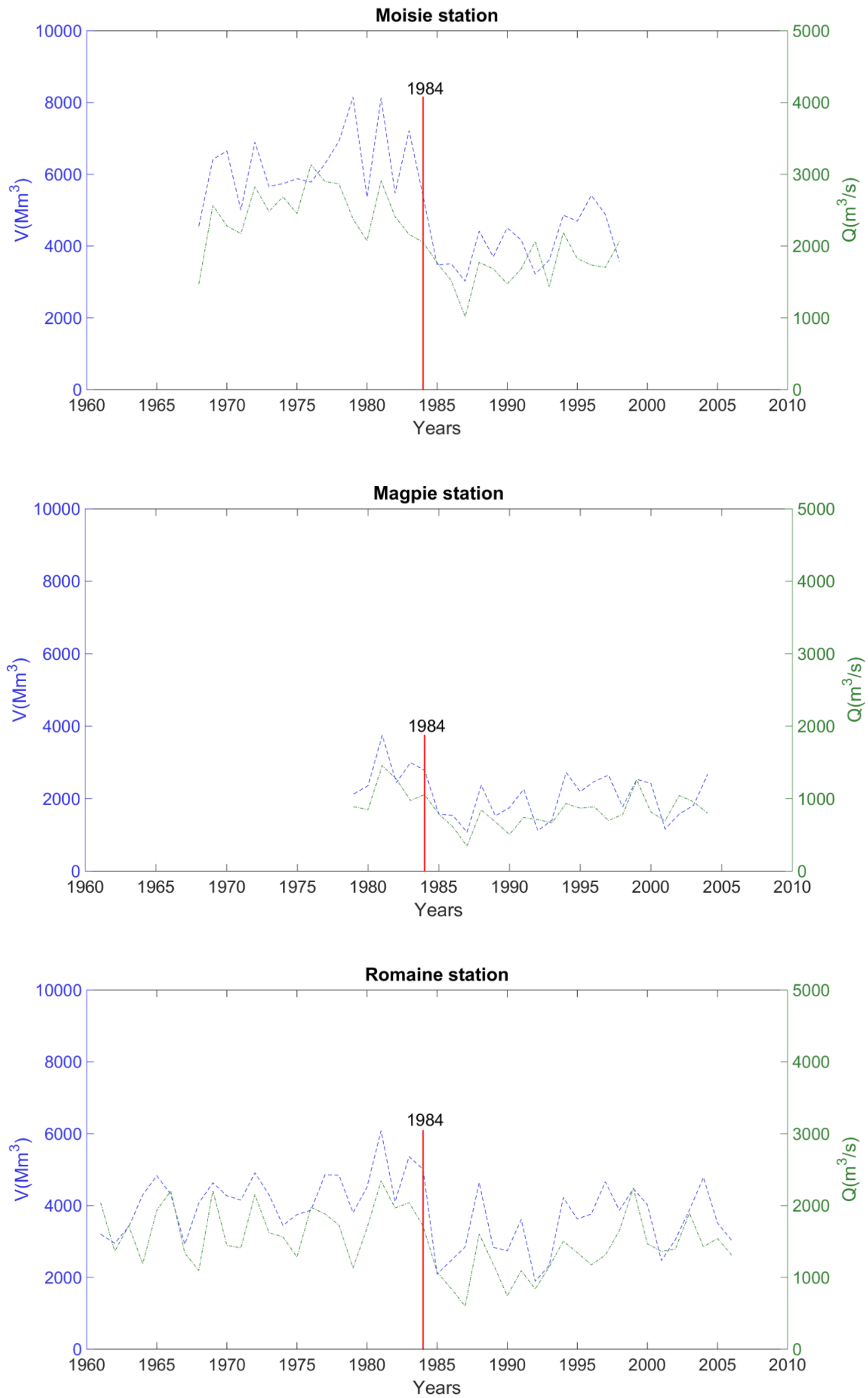


Figure 3 : Geographical location of the *Moisie*, *Magpie* and *Romaine* stations.



a)

b)

c)

Figure 4 : The *V* and *Q* time series of a) *Moisie*, b) *Magpie* and c) *Romaine* stations.