Detection of change-points in dependence structures

Quessy, Jean-François
Université du Québec à Trois-Rivières, Département de mathématiques et d'informatique
3351, boul. des Forges
Trois-Rivières (QC), G9A 5H7, Canada
Jean-Francois.Quessy@uqtr.ca

Saïd, Mériem
Université Laval, département de mathématiques et de statistique
1045, avenue de la Médecine
Québec (QC), G1V 0A6, Canada
meriem.said@mat.ulaval.ca

Champagne, Maryse
Statistique Canada
150, Promenade du pré Tunney
Ottawa (ON), K1A 0T6, Canada
Maryse.Champagne@statcan.gc.ca

Favre, Anne-Catherine
Université Laval, département de mathématiques et de statistique
1045, avenue de la Médecine
Québec (QC), G1V 0A6, Canada
anne-catherine.favre@mat.ulaval.ca

and
Grenoble INP/ENSE3, LTHE UMR 5564
BP 53
Grenoble, F-38041, France
anne-catherine.favre-pugin@ense3.grenoble-inp.fr

1 Introduction

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be observations taken at equal time intervals and for each $i \in \{1, \ldots, n\}$, define $H_i(x, y) = \mathbb{P}(X_i \leq x, Y_i \leq y)$. Suppose one is asking if at some undetermined moment $\mathcal{B} \in \{1, \ldots, n-1\}$, the joint distribution of these vectors changes. In other words, one wants to test for the null and alternative hypotheses

$$\begin{aligned} \mathcal{H}_0' : &\quad H_1 = \cdots = H_n \\ \mathcal{H}_1' : &\quad H_1 = \cdots = H_{\mathcal{B}} \neq H_{\mathcal{B}+1} = \cdots = H_n \text{ for some } \mathcal{B} \in \{1, \ldots, n-1\}. \end{aligned}$$

Here, it is understood that $H_{\mathcal{B}} \neq H_{\mathcal{B}+1}$ means that $H_{\mathcal{B}}(x_0, y_0) \neq H_{\mathcal{B}+1}(x_0, y_0)$ for at least one pair $(x_0, y_0) \in \mathbb{R}^2$.

Some statistical procedures have been proposed to compare $\mathcal{H}_0'$ and $\mathcal{H}_1'$. Many of them are parametric, *i.e.* $H_1, \ldots, H_n$ are supposed to belong to a given parametric family. This approach was privileged by Gombay and Horváth (1997), Jandhyala et al. (2009) and Brodsky and Darkhovsky (2005). Nonparametric methods have also been proposed (see e.g. Zou et al. (2007), Antoch and Hušková (2001), Antoch et al. (2008), Csörgő and Horváth (1988) and Gombay and Horváth (1995)).

It is worth noting that the alternative hypothesis $\mathcal{H}_1'$ hides three possible sources of change-point, namely

(i) a change in the univariate distribution of the series $X_1, \ldots, X_n$;

(ii) a change in the univariate distribution of the series $Y_1, \ldots, Y_n$;

(iii) a change in the dependence structure of (X_i, Y_i), $i \in \{1, \ldots, n\}$.

As a consequence, if one of the methodologies cited above is used, it is not possible to identify the source of the change in the event of a rejection of $\mathcal{H}_0'$. In the study of many phenomena, especially in finance, hydrology and climatology, being able to identify the nature of a change-point, *i.e.* patterns of type (i), (ii) and (iii), would give precious informations on the underlying process that generates the observed data.

In this paper, tests for the detection of change-points of type (iii) are proposed. So far, the only available procedure for that problem is the test of Dias and Embrechts (2009). However, this method is parametric in the sense that the dependence structure is supposed to belong to a given parametric family and the margins are assumed to be known. These assumptions are often unrealistic in practice. The statistics proposed in this work assume nothing about the form of the underlying distributions. The starting point is the possibility to extract the dependence structure of a multivariate distribution, *i.e.* its copula. Specifically, if one assumes that for all $i \in \{1, \ldots, n\}$, the marginal distributions $F_i(x) = \mathbb{P}(X_i \leq x)$ and $G_i(y) = \mathbb{P}(Y_i \leq y)$ are continuous, then Sklar's theorem ensures that there exist unique copulas $C_1, \ldots, C_n$ such that

$$H_i(x, y) = C_i\{F_i(x), G_i(y)\};$$

see Sklar (1959). With this notation, the null and alternative hypotheses of a change-point in the dependence structure of a bivariate time series can be stated as

$$\begin{aligned} \mathcal{H}_0 : &\quad C_1 = \cdots = C_n \\ \mathcal{H}_1 : &\quad C_1 = \cdots = C_{\mathcal{B}} \neq C_{\mathcal{B}+1} = \cdots = C_n \text{ for some } \mathcal{B} \in \{1, \ldots, n-1\}. \end{aligned}$$

The procedures proposed in this work are based on Kendall's measure of association. This dependence index is attractive in our context since its population value depends only on the underlying copula of a random couple. Moreover, its sample version is easy to compute and a quick, valid re-sampling procedure is available for inference purposes. The test statistics that will be investigated in this work are L_1–, L_2–, and L_∞–distances of an empirical process defined as differences of Kendall's measures of association. Their asymptotic distributions are derived and the validity of a re-sampling procedure to compute p-values is established.

The paper is organized as follows. In Section 2, sample and bootstrapped versions of Kendall's tau are described. In Section 3, three test statistics for the detection of a change-point in the copula structure of a bivariate time series are proposed and the required asymptotics are established. Examples on simulation series from the Canadian regional climate model are offered in Section 4.

2 Kendall's tau : population, sample and bootstrapped versions

Consider a pair (X, Y) of random variables with cumulative distribution function $H(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ and marginals F and G. Kendall's tau for this random vector is defined by $\tau(H) = 4\mathbb{E}_H\{H(X, Y)\} - 1$, where $\mathbb{E}_H$ denotes expectation with respect to H. By the change of variable $u = F(x)$ and $v = G(y)$, one can show that

$$\tau(H) = \tau(C) = 4 \int_{[0,1]^2} C(u, v) \mathrm{d}C(u, v) - 1,$$

where C is the unique copula associated to H via $C(u, v) = H\{F^{-1}(u), G^{-1}(v)\}$. It is then clear that Kendall's measure of association is only linked to the dependence structure of H. This feature will be exploited later in order to come up with statistical methodologies for the detection of change-points in the copula that underlies the joint distribution of bivariate observations.

Now suppose that $n \geq 2$ independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y) are observed. An unbiased estimation of $\tau(C)$ based on the subsample (X_I, Y_I), …, (X_J, Y_J), where $1 \leq I < J \leq n$, is given by

$$(1) \qquad \tau_{I:J} = \frac{2}{(J - I + 1)(J - I)} \sum_{i,j=I}^{J} Q_{ij} - 1,$$

where

$$Q_{ij} = \mathbb{I}\{(X_i - X_j)(Y_i - Y_j) > 0\}$$

and $\mathbb{I}(\cdot)$ is the indicator of a set. When $Q_{ij} = 1$, the pairs (X_i, Y_i) and (X_j, Y_j) are said to be concordant. When the number of observations used to estimate Kendall's tau increases to infinity, $\tau_{I:J}$ tends to the normal distribution. In particular, $T_{1:n} = \sqrt{n}\{\tau_{1:n} - \tau(C)\}$ converges in law to $T \sim \mathcal{N}(0, \sigma_\tau^2)$ as $n \to \infty$, where $\sigma_\tau^2 = 16\mathbb{V}\{H(X, Y) - \bar{H}(X, Y)\}$ and $\bar{H}(x, y) = \mathbb{P}(X > x, Y > y)$ is the survival function of (X, Y). This result can be deduced from the theory of U-statistics; see Lee (1990).

3 The test statistics and the computation of p-values

In what follows, consider a series of independent observations (X_1, Y_1), …, (X_n, Y_n), where $(X_i, Y_i) \sim H_i$. It is also supposed that the marginal distributions are stable, *i.e.* $X_i \sim F$ and $Y_i \sim G$ for all $i \in \{1, \ldots, n\}$. Although these requirements are often not met in practice, it is recommended to work with the residuals of models adjusted for the marginals. The idea is detailed in Section 5.

Denote by C_i the underlying copula of H_i, that is

$$C_i(u, v) = H_i\{F^{-1}(u), G^{-1}(v)\}.$$

If the alternative hypothesis $\mathcal{H}_1$ holds, then there exists a $\mathcal{B} \in \{1, \ldots, n-1\}$ and two copulas D_1, D_2 such that $C_i = D_1$ for $i \leq \mathcal{B}$ and $C_i = D_2$ for $i > \mathcal{B}$. As a consequence, $\tau(C_1) = \cdots = \tau(C_\mathcal{B}) \neq \tau(C_{\mathcal{B}+1}) = \cdots = \tau(C_n)$ whenever $\tau(D_1) \neq \tau(D_2)$. Since the empirical version of Kendall's tau based on i.i.d. observations is unbiased for its population version, one has $\mathbb{E}(\tau_{1:\mathcal{B}}) = \tau(D_1)$ and $\mathbb{E}(\tau_{\mathcal{B}+1:n}) = \tau(D_2)$. Hence, in the case when $\mathcal{B}$ is known and as long as $\tau(D_1) \neq \tau(D_2)$, a consistent test would be based on a suitably standardized version of $\tau_{1:\mathcal{B}} - \tau_{\mathcal{B}+1:n}$.

In practice, the true location of the change is generally unknown. In order to test for a change-point in the dependence structure at an unknown location, an idea is to consider all possible differences of empirical Kendall's tau computed for the sets $\{1, \ldots, \mathcal{B}\} \cup \{\mathcal{B}+1, \ldots, n\}$, where $\mathcal{B} \in \{2, \ldots, n-2\}$. To this end, let $\lambda_{n,\mathcal{B}} = \mathcal{B}/n$ and define the empirical process

$$(2) \quad \mathbb{K}_n^{\mathcal{B}} = \lambda_{n,\mathcal{B}}(1-\lambda_{n,\mathcal{B}})\sqrt{n}\,(\tau_{1:\mathcal{B}} - \tau_{\mathcal{B}+1:n})$$

indexed by $\mathcal{B} \in \{2, \ldots, n-2\}$. Note that the factor $\lambda_{n,\mathcal{B}}(1-\lambda_{n,\mathcal{B}})\sqrt{n}$ in the above definition is chosen so that $\mathbb{K}_n^{\mathcal{B}}$ converges in distribution. Indeed, one can show that $\mathbb{K}_n^{\mathcal{B}}$ converges in law to $\mathcal{N}(0, \lambda(1-\lambda)\sigma_\tau^2)$ under $\mathcal{H}_0$, where $\mathcal{B}$ depends on n in such a way that $\lambda = \lim_{n\to\infty} \lambda_{n,\mathcal{B}} \in (0,1)$. Natural statistics based on L_1–, L_2–, and L_∞–distances of the vector $\mathbb{K}_n = (\mathbb{K}_n^2, \ldots, \mathbb{K}_n^{n-2})$ are

$$(3) \quad K_{n1} = \frac{1}{n}\sum_{\mathcal{B}=2}^{n-2} \left|\mathbb{K}_n^{\mathcal{B}}\right|, \quad K_{n2} = \frac{1}{n}\sum_{\mathcal{B}=2}^{n-2} \left|\mathbb{K}_n^{\mathcal{B}}\right|^2 \quad \text{and} \quad K_{n3} = \max_{2\le\mathcal{B}\le n-2} \left|\mathbb{K}_n^{\mathcal{B}}\right|.$$

Their asymptotic behavior is described in the next proposition, whose proof is available upon request. Here and in the sequel, $\rightsquigarrow$ means *convergence in distribution.*

Proposition 1 *Under* $\mathcal{H}_0$,

$$K_{n1} \rightsquigarrow \Upsilon_C \int_0^1 |\mathcal{G}(t)|\,\mathrm{d}t, \quad K_{n2} \rightsquigarrow \Upsilon_C^2 \int_0^1 |\mathcal{G}(t)|^2\,\mathrm{d}t \quad \textit{and} \quad K_{n3} \rightsquigarrow \Upsilon_C \sup_{t\in[0,1]} |\mathcal{G}(t)|,$$

where Υ_C *depends on the common copula of* $(X_1, Y_1), \ldots, (X_n, Y_n)$ *under* $\mathcal{H}_0$ *and* $\mathcal{G}$ *is a Brownian bridge, i.e. a continuous and centered Gaussian process defined on* $[0,1]$ *with* $\mathcal{G}(0) = \mathcal{G}(1) = 0$ *and* $\mathbb{E}\{\mathcal{G}(s)\mathcal{G}(t)\} = \min(s,t) - st$.

In view of Proposition 1, the asymptotic distributions of K_{n1}, K_{n2} and K_{n3} depend on the unknown underlying copula C of the population under $\mathcal{H}_0$. This causes an obvious problem when trying to compute the p-values of the tests. A solution would be to use permutations, as in Horváth and Hušková (2005). This approach is not very attractive due to its computational complexity. This issue will rather be tackled from a multiplier version of the process $\{\mathbb{K}_n^{\mathcal{B}}, 2 \le \mathcal{B} \le n-2\}$. To this end, note that $\mathbb{K}_n^{\mathcal{B}}$ can be re-written, under $\mathcal{H}_0$, as

$$\mathbb{K}_n^{\mathcal{B}} = \sqrt{\lambda_{\mathcal{B},n}}(1-\lambda_{\mathcal{B},n})\,T_{1:\mathcal{B}} - \lambda_{\mathcal{B},n}\sqrt{1-\lambda_{\mathcal{B},n}}\,T_{\mathcal{B}+1:n},$$

where $T_{I:J} = \sqrt{J-I+1}\{\tau_{I:J} - \tau(C)\}$. From this representation, the multiplier bootstrap (see van der Vaart and Wellner, 1996; Kosorok, 2008) versions of $\mathbb{K}_n^{\mathcal{B}}$ can be defined, for $h \in \{1, \ldots, M\}$, by

$$\widehat{\mathbb{K}}^{\mathcal{B},(h)} = \sqrt{\lambda_{\mathcal{B},n}}(1-\lambda_{\mathcal{B},n})\,\widehat{T}_{1:\mathcal{B}}^{(h)} - \lambda_{\mathcal{B},n}\sqrt{1-\lambda_{\mathcal{B},n}}\,\widehat{T}_{\mathcal{B}+1:n}^{(h)},$$

where

$$\widehat{T}_{I:J}^{(h)} = \frac{4}{(J-I+1)^{3/2}} \sum_{i,j\in[I,J]} \left(\frac{g_i^{(h)}}{\bar{g}_{I:J}^{(h)}} - 1\right) Q_{ij}, \quad \text{with} \quad \bar{g}_{I:J}^{(h)} = \frac{g_I^{(h)} + \cdots + g_J^{(h)}}{J-I+1}.$$

Multiplier versions of the three test statistics are then given by

$$\widehat{K}_1^{(h)} = \frac{1}{n}\sum_{\mathcal{B}=1}^{n} \left|\widehat{\mathbb{K}}^{\mathcal{B},(h)}\right|, \quad \widehat{K}_2^{(h)} = \frac{1}{n}\sum_{\mathcal{B}=1}^{n} \left|\widehat{\mathbb{K}}^{\mathcal{B},(h)}\right|^2 \quad \text{and} \quad \widehat{K}_3^{(h)} = \max_{1\le\mathcal{B}\le n} \left|\widehat{\mathbb{K}}^{\mathcal{B},(h)}\right|.$$

The validity of the method under $\mathcal{H}_0$ is stated in the next proposition. Its proof is available upon request.

Proposition 2 *For each $j \in \{1,2,3\}$, one has*

$$\left(K_{nj}, \widehat{K}_j^{(1)}, \ldots, \widehat{K}_j^{(M)}\right) \rightsquigarrow \left(K_j, \tilde{K}_j^{(1)}, \ldots, \tilde{K}_j^{(M)}\right)$$

under $\mathcal{H}_0$, where $\tilde{K}_j^{(1)}, \ldots, \tilde{K}_j^{(M)}$ are independent copies of K_j.

For each $j \in \{1,2,3\}$, let

$$L_{M,j}(t) = \frac{1}{M}\sum_{h=1}^{M} \mathbb{I}\left(\widehat{K}_j^{(h)} \leq t\right).$$

Proposition 2 ensures that under the null hypothesis,

$$p_{M,1} = 1 - L_{M,1}(K_{n1}), \quad p_{M,2} = 1 - L_{M,2}(K_{n2}) \quad \text{and} \quad p_{M,3} = 1 - L_{M,3}(K_{n3})$$

are asymptotically valid p-values, as $\min(M,n) \to \infty$, for the tests based on K_{n1}, K_{n2} and K_{n3}.

4 Illustrations on climatic data sets

We illustrate the proposed tests statistic with climatic data, more precisely with long-run simulations issued from climate models. Hereafter, we first provide a general description of the data and then develop two applications.

4.1 General description of the data

According to IPCC (2007, Chapter 8): "*Climate models are based on well-established physical principles and have been demonstrated to reproduce observed features of recent climate (...) and past climate changes (...). There is considerable confidence that [these models] provide credible quantitative estimates of future climate change, particularly at continental and larger scales*". However, "*confidence in the changes projected by global models decreases at smaller scales*".

To overcome this problem, climatologists have developed regional climate models in order to be able to study climate change at both regional and local scales. A regional climate model is a high-resolution (e.g. 45 x 45 km) limited-area model nested in a low-resolution (typically of 300 x 300 km) global model over the region of interest, e.g. North America. Large-scale information from the global model is transferred to the regional model by forcing its lateral boundaries with the values of the global model; see Caya et al. (1995).

The data sets used in Sections 4.2–4.3 result from simulated observations with respect to a regional climate model. More specifically, the simulated data were performed with the Canadian Regional Climate Model (CRCM4.2.3; Caya and Laprise, 1999; Music and Caya, 2007). The CRCM data has been generated and supplied by *Ouranos, consortium on regional climatology and adaptation to climate change.* The runs were driven by atmospheric fields taken from simulation output of the third generation Canadian Coupled Global Climate Model (CCGCM3; Flato and Boer, 2001; McFarlane et al., 2005; Scinocca et al., 2008).

The domain of simulation covers North-America (AMNO, 201 x 193 grid points) with a horizontal grid-sized mesh of 45 km (true at 60 N). A spectral nudging technique was applied to large-scale winds (Riette and Caya, 2002) within the interior of the regional domain to keep CRCM's large scale flow close to its driving data.

To take into account a potential variability in the global climate model, the initial conditions are slightly modified. In the terminology used in climatology, different *members* of the simulation are thus obtained. Both global and regional simulations were performed according to IPCC "observed 20th century" scenario for years 1961-2000 and scenario A2 for greenhouse gas and aerosol projected evolution for years 2001-2100 (Nakicenovic et al., 2000).

In climatic projections, many variables are simulated in order to model most physical processes. In the applications presented below, we focus on precipitation and runoff for small watersheds located in the northern part of the province of Québec, namely Pyrite and Arnaud. It is important to note that runoff is obtained by putting in adequacy the atmospheric balance and the hydrologic balance (Peixoto and Oort, 1992), without making use of a rainfall-runoff model.

The two applications that are detailed in Sections 4.2–4.3 highlight the importance of stabilizing the margins before testing for a change in the dependence structure. For the two examples, various kinds of changes in the mean are apparent. Hence, the following steps were accomplished:

(S_1) Hypothesis testing based on the statistics K_{n1}, K_{n2} and K_{n3} computed from the original sample $(X_1, Y_1), \ldots, (X_n, Y_n)$;

(S_2) Test for a smooth change in the mean of the two marginal series based on the following model

$$(4) \quad \theta_i = \begin{cases} \mu_1, & 1 \le i \le \mathcal{K}_1; \\ \left(\dfrac{\mathcal{K}_2 - i}{\mathcal{K}_2 - \mathcal{K}_1} \right) \mu_1 + \left(\dfrac{i - \mathcal{K}_1}{\mathcal{K}_2 - \mathcal{K}_1} \right) \mu_2, & \mathcal{K}_1 < i < \mathcal{K}_2; \\ \mu_2, & \mathcal{K}_2 \le i \le n\,. \end{cases}$$

(S_3) In the case of a significant marginal change-points identified in step (ii), estimate $(\widehat{\mathcal{K}}_1, \widehat{\mathcal{K}}_2)$ as advocated by Lombard (1987) via

$$(5) \quad (\widehat{\mathcal{K}}_1, \widehat{\mathcal{K}}_2) = \arg \max_{1 \le \mathcal{K}_1 < \mathcal{K}_2 \le n} \left\{ \left| \sum_{j=\mathcal{K}_1+1}^{\mathcal{K}_2} \sum_{\ell=1}^{j} \left\{ \frac{\phi(X_\ell) - \bar{\phi}}{\sigma_\phi} \right\} \right| \Big/ \sigma \left(\frac{\mathcal{K}_1}{n}, \frac{\mathcal{K}_2}{n} \right) \right\},$$

where $\phi(u) = 2u - 1$ is Wilcoxon's score function, $\bar{\phi}$, σ_ϕ are the mean and standard deviation of $\phi\{1/(n+1)\}, \ldots, \phi\{n/(n+1)\}$, and

$$\sigma^2(u, v) = \frac{(1-u)^3(1+3u)}{12} - \frac{(1-v)^3(1+3v)}{12} - \frac{(1-v)^2(v^2-u^2)}{2}\,.$$

and stabilize the mean based on the least-square estimators $(\hat{\mu}_1, \hat{\mu}_2)$ (see Quessy et al., 2011, equ. 6 and 7);

(S_4) Perform the three tests for the detection of change-points in the dependence based on the stabilized pseudo-sample $(X_{1,n}, Y_{1,n}), \ldots, (X_{n,n}, Y_{n,n})$.

Step S_1 has been included in order to illustrate the importance of stabilizing the marginal series. Step S_2 uses a Cramér–von Mises statistics proposed by Lombard (1987) for the detection of smooth changes in the mean. As shown by Quessy et al. (2011) in their extensive simulation study, this test statistic performs very well to identify many kinds of change-points, including the general smooth-change pattern, as well as abrupt and onset of trend changes.

4.2 Application 1: testing change in dependence between two members

The first dataset concerns a bivariate time series representing the simulated mean annual precipitation driven by two different members of the MCCG3.1 for the Pyrite watershed. The latter are called respectively `aet` and `aev`. The simulation period ranges from 1961 to 2099. The three tests for the detection of a change in the dependence structure were first performed on the original series shown in Figure 1. The estimated p-values, based on $M = 10\ 000$ multiplier samples, were $(p_{M,1}, p_{M,2}, p_{M,3}) = (0.0020, 0.0031, 0.0210)$. Hence, the three tests concluded to a significant change-point in the dependence structure. Based on the estimator

$$(6) \quad \widehat{\mathcal{B}} = \arg \max_{2 \le \mathcal{B} \le n-2} \left| \mathbb{K}_n^{\mathcal{B}} \right|,$$

implicit in the definition of K_{n3} this change happened in 2004.

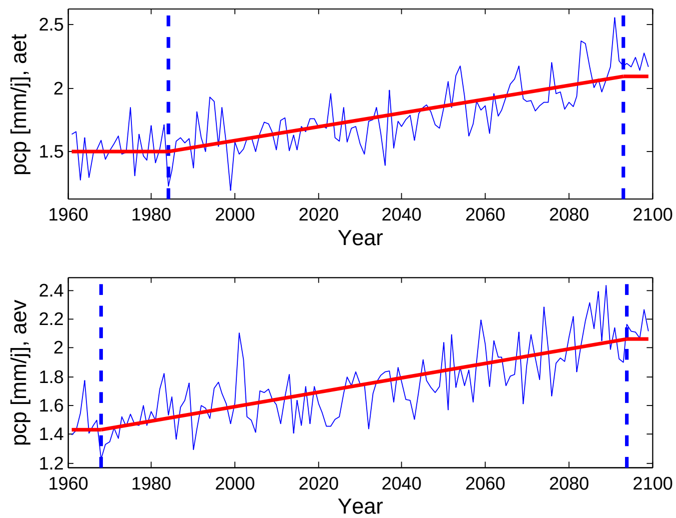

Figure 1: Time series of the mean annual precipitation (simulations `aet` and `aev`) for the Pyrite watershed for the period 1961–2099

In view of Figure 1, however, the conclusion of the tests may have been influenced by the occurrence of change-points in the marginal series. In order to remove these undesirable effects, Lombard's test was performed. For `aet`, the latter concluded to a significant change-point estimated, via Equation (5), by $(\widehat{\mathcal{K}}_1, \widehat{\mathcal{K}}_2) = (1983, 2093)$. As the second change-point is located near the upper boundary of the time series, this is an example of the so-called onset of trend model. Lombard's test also detected a significant change-point in `aev` at $(\widehat{\mathcal{K}}_1, \widehat{\mathcal{K}}_2) = (1968, 2094)$. Here again, the change corresponds to an onset of trend pattern. The stabilized series based on the residuals of the smooth-change model (4) are illustrated in Figure 2.

This time, the tests based on K_{n1}, K_{n2} and K_{n3} are no more significant when computed from the stabilized series. Indeed, the estimated p-values were greater than 0.97 for the three tests. Hence, the change-point detected on the original series was only an artefact induced by the unstability of the margins with respect to their mean.

In figure 3, one can see the value of $\mathbb{K}_n^{\mathcal{B}}$ as a function of $\mathcal{B} \in \{2, \dots, n-2\}$ both for the original

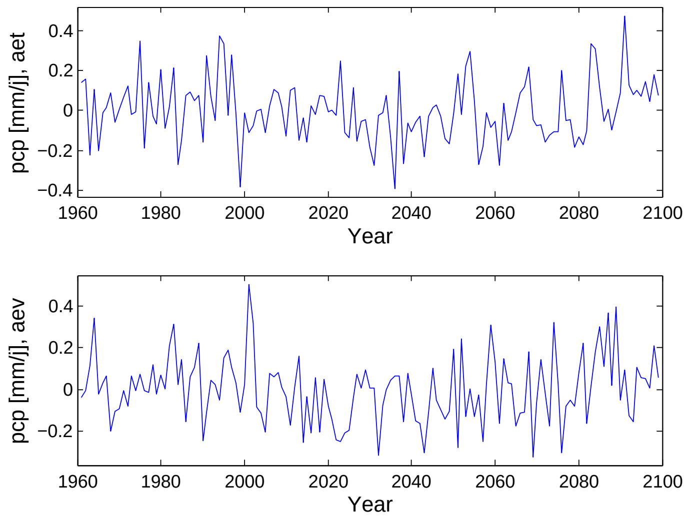

Figure 2: Stabilized time series of the mean annual precipitation (simulations `aet` and `aev`) for the Pyrite watershed for the period 1961–2099

and the stabilized series. This figure relies to statistic K_{n3} given by Equation (3). The lack of stability in margins, as shown in Figure 1, yields to a global maximum in the corresponding series $\mathbb{K}_n^{\mathcal{B}}$ (in blue) and also several "local maxima" (values of $\mathbb{K}_n^{\mathcal{B}}$ greater than the estimated critical value). However, after stabilizing the margins (see Figure 2), all values of $\mathbb{K}_n^{\mathcal{B}}$ (in red) are smaller than the estimated critical value.

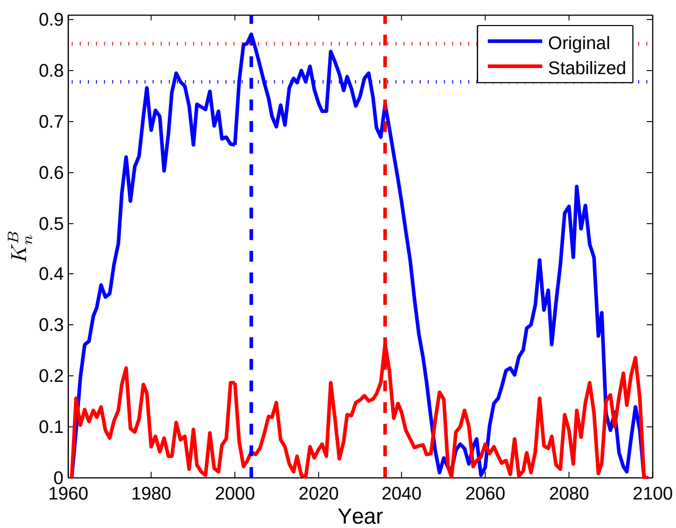

Figure 3: Values of $\mathbb{K}_n^{\mathcal{B}}$ as a function of $\mathcal{B}$ computed from the original series (in blue) and from the stabilized series (in red); the horizontal dotted lines correspond to the estimated critical values while vertical lines identify the detected change-point in dependence

In the former example, we illustrate how the lack of stability in margins may drastically affect the results of tests based on K_{n1}, K_{n2} and K_{n3} and yield to the wrong conclusion that there exists a change-point in dependence. In the next example, we show the effect of stabilization on the estimated change-point $\widehat{\mathcal{B}}$ defined by (6) when a change in dependence really occurred.

4.3 Application 2: testing change in dependence between two variables

The second example relates to the mean annual precipitation and runoff simulated from the Canadian Regional Climate Model (CRCM) from 1961 to 2099 for the Arnaud watershed. The corresponding bivariate time series, involving two variables, is presented in Figure 4.

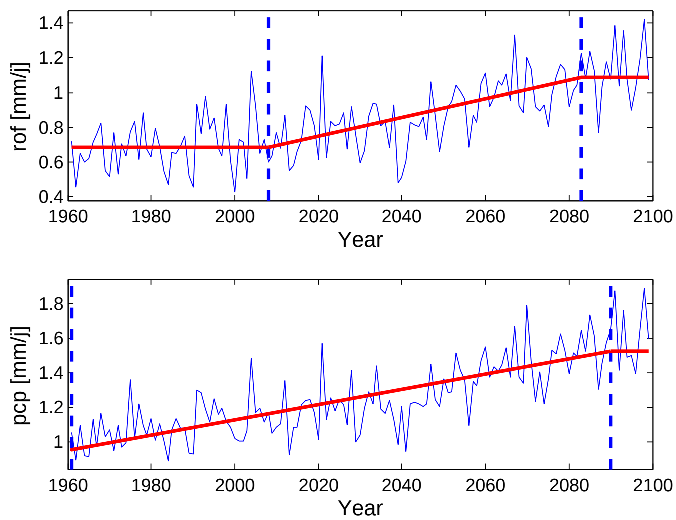

Figure 4: Time series of the mean annual runoff and precipitation (simulation `aet`) for the Arnaud watershed for the period 1961–2099

The three tests of change-point detection in the dependence yield estimated p-values inferior to 0.0001, indicating a clear change-point in the dependence structure. The location of the change, estimated via (6), is identified in 2011. Since the two univariate series have a clear tendency to increase following a smooth-change pattern, these conclusions could be erroneous, however. Indeed, Lombard's test detected significant marginal change-points at $(\widehat{\mathcal{K}}_1, \widehat{\mathcal{K}}_2) = (2008, 2083)$ for the runoff variable. For the precipitation variable, significant change-points are located at $(\widehat{\mathcal{K}}_1, \widehat{\mathcal{K}}_2) = (1961, 2090)$. Figure 5 shows the stabilized time series based on the residuals of the smooth-change model (4). The estimated p-values of the tests for the detection of change-point in the dependence based on this transformed series are $(p_{M,1}, p_{M,2}, p_{M,3}) = (0.0003, 0.0005, 0.0088)$. Hence, the conclusion of a significant change-point still holds, but the year when it occurred is now estimated at $\widehat{\mathcal{B}} = 2038$. Kendall'tau before the change-point in dependence is $\tau_{1:\widehat{\mathcal{B}}} = 0.363$ while this measure of association almost doubles after year 2038 with $\tau_{\widehat{\mathcal{B}}+1:n} = 0.681$.

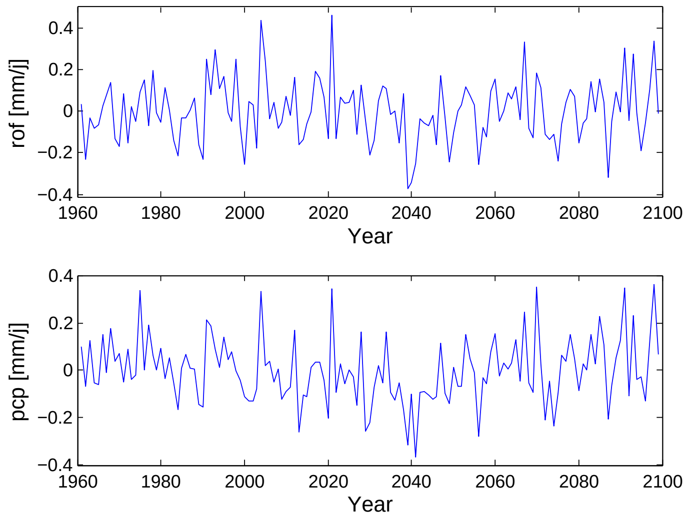

Figure 5: Stabilized time series of the mean annual runoff and precipitation (simulation `aet`) for the Arnaud watershed for the period 1961–2099

References

Antoch, J. and Hušková, M. (2001). Permutation tests in change point analysis. *Statistics & Probability Letters*, 53(1):37–46.

Antoch, J., Hušková, M., Janic, A., and Ledwina, T. (2008). Data driven rank test for the change point problem. *Metrika*, 68(1):1–15.

Brodsky, B. and Darkhovsky, B. (2005). Asymptotically optimal methods of change-point detection for composite hypotheses. *Journal of Statistical Planning and Inference*, 133(1):123–138.

Caya, D. and Laprise, R. (1999). A semi-implicit semi-Lagrangian regional climate model: The Canadian RCM. *Monthly Weather Review*, 127(3):341–362.

Caya, D., Laprise, R., Gigure, G., Bergeron, G., Blanchet, J. P., Stocks, B. J., Boer, G. J., and McFarlane, N. A. (1995). Description of the Canadian regional climate model. *Water, Air, & Soil Pollution*, 82(1):477–482.

Csörgő, M. and Horváth, L. (1988). Invariance principles for changepoint problems. *Journal of Multivariate Analysis*, 27(1):151–168.

Dias, A. and Embrechts, P. (2009). Testing for structural changes in exchange rates dependence beyond linear correlation. *The European Journal of Finance*, 15(7):619–637.

Flato, G. M. and Boer, G. J. (2001). Warming asymmetry in climate change simulations. *Geophysical Research Letters*, 28(1):195–198.

Gombay, E. and Horváth, L. (1995). An application of U-statistics to change-point analysis. *Acta Scientiarum Mathematicarum (Szeged)*, 60(1-2):345–357.

Gombay, E. and Horváth, L. (1997). An application of the likelihood method to change-point detection. *Environmetrics*, 8(5):459–467.

Horváth, L. and Hušková, M. (2005). Testing for changes using permutations of U-statistics. *Journal of Statistical Planning and Inference*, 128(2):351–371.

IPCC (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Jandhyala, V. K., Liu, P., and Fotopoulos, S. B. (2009). River stream flows in the northern Québec Labrador region: A multivariate change point analysis via maximum likelihood. *Water Resources Research*, 45. W02408, doi:10.1029/2007WR006499.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York, NY, USA.

Lee, A. J. (1990). *U-statistics - Theory and practice*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York, NY, USA.

Lombard, F. (1987). Rank tests for changepoint problems. *Biometrika*, 74(3):615–624.

McFarlane, N., Scinocca, J. F., Lazare, M., Harvey, R., Verseghy, D., and Li, J. (2005). The CCCma third generation atmospheric general circulation model. Cccma internal report. http://www.cccma.ec.gc.ca/papers/jscinocca/AGCM3_report.pdf. Last visited July, 2010, 25 pp.

Music, B. and Caya, D. (2007). Evaluation of the hydrological cycle over the Mississippi River basin as simulated by the Canadian Regional Climate Model (CRCM). *Journal of Hydrometeorology*, 8(5):969–988.

Nakicenovic, N., Swart, S., and al. (2000). *IPCC special report on emissions scenarios: a special report of Working Group III of the IPCC*. Cambridge University Press, Cambridge, UK. http://www.ipcc.ch/ipccreports/sres/emission/index.php?idp=0. Last visited July, 2010.

Peixoto, J. and Oort, A. (1992). *Physics of climate*. Springer-Verlag, New York, NY, USA.

Quessy, J.-F., Favre, A.-C., Saïd, M., and Champagne, M. (2011). Statistical inference in Lombard's smooth-change model. *Environmetrics*. DOI:10.1002/env.1108, in press.

Riette, S. and Caya, D. (2002). Sensitivity of short simulations to the various parameters in the new CRCM spectral nudging. Research activities in Atmospheric and Oceanic Modelling. Edited by H. Ritchie, WMO/TD - No 1105, Report No. 32: 7.39-7.40.

Scinocca, J. F., McFarlane, N. A., Lazare, M., Li, J., and Plummer, D. (2008). The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmospheric Chemistry and Physics Discussions*, 8(2):7883–7930.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, 8:229–231.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes – With applications to statistics*. Springer Series in Statistics. Springer-Verlag, New York, NY, USA.

Zou, C., Liu, Y., Qin, P., and Wang, Z. (2007). Empirical likelihood ratio test for the change-point problem. *Statistics and Probability Letters*, 77(4):374–382.