



Intercomparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change point models

Claudie Beaulieu,^{1,2} Ousmane Seidou,³ Taha B. M. J. Ouarda,¹ and Xuebin Zhang⁴

Received 2 October 2008; revised 17 May 2009; accepted 22 May 2009; published 11 August 2009.

[1] In this paper, two new Bayesian change point techniques are described and compared to eight other techniques presented in previous work to detect inhomogeneities in climatic series. An inhomogeneity can be defined as a change point (a time point in a series such that the observations have a different distribution before and after this time) in the data series induced from changes in measurement conditions at a given station. It is important to be able to detect and correct an inhomogeneity, as it can interfere with the real climate change signal. The first technique is a Bayesian method of multiple change point detection in a multiple linear regression. The second one allows the detection of a single change point in a multiple linear regression. These two techniques have never been used for homogenization purposes. The ability of the two techniques to discriminate homogeneous and inhomogeneous series was evaluated using simulated data series. Various sets of synthetic series (homogeneous, with a single shift, and with multiple shifts) representing the typical total annual precipitation observed in the southern and central parts of the province of Quebec, Canada, and nearby areas were generated for the purpose of this study. The two techniques gave small false detection rates on the homogeneous series. Furthermore, the two techniques proved to be efficient for the detection of a single shift in a series. For the series with multiple shifts, the Bayesian method of multiple change point detection performed better. An application to a real data set is also provided and validated with the available metadata.

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

1. Introduction

[2] The increasing interest in climate modeling has triggered the development and testing of more performing homogenization techniques. These techniques are intended to identify and sometimes correct inhomogeneities in data series that do not reflect the real climate variations. An inhomogeneity can be defined as a change point (a time point in a series such that the observations have a different distribution before and after this time) in the data series induced from changes in measurement conditions at a given station. These changes include relocations, instrument replacements, changes in observation procedures or modifications in the immediate environment of the site. There is a difference between an inhomogeneous series and a series that is nonstationary, as a series can be nonstationary and

homogeneous at the same time (the change points are only caused by real climatic variations). The user of climatic data series is often unaware of the presence or absence of inhomogeneities in the series, which is why the homogenization techniques are needed. The inhomogeneities can interfere with the real climate change signal and lead to poor climatic or hydrological model calibration or biased studies of climate trends and variability. Therefore, the detection and correction of these inhomogeneities is important before undertaking any kind of climate analysis.

[3] Several techniques have been developed for the detection of inhomogeneities in climate series. For a complete review, the reader is referred to *Peterson et al.* [1998], *World Meteorological Organization* [2003], *Beaulieu et al.* [2007], and *Reeves et al.* [2007]. Most of the homogenization techniques are addressed in classical or Bayesian statistical frameworks, supported by parametric or nonparametric models. The change point model often represents a shift in the mean of the base series (the series that is tested for a potential inhomogeneity) and that does not occur in the neighbor series. Some techniques allow to identify the position of the shift and to determine its significance using a procedure based on the likelihood ratio test [*Maronna and Yohai*, 1978; *Potter*, 1981; *Alexandersson*, 1986]. The change point analysis can also be performed with a regression model which represents a change in the mean and/or in

¹Centre Eau, Terre et Environnement, Institut National de la Recherche Scientifique, University of Quebec, Quebec, Quebec, Canada.

²Now at Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey, USA.

³Department of Civil Engineering, University of Ottawa, Ottawa, Ontario, Canada.

⁴Climate Research Division, Science and Technology Branch, Environment Canada, Downsview, Ontario, Canada.

the trend of the series [Easterling and Peterson, 1995; Vincent, 1998; Lund and Reeves, 2002; Wang, 2003]. The neighbor series are used to identify shifts in the base series that are not reproduced in the immediate neighborhood. These shifts are assumed to be artificially caused and not representative of the regional climatic signal. Hence, the neighbor series are climatically similar to the base series and must be homogeneous. Otherwise, an inhomogeneity in a neighbor series could be allocated to the base series. However, there are a few techniques developed for isolated stations that do not rely on neighbor series. For more details on these techniques, the reader is referred to Peterson et al. [1998].

[4] Most homogenization techniques were developed for data series which contain at most one change point. In general, for the detection of multiple changes, a segmentation procedure is applied. The drawback with the segmentation approach is that the same test applied several times on the same observations can increase the risk of false detection. Furthermore, the size of the segments is reduced during the analysis, each time diminishing the detection power of the method. To avoid these inconveniences, a few techniques were developed to detect multiple shifts in several series [Szentimrey, 1999; Caussinus and Mestre, 2004]. The series are mutually compared to each other by switching the role (neighbor or base) of the series at each step of the procedure. The technique proposed by Szentimrey [1999] allows the inference on the number of shifts by minimizing the variance of the difference series. The technique of Caussinus and Mestre [2004] allows the determination of the number of change points with a criterion developed by Caussinus and Lyazrhi [1997]. These techniques are suitable for cases with several neighbor series, but they are less efficient with few series (less than six) [Slonosky et al., 1999].

[5] Several change point detection techniques have been presented in the literature. For a complete review of the different approaches, the reader is referred to Chen and Gupta [2001]. Bayesian linear regression models were addressed among others by Bacon and Watts [1971], Smith and Cook [1980], Holbert [1982], Solow [1988], and Rasmussen [2001]. Seidou et al. [2007] presented a technique which has a general formulation and that allows testing the hypothesis of no change. In the case of multiple shifts, Bayesian techniques that infer the positions of the changes when the number of shifts is known were proposed by Stephens [1994]. Barry and Hartigan [1992, 1993] developed a technique to detect multiple change points with a product partition models-based algorithm. Green [1995] proposed the use of reversible jump Markov chain Monte Carlo (MCMC) methods for multiple change point analysis. Chib [1998] and Lavielle and Lebarbier [2001] used a new formulation for the change point problem and proposed to solve it with MCMC methods. Seidou and Ouarda [2007] applied the algorithm on the basis of recursions proposed by Fearnhead [2006], to detect an unknown number of shifts in a multiple linear regression model. Change point linear regression models can be used to detect inhomogeneities in climatic series by setting the base series as the dependent variable and the neighbor series as the independent variables.

[6] Few Bayesian change point techniques have been specifically applied to homogenization problems. The technique proposed by Lee and Heghinian [1977] and extended by Perreault et al. [1999, 2000] was applied in comparison studies to detect inhomogeneities in synthetic temperature series. This technique performed well to detect a single shift, but tends to combine the sequential shifts in the presence of multiple inhomogeneities [Ducré-Robitaille et al., 2003; DeGaetano, 2006]. Beaulieu et al. [2008] applied the Bayesian technique proposed by Rasmussen [2001] to identify inhomogeneities in synthetic precipitation series. This technique performed well to detect a single or multiple shifts, but lead to a very high percentage of false detections when the series are homogeneous. Bayesian techniques that were compared in simulation studies, such as Ducré-Robitaille et al. [2003], DeGaetano [2006], and Beaulieu et al. [2008], did not lead to better results than popular techniques such as the standard normal homogeneity test [Alexandersson, 1986] or the bivariate test [Potter, 1981].

[7] The objectives of this work are to introduce two new Bayesian techniques [Seidou and Ouarda, 2007; Seidou et al., 2007] for the homogenization of climate series and to verify their capacity to discriminate homogeneous and inhomogeneous (with a shift in the mean) series having the statistical characteristics of precipitation data series observed in the southern and central parts of the province of Quebec and nearby areas, Canada. These two techniques were applied to various types of synthetic precipitation series: a set of homogeneous series, a set of series with a single shift and a set of series with multiple shifts, to evaluate their performances. For each type of synthetic series (homogeneous, one change point, two change points or three change points), the performances of these two new Bayesian techniques were compared to those of the best of eight techniques reviewed by Beaulieu et al. [2008]. These two methods were not included in the comprehensive intercomparison study of Beaulieu et al. [2008] because they were under development at the time the study was carried out. However, they were shown to be applicable to several practical problems that previously available methods could not handle. An application to real world data series is also presented.

[8] The paper is organized as follows: section 2 gives a detailed description of the two techniques. The methodology is described in section 3. The results of the simulation study are presented in section 4. In section 5, an application to series of total annual precipitation is presented. Finally, the discussion and conclusions are presented in section 6.

2. Theoretical Background

[9] In this paper, the Bayesian approach to change point detection is considered. Bayesian methods have a different approach from classical techniques. Through a prior distribution, they allow the integration of some knowledge (informative prior distribution) or lack of information (non-informative prior distribution) about the phenomenon being studied. Then, the information provided by the prior distribution and the information provided by the observations are combined into a posterior distribution, which is used to make inference about the parameters. Some advantages of using Bayesian analysis over classical analysis is that it

allows a formal use of nonexperimental sources of information and it provides the full posterior probability distribution for the parameter of interest [Berger, 1985]. For example, Bayesian change point methods provide the full posterior probability distribution of the position of the shifts, which can be multimodal or skewed. After specifying a loss function, an estimate of the shift's position can be obtained. This provides more information than an estimate of the shift's position and its significance level, which is obtained with the classical change point techniques.

[10] In this section, the two techniques are presented. The first technique proposed in the present work is a Bayesian linear regression model designed to detect multiple change points [Seidou and Ouarda, 2007]. Its main feature is that it allows the detection of an unknown number of shifts. The second technique proposed in the present work is a Bayesian linear regression model which allows for the detection of a single change [Seidou et al., 2007]. It improves the model presented by Rasmussen [2001], which assumes the presence of a change point and infers its position, by extending it to the case for which a change point does not occur with certainty. The two techniques are briefly described in sections 2.1 and 2.2. For more details, the reader is referred to Seidou and Ouarda [2007] and Seidou et al. [2007].

2.1. Bayesian Multiple Change Point Detection in Multiple Linear Regression (BAMS)

[11] This technique is an adaptation of Fearnhead [2006] which presents a general procedure to detect the number and positions of the change points in a data series with a nonparametric approach. The procedure was adapted by Seidou and Ouarda [2007] to multiple linear regression. The response variable is noted by y_j ($j = 1, \dots, n$), or $\mathbf{y}_{n \times 1}$ in vectorial form, while x_{ij} ($i = 1, \dots, d^*$; $j = 1, \dots, n$) represents the j th observation of the i th explanatory variable ($\mathbf{X}_{d^* \times n}$ in matricial form). There are n observations and d^* explanatory variables. For homogenization problems, the response variable is the base series, while the neighbor series are the explanatory variables. The relationship between the base series and the neighbor series can be represented by

$$y_j = \sum_{i=1}^{d^*} \theta_i x_{ij} + \varepsilon_j, \quad j = 1, \dots, n. \quad (1)$$

The response variable is normally distributed with the mean $\sum_{i=1}^{d^*} \theta_i x_{ij}$ and variance σ^2 . Given the parameter vector $\Phi = [\theta_1, \theta_2, \dots, \theta_{d^*}, \sigma]$, the density of the response variable is

$$f(y_i | \Phi) = (2\pi\sigma^2)^{-1/2} \exp \left(-\frac{\left(y_i - \sum_{j=1}^{d^*} \theta_j x_{ij} \right)^2}{2\sigma^2} \right). \quad (2)$$

Let m be the number of change points and $\tau_0 = 0, \tau_1, \dots, \tau_m, \tau_{m+1} = n$ the positions of the change points. The prior

probability distribution for the position of the change is uniform, given the preceding change point:

$$\pi(\tau_{i+1} | \tau_i) = \frac{1-p}{n-\tau_i}, \quad (3)$$

where p represents the prior probability for no change. Uniform priors are assumed for θ . Therefore, the prior distribution of Φ depends only on σ ($\pi(\Phi) = \pi(\theta_1, \dots, \theta_{d^*}, \sigma) \propto \pi(\sigma)$). It is assumed that the prior distribution of Φ has this particular form:

$$\pi(\Phi) = \pi(\sigma) = p(\sigma | a, c) = \frac{\sigma^{-a} \exp\left(-\frac{c}{2\sigma^2}\right)}{2^{\frac{a-3}{2}} c^{-\frac{a-1}{2}} \Gamma\left(\frac{a-1}{2}\right)}, \quad a > 1, c > 0, \quad (4)$$

where a and c are the hyperparameters and Γ represents the Gamma function. The positions of the change points are inferred on the basis of the coherence $P(t, s)$ of all segments $[t, s]$, where $P(t, s)$, $s \geq t$ is the probability that t and s (representing the lower and upper limits of a given segment) are in the same segment and where t varies from 1 to n and s varies from t to n . The posterior probability that t and s ($s \geq t$) are in the same segment is given by

$$P(t, s) = 2^{\frac{d^*}{2}} \pi^{\frac{d^* - t + s - 1}{2}} \frac{(\varepsilon_{s:t}^T \varepsilon_{s:t} + c)^{-\frac{(t-s+a)}{2}} \Gamma\left(\frac{t-s+a}{2}\right)}{c^{-\frac{(a-1)}{2}} |\mathbf{X}_{s:t}^T \mathbf{X}_{s:t}|^{1/2} \Gamma\left(\frac{a-1}{2}\right)}, \quad (5)$$

where $\varepsilon_{t:s}$ is a part of the vector of random errors between t and s and $\mathbf{X}_{t:s}$ represents the rows between t and s of the explanatory variables. The superscript T indicates the transpose.

[12] The hyperparameter a was chosen after a sensitivity analysis. The level of information in the prior can be controlled by changing the value of a . It was noticed, after a few trials, that with the noninformative prior $a = 1 + \varepsilon$, this technique performed well on homogeneous series with a weak percentage of false detection. When an informative prior ($a \geq 3$) was used, the technique had a high percentage of false detection on the homogeneous series and a high percentage of detection on series with real shifts. For the present study, a different approach was used. It consists in applying the technique with a noninformative prior ($a = 1.1$) to determine if the series is homogeneous or not. If the series is considered inhomogeneous, the technique is reapplied with informative priors ($a = 5$) to detect the position and magnitude of the shifts. The parameter c , representing the dimension of the variance, was set to the variance obtained by least square estimates of the linear regression equations, ($c = \varepsilon_{1:n}^T \varepsilon_{1:n} = \mathbf{y}_{1:n}^T \mathbf{y}_{1:n} - \mathbf{X}_{1:n} \mathbf{X}_{1:n}^T (\mathbf{X}_{1:n}^T \mathbf{X}_{1:n})^{-1} \mathbf{y}_{1:n}^T \mathbf{y}_{1:n}$). More details concerning the development of BAMS and the inference of the number and position of the change points are presented in Appendix A.

2.2. Bayesian Change Point in Multiple Linear Regression (BARE)

[13] This technique is presented by Seidou et al. [2007]. The model was designed to infer the position of a single change point in the parameters of a multiple linear regres-

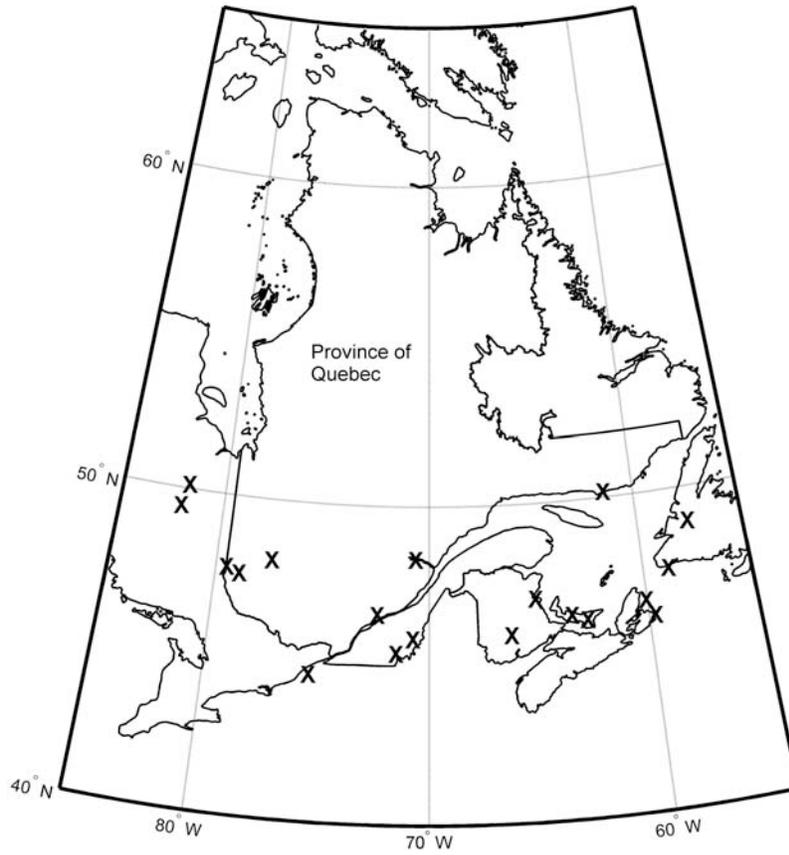


Figure 1. Locations of the stations used to generate the synthetic series.

sion equation. The model can be represented in matrix form by

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\theta}_t^{(p)} + \mathbf{v}_t, \quad (6)$$

where \mathbf{y}_t represents the response vector ($r \times 1$), \mathbf{X}_t is the matrix of explanatory variables matrix ($r \times d^*$) of the neighbor series, \mathbf{v}_t represents the vector of residuals that are independent and identically normally distributed ($N[0, \Sigma_{\mathbf{v}}]$) and $\boldsymbol{\theta}_t^{(p)}$ is the parameter vector, where

$$\boldsymbol{\theta}_t^{(p)} = \begin{cases} \boldsymbol{\beta}_1^*, & 1 \leq t < p, \\ \boldsymbol{\beta}_2^*, & p \leq t \leq n, \end{cases} \quad (7)$$

under the following constraints:

$$\boldsymbol{\beta}_1^* = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)^T \text{ and } \boldsymbol{\beta}_2^* = (\boldsymbol{\beta}_2, \boldsymbol{\beta}_0)^T. \quad (8)$$

The dimensions of the vectors $\boldsymbol{\theta}_t^{(p)}$, $\boldsymbol{\beta}_1^*$, $\boldsymbol{\beta}_2^*$, $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ are $(d^* \times 1)$, $(d^* \times 1)$, $(d^* \times 1)$, $(d_0^* \times 1)$, $(d_1^* \times 1)$ and $(d_2^* \times 1)$, respectively. The constraints imply that $d^* = d_0^* + d_1^*$. The model infers a change in the vector $\boldsymbol{\theta}_t^{(p)}$ at the position p . The vector of parameters $\boldsymbol{\beta}_0$ is assumed constant before and after the change point. The vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ contain the values of the remaining parameters before and after the change point. Finally, the vectors $\boldsymbol{\beta}_1^*$ and $\boldsymbol{\beta}_2^*$ represent the regression parameters before and after the change. Hence, the model represents a change point in the vector $\boldsymbol{\theta}_t^{(p)}$, from the subvector $\boldsymbol{\beta}_1$ to the subvector $\boldsymbol{\beta}_2$. The subvector $\boldsymbol{\beta}_0$ remains part of $\boldsymbol{\theta}_t^{(p)}$. When this model is

applied to detect inhomogeneities in a climate series, the base series is represented by \mathbf{y}_t , and \mathbf{X}_t is the matrix of the neighbor series.

[14] *Seidou et al.* [2007] considered noninformative prior distributions for the regression parameters and the variance. The prior for the change point position is a uniform distribution. The posterior distributions were obtained by Gibbs sampling [Gelfand and Smith, 1990]. The method was developed to detect at most one change point, but it can be applied for multiple change points using a segmentation approach. As the Gibbs sampling is used, the model can take into account missing observations.

3. Simulation Study

3.1. Synthetic Series

[15] The two techniques were applied to synthetic series of precipitation that were generated for previous comparative studies of several homogenization techniques. The statistical properties reproduced in these series are (1) a mean total annual precipitation of 1089 mm, (2) a standard deviation of 142 mm and (3) a lag one autocorrelation of 0.02. These values are the average characteristics of a set of selected stations located in central Quebec and surroundings having long time series with few missing data (Figure 1). For each base series, three correlated neighbor series are generated to reproduce a spatial cross correlation of 0.55. This value is the mean spatial cross correlation in the set of selected stations that are located at a distance less than 300 km. Different data sets are used here: homogeneous

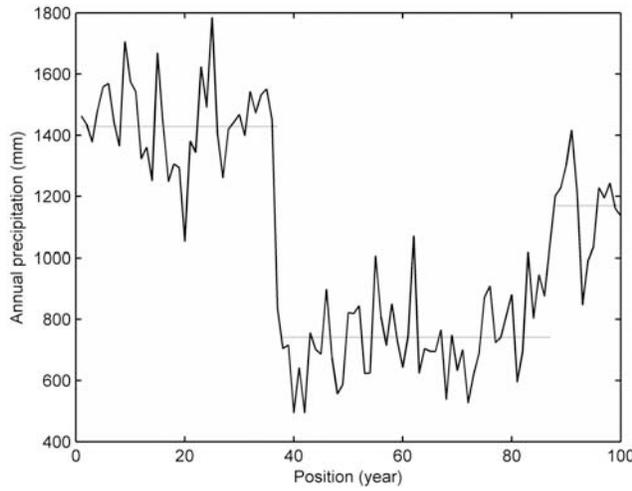


Figure 2. Example of synthetic series with two shifts (the dashed line represents the mean of each segment).

series and series with a single or multiple shifts in the mean. New synthetic series were also generated for the purpose of this work to verify the performance of the two techniques in the presence of autocorrelation: homogeneous series with a lag one autocorrelation of 0.2 and 0.4 (according to the set of selected stations, an autocorrelation higher than 0.4 for total annual precipitation data is unlikely to be observed). A total number of 15,000 homogeneous series, 25,000 series with a single shift, 15,000 series with two shifts and 15,000 series with three shifts was generated. All the synthetic series are 100 years long. The magnitude and positions of the shifts were generated randomly to represent as many configurations as possible. The detailed generation scheme of the synthetic series is presented in *Beaulieu et al.* [2008]. An example of series with two shifts is presented in Figure 2.

3.2. Application of the Techniques

[16] The two techniques were applied to the synthetic series. The mode of the posterior distributions was used as the Bayes parametric estimator for the presence and positions of the shifts. In practice, it is common to apply sequentially a technique developed for a single shift, to be able to detect multiple shifts. This procedure was used to detect multiple shifts in the comparative studies of *Ducré-Robitaille et al.* [2003] and *Beaulieu et al.* [2008]. To detect multiple shifts with BARE, this procedure was used. When a shift is detected, the series is divided into two segments, the technique is reapplied to each segment, and this procedure is repeated until all segments of the series are considered homogeneous or too short to be tested. For both techniques, the magnitudes were estimated by computing the difference of the means of the segments preceding and following a shift. The estimated magnitudes are then compared to the real magnitudes introduced in the synthetic series. A prior probability of change of 0.5 was used with both techniques, to be able to compare the results with classical techniques. It can be considered as noninformative as it represents the case for which the user has no information concerning the presence or absence of a shift in the series. The minimal length between two consecutive shifts was set to 10 observations and the shifts identified in the first ten or last ten observations in the series were

ignored to make the results comparable to those obtained by *Beaulieu et al.* [2008].

3.3. Evaluation of the Detection Skills

[17] The detection skills of the techniques were assessed on all the synthetic series. Once again, the detection skills are evaluated using the same thresholds that were used by *Beaulieu et al.* [2008] to have comparable results. For all types of synthetic series, the number of shifts detected is compared to the real number of shifts without taking into account the position or the magnitude of the shifts detected.

[18] For the homogeneous series (with and without autocorrelation), the performance is verified by computing the false detection (type I error) rate. It is the percentage of cases for which the null hypothesis is rejected (homogeneity) when it is true.

[19] For the series with a single shift, the errors in position and magnitude were computed (real position/magnitude minus estimated position/magnitude). The absolute errors were also calculated. When a shift is not detected, the position error and the absolute position error are fixed to the length of the series (100) and the magnitude error and the absolute magnitude error are fixed to 3 standard deviations (highest generated magnitude). The numbers of correctly identified, well-identified and well-positioned shifts were also computed. A shift is correctly identified when the estimated position is exact and the relative difference between the estimated magnitude and the real magnitude is less than 20% of the real magnitude. A shift is considered well identified when it is located within ± 2 years of the true location and the absolute error on the estimation of the magnitude is lower or equal to 50% of the real magnitude. A shift is well positioned when it is located within ± 2 years of the true position without any measure of accuracy for the magnitude.

[20] For the series with multiple shifts, the ability of the techniques to correctly position all the shifts, without omission or false detection is evaluated with a performance criterion proposed by *Beaulieu et al.* [2008]. This criterion measures a distance between the locations of the true shifts and the detected shifts. It can be expressed as follows:

$$C = \begin{cases} \frac{1}{nd} \sum_{i=1}^{nd} (p_i^d - p_i)^2, & nr = nd \\ \frac{1}{nr} \left[\sum_{i=1}^{nd} (p_i^d - p_i)^2 + |nr - nd|(n-1)^2 \right], & nr > nd \\ \frac{1}{nd} \left[\sum_{j=1}^{nr} (p_j^d - p_j)^2 + |nr - nd|(n-1)^2 \right], & nr < nd, \end{cases} \quad (9)$$

where $p_i^d, i = 1, \dots, nd$ and $p_j, j = 1, \dots, nr$ represent the positions of the detected and real shifts, respectively; nd is the number of detected shifts; nr is the number of real shifts; and n is the length of the series. The pairs (p_i^d, p_j) are chosen to minimize the criterion. When the exact number of shifts is detected ($nr = nd$), the criterion is the sum of squares of the differences between the pairs which minimize the criterion. When $nr < nd$ or $nr > nd$, $(n-1)^2$ is added each time that a true shift is neglected or that a false shift is detected. This value corresponds to the square of the maximum possible distance between two shifts. C is equal

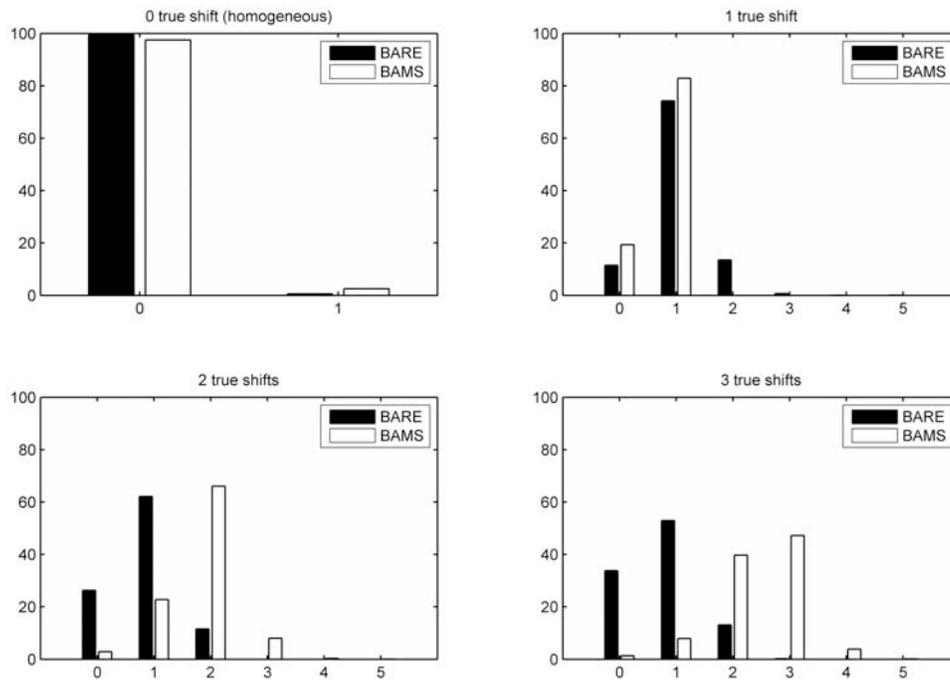


Figure 3. Percentage of series for which 0, 1, . . . , 5 shifts were detected by the two techniques when there are zero (homogeneous without autocorrelation), one, two, or three true shifts. BAMS represents the Bayesian technique for multiple shifts of *Seidou and Ouarda* [2007], and BARE represents the Bayesian technique for a single shift of *Seidou et al.* [2007].

to zero when all shifts are correctly positioned ($p_i^d = p_i, i = 1, \dots, nd = nr$). When C is close to zero, the detected shifts are located near the true positions ($p_i^d \approx p_i, i = 1, \dots, nd = nr$). A high value of C indicates that some shifts in the series are not detected or are wrongly detected ($nd \neq nr$). The performance criterion C was computed for all series with two and three shifts and the overall performance is the mean of the criterion over each set of synthetic series.

[21] The performances of the two techniques are compared with those of the techniques that were found the best to identify each type of synthetic series by *Beaulieu et al.* [2008]. These techniques are: the Jaruskova’s technique [*Jaruskova, 1996*], the bivariate approach [*Maronna and Yohai, 1978; Potter, 1981*] and a Bayesian technique [*Rasmussen, 2001*]. In the remainder of the text, these techniques are named JARU, BIVT and BAYE, respectively.

[22] The criteria used to assess the detection skills were chosen in order to be able to compare the results with those of *Beaulieu et al.* [2008]. However, from a Bayesian point of view, the use of these criteria could be contested, as the information given by the posterior probability distribution is not used in its full potential. However, to be able to compare Bayesian methods with classical methods, some criteria have to be used. In the comparative studies of homogenization techniques which involved Bayesian and classical techniques [see *Ducré-Robitaille et al., 2003; DeGaetano, 2006*] similar criteria were used to assess the performance of the different techniques.

4. Results

4.1. Ability to Identify the Exact Number of Shifts

[23] For all synthetic series, the number of shifts detected was compared with the true number of shifts that were

introduced in the series. Figure 3 presents the percentage of series for which 0, 1, . . . , 5 shifts were detected for the homogeneous series (0 shift) and the series with 1, 2 and 3 shifts. The position and magnitude of the detected shifts are not the focus here. For the homogeneous series and the series with a single shift, both techniques detected, in a large majority, the exact number of shifts (0 and 1). For the series with two and three shifts, BAMS detected, almost systematically, the exact number of shifts, but BARE detected one shift in most cases.

4.2. False Detection Rates in the Homogeneous Series

[24] Table 1 presents the false detection rates obtained with the two techniques. The best results observed in the work of *Beaulieu et al.* [2008] are also presented in Table 1 for comparison purposes. Both techniques gave a percentage of falsely detected shifts significantly smaller than 5%, the same as for JARU. Figure 4 presents the magnitude and

Table 1. Falsely Detected Shifts in the Homogeneous Series According to the Estimated Magnitude of the Shift

Magnitude Standard Deviation	BAMS ^a (%)	BARE ^b (%)	JARU ^c (%)
0–0.25	1.04	0.18	0.08
0.25–0.5	0.96	0.16	0.46
0.5–1	0.44	0.16	0.64
1–2	0.06	0.00	0.00
>2	0.00	0.00	0.00
Total	2.50 ^d	0.50 ^d	1.18 ^d

^aBayesian technique for multiple shifts [*Seidou and Ouarda, 2007*].

^bBayesian technique for a single shift [*Seidou et al., 2007*].

^cJaruskova’s technique [*Jaruskova, 1996*].

^dSignificantly smaller than the expected percentage of type I error obtained with a classical technique (5% critical level).

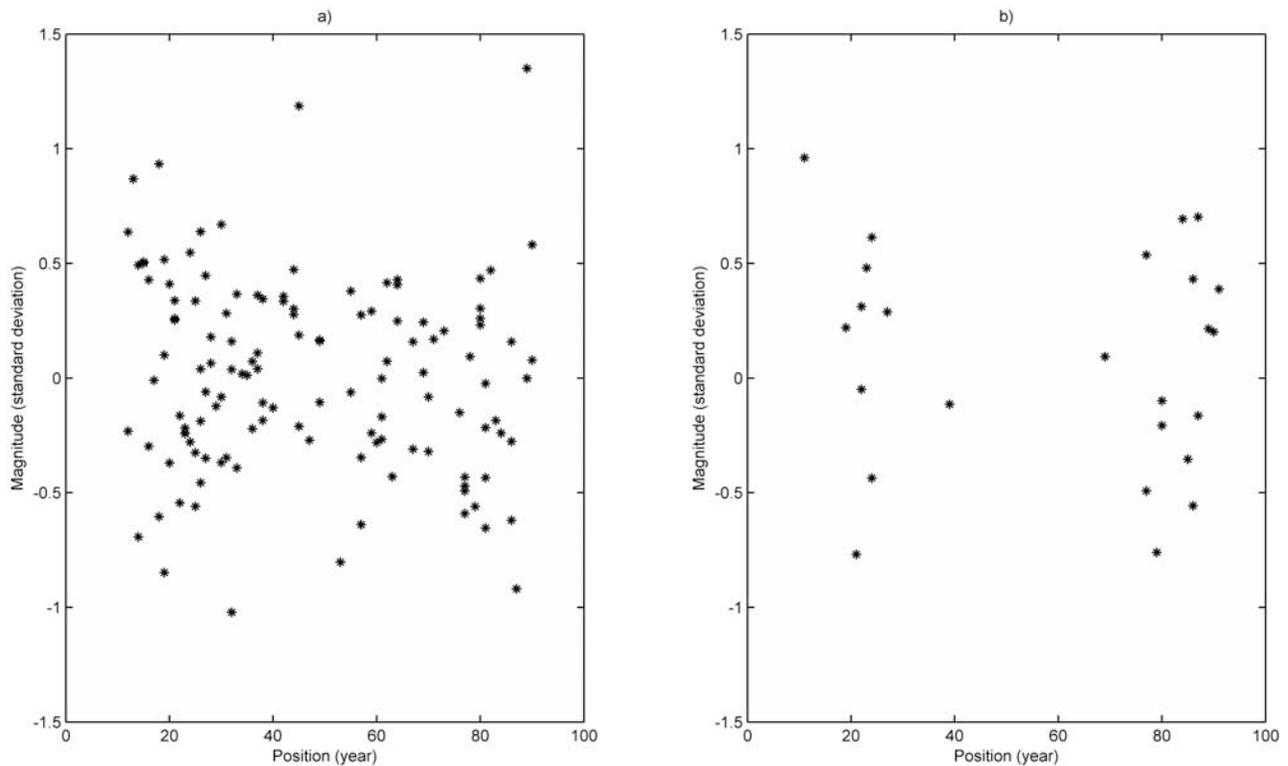


Figure 4. Falsely detected shifts by the two techniques when applied to the homogeneous series: (a) BAMS, the Bayesian technique for multiple shifts of *Seidou and Ouarda* [2007], and (b) BARE, the Bayesian technique for a single shift of *Seidou et al.* [2007].

position of the falsely detected shifts on the homogeneous series. It can be seen in Figure 4 that the number of falsely detected shifts tends to increase at the beginning or end of the series as well as their magnitudes. It is also interesting to note that the magnitudes of the falsely detected shifts rarely exceed 1 standard deviation. Figure 5 presents the percentage of falsely detected shifts obtained with the two techniques as a function of the autocorrelation coefficient. It can be seen that the presence of autocorrelation increases the risk of false detection.

4.3. Ability to Identify a Series With a Single Shift

[25] Table 2 presents the descriptive statistics of the absolute errors in position and magnitude for the two techniques. The mean absolute errors in position/magnitude are 12.7/0.4 and 19.2/0.6 for BAMS and BARE, respectively. The results were compared with those obtained by the three techniques (BIVT, JARU and BAYE) that led to the best performances in the work by *Beaulieu et al.* [2008]. The results are also presented in Table 2. The mean absolute errors in position/magnitude obtained with the various techniques were significantly different (Kruskal-Wallis test, 5% critical level). The Conover-Inman procedure was used (5% critical level) to make multiple pairwise comparisons between the mean absolute errors in position/magnitude obtained with the five techniques. The ranking of the absolute errors in position can be expressed as follows:

$$\text{BAYE BIVT JARU} < \text{BAMS} < \text{BARE}. \quad (10)$$

The techniques on the left are not significantly different. The same ranking was obtained for the absolute errors in

magnitude. The techniques BIVT, JARU and BAYE remain those with the smallest errors. The absolute errors in position and magnitude obtained with BAMS and BARE are significantly different, BAMS having a better performance than BARE. Consequently, the two Bayesian techniques have a good capacity to detect series with a

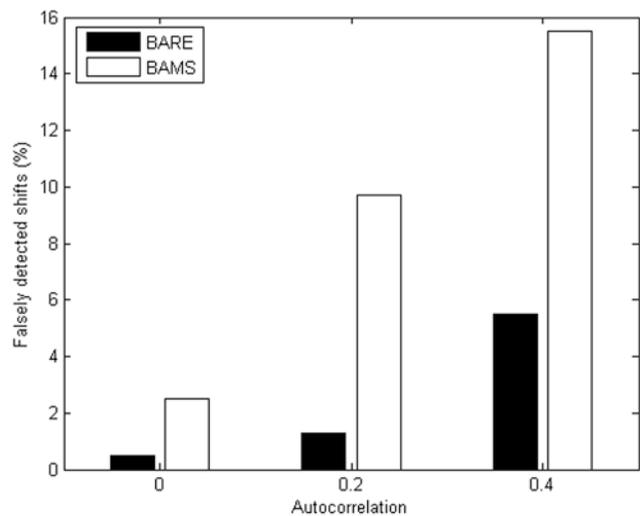


Figure 5. Falsely detected shifts by the two techniques when applied to the homogeneous series with different autocorrelation coefficients. BAMS represents the Bayesian technique for multiple shifts of *Seidou and Ouarda* [2007], and BARE represents the Bayesian technique for a single shift of *Seidou et al.* [2007].

Table 2. Descriptive Statistics of the Absolute Errors in Position and Magnitude Obtained When the Techniques Are Applied to Series With a Single Shift^a

Absolute Error	Statistic	BAMS ^b	BARE ^c	BAYE ^d	BIVT ^e	JARU ^f
Position	mean	12.7	19.2	5.2	7.5	9.2
	median	0	1	0	0	0
	standard deviation	32.0	38.4	19.6	24.7	27.6
Magnitude	mean	0.4	0.6	0.1	0.2	0.3
	median	0	0	0	0	0
	standard deviation	1.0	1.1	0.6	0.7	0.8

^aThe absolute errors in position and magnitude obtained with all techniques differ significantly (Kruskal-Wallis test, 5% critical level).

^bBayesian technique for multiple shifts [Seidou and Ouarda, 2007].

^cBayesian technique for a single shift [Seidou et al., 2007].

^dBayesian technique for a single shift [Rasmussen, 2001].

^eBivariate approach [Maronna and Yohai, 1978; Potter, 1981].

^fJaruskova's technique [Jaruskova, 1996].

single shift, but some classical techniques perform slightly better. The type II error rates (failure to reject the homogeneity hypothesis when it is false) are 11.5% and 17.1% for BAMS and BARE, respectively.

[26] Table 3 presents the rates of correctly identified, well-identified and well-positioned shifts for the two techniques as well as BAYE, BIVT and JARU. BAMS and BARE do not give the best performances. Nevertheless, the two techniques are able to well position a shift (75% and more). The percentage of well-identified shifts is similar to the percentage of well-positioned shifts. Hence, when the detected shift is well positioned (± 2 years of the true location), the magnitude is well estimated most of the time. However, there is a large gap between the rates of correctly identified shifts and the well-identified or well-positioned shifts (between 23% and 28%). Figure 6 presents the percentages of well-identified shifts for various classes of magnitude and position obtained on series with a single shift. The two techniques have a high percentage of well-identified shifts when the magnitude is larger than two standard deviations (almost 100%). When the magnitude is less than a standard deviation, the percentage of well-identified shifts decreased quickly (50% and less). The percentage of well-identified shifts does not seem to be affected by the position of the shift.

4.4. Ability to Identify a Series With Multiple Shifts

[27] Tables 4 and 5 present the descriptive statistics of the positioning criterion C (equation (9)) obtained on the series

with two and three shifts with BAMS and BARE. The results of the techniques that had the best performances in the study of Beaulieu et al. [2008] are also presented. These techniques are: BIVT for the series with two shifts (Table 4) and BAYE for the series with three shifts (Table 5). The most successful techniques are those with the lowest values of the performance criterion. The criteria obtained with the different techniques are significantly different (Kruskal-Wallis test, 5% critical level). Hence, multiple pairwise comparisons were used to rank the results obtained with the different techniques (Conover-Inman procedure, 5% critical level). For the series with two shifts, the multiple pairwise comparison procedure gives

$$\text{BAMS BIVT} < \text{BARE}. \quad (11)$$

The techniques on the left are not significantly different. BAMS led to the smallest criterion and its value is not significantly different from the value obtained with BIVT. For the series with three shifts, the ordering is

$$\text{BAMS} < \text{BAYE} < \text{BARE}. \quad (12)$$

BAMS led to the smallest criterion. The positioning criterion obtained with BAMS is significantly smaller than those obtained with all others techniques. It can be concluded that, BAMS is the most performing technique for the series with two and three shifts. In Table 5, the median criterion is very low. It means that in half of the

Table 3. Correctly Identified, Well-Identified, and Well-Positioned Shifts Obtained When the Techniques Are Applied to Series With a Single Shift

Identification of the Shift	BAMS ^a (%)	BARE ^b (%)	BAYE ^c (%)	BIVT ^d (%)	JARU ^e (%)
Correctly identified	56.5	46.2	57.8	62.7	63.0
Well identified	79.2	74.3	83.2	83.9	83.4
Well positioned	80.3	74.3	85.2	84.3	83.6

^aBayesian technique for multiple shifts [Seidou and Ouarda, 2007].

^bBayesian technique for a single shift [Seidou et al., 2007].

^cBayesian technique for a single shift [Rasmussen, 2001].

^dBivariate approach [Maronna and Yohai, 1978; Potter, 1981].

^eJaruskova's technique [Jaruskova, 1996].

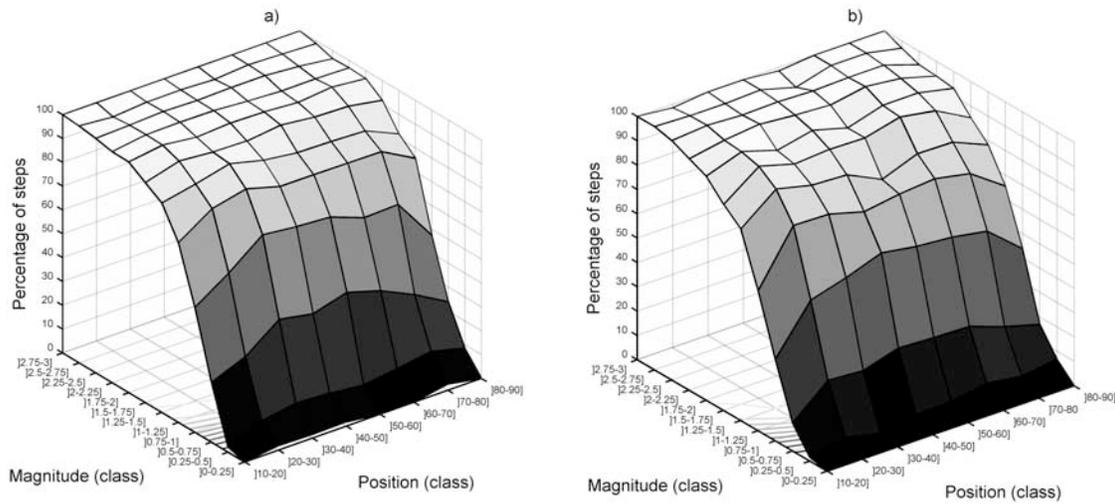


Figure 6. Percentage of well-identified shifts according to their position and magnitude obtained by the two techniques when applied to series with a single shift: (a) BAMS, the Bayesian technique for multiple shifts of *Seidou and Ouarda* [2007], and (b) BARE, the Bayesian technique for a single shift of *Seidou et al.* [2007].

series with two shifts, BAMS well positioned all the shifts without detecting nonexistent shifts or omitting real ones. In Table 6, the median criterion is higher. Indeed, the difficulty to identify all the shifts increases with the number of real shifts. The maximum criterion is 9801 and corresponds to the case for which all the real shifts were not detected.

5. Example of Application to Real Data

[28] In this section, an application of BAMS and BARE to detect a shift in the mean of a precipitation series is presented. A comparison is also made with some classical techniques which already proved their efficiency. The purpose of this application was to assess the appropriateness of the methods for practical use.

[29] The total annual precipitation series of Quebec City (station 7016294), Quebec, Canada, was tested for homogeneity. This station was chosen among 35 stations located in southern and central Quebec and nearby areas (Canada) with high-quality data (few missing values, long observation series) that were extracted from Environment Canada’s

database. The Quebec station was chosen as the base series because of the availability of neighbor series and of metadata (station history) for this station. The metadata are used to identify the cause of an inhomogeneity.

[30] The base station is located at latitude 46.8, longitude -71.38 and at an altitude of 70 m. Two neighbor stations, Shawinigan Falls (7018000) and La Pocatière (7054095), were identified in the set of 35 stations by considering the distance from the base station, the elevation difference, the observation period, the correlations and the correlations computed from the first difference series (Table 6). The two neighbor series are used as explanatory variables in the two regression models (BAMS and BARE), while the dependent variable is the base series. The neighbor series allow searching for a change which occurs only in the base series. When the neighbor series are too far from the base series or have a high elevation difference, the inhomogeneities in the base series may be hidden by the discordance between the series because of the large spatial variability of the precipitation series. The correlations are computed for the period 1944–1982, the common observation period for

Table 4. Descriptive Statistics of the Positioning Criterion C Obtained When the Techniques Are Applied to Series With Two Shifts^a

Statistic	BAMS ^b	BARE ^c	BIVT ^d
Mean	1702	5625	1730
Median	3	4901	2
Standard deviation	2508	2928	2789
Minimum	0	0	0
Maximum	9801	9801	9801

^aThe positioning criterion (equation (9)) obtained with all techniques differs significantly (Kruskal-Wallis test, 5% critical level).

^bBayesian technique for multiple shifts [*Seidou and Ouarda*, 2007].

^cBayesian technique for a single shift [*Seidou et al.*, 2007].

^dBivariate approach [*Maronna and Yohai*, 1978; *Potter*, 1981].

Table 5. Descriptive Statistics of the Positioning Criterion C Obtained When the Techniques Are Applied to Series With Three Shifts^a

Statistic	BAMS ^b	BARE ^c	BAYE ^d
Mean	2056	7198	2216
Median	2453	6534	2451
Standard deviation	2210	2156	2488
Minimum	0	0	0
Maximum	9801	9801	9801

^aThe positioning criterion (equation (9)) obtained with all techniques differs significantly (Kruskal-Wallis test, 5% critical level).

^bBayesian technique for multiple shifts [*Seidou and Ouarda*, 2007].

^cBayesian technique for a single shift [*Seidou et al.*, 2007].

^dBayesian technique for a single shift [*Rasmussen*, 2001].

Table 6. Distance, Elevation Difference and Correlation Between the Base Series (Quebec City) and the Neighbor Series

Neighbor Series	Distance (km)	Elevation Difference (m)	Correlation	Correlation, First Difference Series
Shawinigan Falls (7018000)	105	23	0.72	0.89
La Pocatière (7054095)	119	40	0.53	0.88

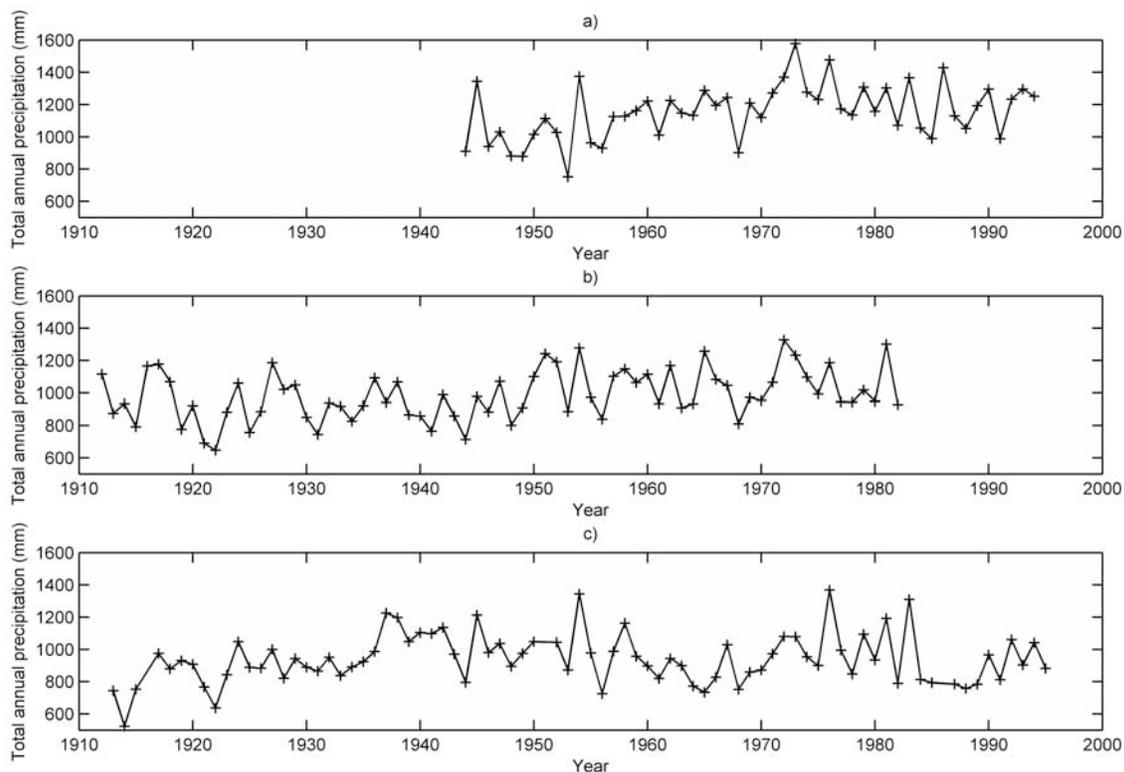
all series. The correlations computed from the first difference series are high (0.89 and 0.88), which indicates a strong relationship between the neighbor series and the base series without inhomogeneities. The correlations are lower (0.72 and 0.53); hence the base or the neighbor series may be inhomogeneous. The disagreement between the correlations and the correlations computed from the first difference series is an indicator of the presence of potential inhomogeneities in the base or in the neighbor series [Vincent, 1998]. Figure 7 presents the time series of the base and the two neighbor stations.

[31] BAMS and BARE were applied to the base series with the neighbor series as the explanatory variables. Both techniques assume IID Gaussian time series. All the series seemed normally distributed (Shapiro-Wilks test, 5% critical level). Since annual series were used, the independence can be assumed. However, it must be noted that the presence of a change point can bias the results of the goodness of fit tests. If these tests are rejected when applied to the entire series, it is important to apply them to each

segment of the series after the change point analysis. This issue is discussed by Perreault *et al.* [1999]. As the two techniques are applied along with classical techniques, noninformative priors are necessary to allow a reliable comparison. The prior probability of change was set to 0.5 and the minimal length between two shifts to 10 years. The posterior probability distributions for the number of shifts and for the positions of the shifts given the number of shifts are presented in Figure 8. By using the mode as the Bayes estimator, a single change point is detected with both techniques. Given the presence of a single change point, the change is detected in 1959 with BAMS. With BARE, the change is detected in 1956.

[32] The posterior distributions of the date of change obtained by the two techniques are bimodal; they both have two peaks in 1956 and 1959. The two detected shifts are in the same neighborhood. The two techniques could have detected the same shift at 3 years' interval. The presence of two shifts in the series is also possible, but only one is detected because of their proximity. When two change points are too close in a time series, several homogenization techniques may detect only one shift, which is in fact a blend of the two change points [Menne and Williams, 2005]. If they have different magnitudes, the larger shift could obliterate the smaller one. This is emphasized if the magnitudes of the shifts have the same signs. Consecutive shifts having the same signs are more difficult to detect than successive shifts with alternate signs [Beaulieu *et al.*, 2008].

[33] For comparison purposes, classical techniques that were found efficient in previous work [Beaulieu *et al.*, 2008] were also applied at a 5% critical level: BIVT, JARU and SNHT; the standard normal homogeneity test

**Figure 7.** Total annual precipitation of the base station and the two neighbor stations: (a) Quebec city (base), (b) Shawinigan Falls (neighbor), and (c) La Pocatière (neighbor).

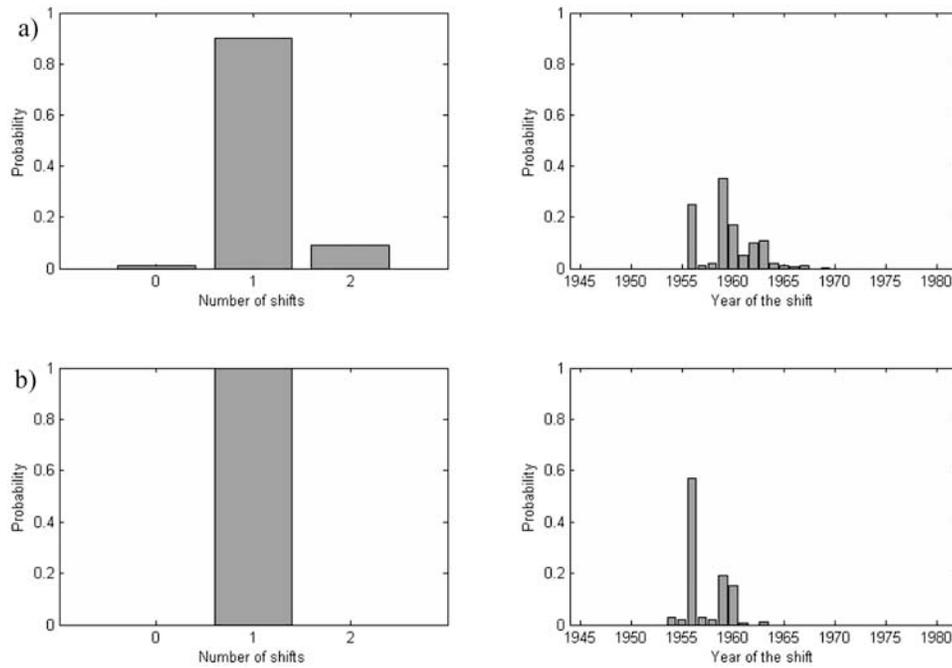


Figure 8. Posterior probability distributions for (left) the number of shifts and (right) the position of the shift obtained with the techniques (a) BAMS, the Bayesian technique for multiple shifts of *Seidou and Ouarda* [2007], and (b) BARE, the Bayesian technique for a single shift of *Seidou et al.* [2007].

[Alexandersson, 1986]. All these techniques also detected a single shift in 1959.

[34] The metadata was investigated to identify the cause of the detected shifts. There is no information about a potential cause of change for the years 1955–1956, but it is indicated that the Quebec station was relocated around 1958, with a change in its elevation. For precipitation series, a small change in the instrument height can induce a large shift in the observed amount of precipitation [Heino, 1997]. Hence, this relocation could be the cause of the detected shifts. The metadata could have been used to build the prior distribution. As changes are documented in the metadata, a higher prior probability of change could have been used. In this case, it would not affect the results and the shift would remain detected. However, shifts having a smaller magnitude could be more easily detected using an informative prior distribution. The metadata could also be used to set the minimum length between two segments as the minimum length between two documented changes.

6. Discussion, Conclusions, and Future Work

[35] This work aimed to verify the capacity of the two Bayesian techniques to discriminate homogeneous and inhomogeneous series. This was verified by applying the two techniques to several sets of synthetic series representing the typical total annual precipitation of the southern and central regions of the province of Quebec (Canada) and nearby areas. The synthetic series were also used in a previous comparative study aiming to identify the most appropriate techniques to homogenize precipitation series. The use of the same synthetic series allowed the comparison of the two Bayesian techniques with other techniques that were found to be the most suitable in the previous work.

The performance observed with the two techniques is summarized here.

[36] First, a weak percentage of false detection is very important in the homogeneous series. The two techniques presented in this work gave false detection rates smaller than 5%, which is the usual critical level used with classical techniques. Furthermore, the magnitudes of the falsely detected shifts rarely exceed 1 standard deviation (Table 1). Hence, the probability to introduce large artificial shifts is very low.

[37] Second, the techniques should be able to identify a single shift in a series. The percentages of positioned shifts were around 80% (BAMS) and 74% (BARE) for the series with a single shift (Table 3). The large shifts (magnitudes of 2 standard deviations and more) were well identified most of the time (Figure 5). The probability to identify a shift with a small magnitude is lower. The absolute errors in position and magnitude obtained from each technique were also studied (Table 2). In the comparative study of *Beaulieu et al.* [2008], some techniques had mean absolute errors in position and magnitude smaller than the two Bayesian techniques. The two techniques have shown to be efficient for the detection of series with a single shift, but some techniques are more powerful (BAYE, BIVT and JARU).

[38] Third, the techniques should be able to detect multiple shifts in a series, without omission or false detection. BARE seems disadvantaged in the presence of multiple shifts and BAMS gave the best performance. In the series with two and three shifts, BAMS gave the smaller criteria (Tables 4 and 5). Even though the mean criterion obtained with BAMS was the smallest for the series with two shifts, it was not significantly different from BIVT, the technique that was identified as the best in this case by *Beaulieu et al.* [2008]. For the series with three shifts (Table 5), BAMS was

better than the best technique (BAYE) presented by *Beaulieu et al.* [2008] for the same case.

[39] Finally, the number of shifts detected with the two techniques versus the number of true shifts was analyzed. For all synthetic series, BAMS often detected the exact number of shifts. BARE generally detected the exact number of shifts in homogeneous series and in the series with a single shift. However, in the series with two and three shifts, BARE detected a single shift in the majority of the cases (Figure 3).

[40] An example of application to a real data set was included to illustrate the homogenization procedures and to compare the two approaches to other performing techniques. The two Bayesian techniques were applied using non-informative priors for the real example as well as for the synthetic series to have results comparable to those obtained with classical methods. In homogenization, the metadata, some data not used in the analysis (the beginning or end of the record) and regional information can be incorporated in the prior distribution to give information about the parameters and to reduce the uncertainty. The metadata can indicate the position of a potential shift as well as its associated incertitude. Usually, the metadata is used to validate and identify the cause of an inhomogeneity, as it was done in the application presented in the present work (section 5). The use of metadata in the choice of the prior distribution is possible with Bayesian techniques and could improve their performances. For example, if some modifications are documented, the prior probability for a change could be set higher and if no shifts are documented, the prior probability for a change could be set lower. The minimum distance between two consecutive shifts could be set as the minimum length between two documented changes in the metadata. It can be expected that the shifts having a small magnitude will be more easily detected with informative prior distributions.

[41] The main advantage of BAMS is the simultaneous detection of several shifts. Most homogenization techniques were developed to detect at most one change point and are applied iteratively to detect multiple shifts. It was shown that it is advantageous to use a technique which allows the simultaneous detection of several shifts, such as BAMS, as it leads to best performances in this case. Furthermore, this technique does not require a large number of neighbor series to perform properly (in opposition to the technique developed by, e.g., *Caussinus and Mestre* [2004]). Another advantage of this technique is that it allows the detection of different types of inhomogeneities. It was tested for the detection of one or multiple shifts in the mean, but trends and shifts in the variance can also be detected with this technique. Future work should focus on this subject.

[42] These two techniques were applied with neighbor series to be able to compare the results with those obtained with classical methods. An important benefit of the two techniques is that it is also possible to apply them without neighbor stations. Most homogenization techniques require the presence of homogeneous neighboring series to represent the regional climate. The use of a neighbor series that is inhomogeneous, could alter the detection and correction of a shift. This issue is discussed by *Menne and Williams* [2005] with simulated series and by *Reeves et al.* [2007] with an example of application to real data series.

[43] A limitation of the two techniques, as well as most other homogenization techniques, is that they require data to be normally distributed. For the synthetic series and the case study, this hypothesis was not problematic, as the normality was respected. For the total annual precipitation series, the normality assumption is reasonable. For monthly or daily precipitation series, the normality hypothesis is less realistic [*Wilks*, 2006]. The normality assumption can be contested for several types of climatic series. Nevertheless, a Box-Cox transformation [*Box and Cox*, 1964] can be used to normalize the data. The robustness of the two techniques to other distributions was not verified in the present work and no existing work has been done in this direction. The independence hypothesis is also made with these two techniques. Once again, with a smaller time scale, this assumption may be less realistic. It was shown that the two techniques tend to detect more false shifts in the presence of autocorrelation (Figure 5). These techniques could be extended to accommodate series with autocorrelation with an approach similar to the one used by *Lund et al.* [2007]. The extension of the two techniques to less restrictive models is highly desirable.

[44] *Wang et al.* [2007] showed that the false detection rates and the probabilities of detection according to the shift positions generally have a U shape. For the false detection rates, it can be seen in Figure 4 that the two techniques tend to detect more false shifts at the beginning or the end of the series. The magnitudes of the falsely detected shifts tend also to increase at the extremities of the series. However, the power of detection seems mainly affected by the magnitude of the shift, and slightly affected by the position of the shift (Figure 6). *Wang et al.* [2007] proposed a penalized statistic to correct the U shape problem for a classic homogenization technique. With this penalized statistic, the false detection probabilities are the same for all positions in a homogeneous series and the confidence levels are the same for all positions of the detected change points [*Wang et al.*, 2007]. This issue was not addressed in the present work, but should be the subject of a future study.

[45] These two techniques were applied under the specific conditions of the province of Quebec, Canada, and nearby areas. The results of this work may only be valid under these conditions. The same techniques could lead to different performances on series with a different length, more than three shifts, other shift magnitudes, a different distribution, a different autocorrelation structure and/or different correlations with neighbor series. However, these techniques will be useful to detect a single (BARE) or multiple shifts (BAMS) in precipitation series similar to the synthetic series generated. They can also be used with other climatic variables (e.g., temperature), with hydrological variables (e.g., streamflow) or more generally, to solve different change point problems which can be expressed in a multiple linear regression form as long as the hypotheses of the linear regression are respected.

Appendix A: Detailed Presentation of BAMS

[46] Let $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ be the sample, n the sample size, m the number of change points, $\tau_0 = 0$, $\tau_1, \dots, \tau_m, \tau_{m+1} = n$ the positions of the change points and \mathbf{y}_{ij}

the observations from time i to time j . The superscript T indicates the transpose. We also denote $g(\cdot)$ the probability distribution of the time interval between consecutive change points and $g_0(\cdot)$ the probability distribution of the first change point. The j th segment is then $\mathbf{y}_{(\tau_{j-1}+1):\tau_j}$ with parameter Φ_j . The posterior probabilities of the change points positions were derived by *Fearnhead* [2006]:

$$\begin{cases} \Pr(\tau_1 | \mathbf{y}_{1:n}) = P(1, \tau_1) Q(\tau_1 + 1) g_0(\tau_1) / Q(1) \\ \Pr(\tau_j | \tau_{j-1}, \mathbf{y}_{1:n}) = P(\tau_{j-1} + 1, \tau_j) Q(\tau_j + 1) g(\tau_j - \tau_{j-1}) / Q(\tau_{j-1} + 1), \end{cases} \quad (\text{A1})$$

where $P(t, s)$, $s \geq t$ is the probability that t and s (representing the lower and upper limits of a given segment) are in the same segment and

$$\begin{aligned} P(t, s) &= \Pr(\mathbf{y}_{t:s}; t, s \text{ in the same segment}) \\ &= \int \prod_{i=t}^s f(y_i | \Phi) \pi(\Phi) d\Phi \end{aligned} \quad (\text{A2})$$

and $Q(t)$ is the likelihood of the segment $\mathbf{y}_{t:n}$ given a change point at $t - 1$. $Q(t)$ $t = 1, \dots, n$ and $P(t, s)$, $s \geq t$ are linked by these recursive equations:

$$\begin{cases} Q(1) = \sum_{s=1}^{n-1} P(1, s) Q(s+1) g_0(s) + P(1, n) (1 - G_0(n-1)) \\ Q(t) = \sum_{s=1}^{n-1} P(t, s) Q(s+1) g_0(s+1-t) + P(t, n) (1 - G(n-t)), \end{cases} \quad (\text{A3})$$

where $G(t) = \sum_{i=1}^t g(i)$ and $G_0(t) = \sum_{i=1}^t g_0(i)$. It is assumed that the observations are independent conditional on the change points and parameter values. The position of the change point is inferred on the basis of the coherence $P(t, s)$ of all segments $[t, s]$, where t varies from 1 to n and s varies from t to n .

[47] The procedure was adapted by *Seidou and Ouarda* [2007] to multiple linear regression. The response variable is noted by y_j ($j = 1, \dots, n$), or $\mathbf{y}_{n \times 1}$ in vectorial form, while x_{ij} ($i = 1, \dots, d^*$; $j = 1, \dots, n$) represents the j th observation of the i th explanatory variable ($\mathbf{X}_{d^* \times n}$ in matrix form). There are n observations and d^* explanatory variables. The relationship between the response variable and the explanatory variables is

$$y_j = \sum_{i=1}^{d^*} \theta_i x_{ij} + \varepsilon_j, \quad j = 1, \dots, n. \quad (\text{A4})$$

The response variable is normally distributed $N(\sum_{i=1}^{d^*} \theta_i x_{ij}, \sigma^2)$. Given the parameter vector $\Phi = [\theta_1, \theta_2, \dots, \theta_{d^*}, \sigma]$, the density of the response variable is

$$f(y_i | \Phi) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{\left(y_i - \sum_{j=1}^{d^*} \theta_j x_{ij}\right)^2}{2\sigma^2}\right). \quad (\text{A5})$$

Following *Rasmussen* [2001],

$$\begin{aligned} \Pr(\mathbf{y}_{t:s} | \Phi) &= \prod_{i=t}^s f(y_i | \Phi) = (2\pi\sigma^2)^{-(s-t+1)/2} \\ &\cdot \exp\left[-\frac{(\mathbf{y}_{t:s} - \mathbf{X}_{t:s}\boldsymbol{\theta})^T (\mathbf{y}_{t:s} - \mathbf{X}_{t:s}\boldsymbol{\theta})}{2\sigma^2}\right]. \end{aligned} \quad (\text{A6})$$

From equations (A3) and (A6),

$$\begin{aligned} P(t, s) &= \Pr(\mathbf{y}_{t:s}; t, s \text{ in the same segment}) = \int \prod_{i=t}^s f(y_i | \Phi) \pi(\Phi) d\Phi \\ &= \int (2\pi\sigma^2)^{-(s-t+1)/2} \\ &\cdot \exp\left[-\frac{(\mathbf{y}_{t:s} - \mathbf{X}_{t:s}\boldsymbol{\theta})^T (\mathbf{y}_{t:s} - \mathbf{X}_{t:s}\boldsymbol{\theta})}{2\sigma^2}\right] \pi(\Phi) d\Phi, \end{aligned} \quad (\text{A7})$$

where $P(t, s)$ is the posterior probability that t and s ($s \geq t$) are in the same segment. The prior probability distribution for the position of the change is uniform, given the preceding change point:

$$\pi(\tau_{i+1} | \tau_i) = \frac{1-p}{n-\tau_i}, \quad (\text{A8})$$

where p represents the prior probability for no change. Uniform priors are assumed for $\boldsymbol{\theta}$. Therefore, the prior distribution of Φ depends only on σ ($\pi(\Phi) = \pi(\theta_1, \dots, \theta_{d^*}, \sigma) \propto \pi(\sigma)$). It is assumed that the prior distribution of Φ has this particular form:

$$\pi(\Phi) = \pi(\sigma) = p(\sigma | a, c) = \frac{\sigma^{-a} \exp\left(-\frac{c}{2\sigma^2}\right)}{2^{\frac{a-3}{2}} c^{-\frac{a-1}{2}} \Gamma\left(\frac{a-1}{2}\right)}, \quad a > 1, c > 0, \quad (\text{A9})$$

where a and c are the hyperparameters and Γ represents the Gamma function. This prior was used instead of the inverted gamma distribution, which is the standard prior for Bayesian linear models, because it allows to obtain an analytic form for $P(t, s)$. The analytic expression of $P(t, s)$ is obtained after substituting equation (A8) in equation (A7) and integrating out σ and $\boldsymbol{\theta}$ in equation (A7):

$$P(t, s) = 2^{\frac{d^*}{2}} \pi^{\frac{d^*-t+s-1}{2}} \frac{(\boldsymbol{\varepsilon}_{s:t}^T \boldsymbol{\varepsilon}_{s:t} + c)^{-\frac{(t-s+a)}{2}} \Gamma\left(\frac{t-s+a}{2}\right)}{c^{-\frac{(a-1)}{2}} |\mathbf{X}_{s:t}^T \mathbf{X}_{s:t}|^{1/2} \Gamma\left(\frac{a-1}{2}\right)}, \quad (\text{A10})$$

where $\boldsymbol{\varepsilon}_{t:s}$ is a part of the vector of random errors between t and s and $\mathbf{X}_{t:s}$ represents the rows between t and s of the explanatory variables.

[48] To infer on the positions of change points, a set of M possible scatter schemes ($E = \{S_k, k = 1:M\}$) of the change points on the segment using the posterior probability mass of the first change point, and the conditional probability mass of subsequent change points was simulated. The set of scatter schemes is represented by E . Indeed, M should be large enough to obtain a reliable distribution for the positions of the change points. The k th element of E , S_k , is a set of \tilde{m}_k change points $S_k = \{\tilde{t}_1^k, \tilde{t}_2^k, \dots, \tilde{t}_{\tilde{m}_k}^k\}$. An efficient simulation algorithm for E is given by *Fearnhead* [2006]:

[49] 1. For a sample of size M , initiate M samples with a change point at $t = 0$.

[50] 2. For $t = 0, \dots, n - 2$, repeat the following steps.

[51] Compute the number n_t of samples for which the last change point was at time t ;

[52] If $n_t > 0$, compute $\Pr(\tau|\tau_{j-1} = t, \mathbf{y}_{1:n})$;

[53] Sample n_t times from $\Pr(\tau|\tau_{j-1} = t, \mathbf{y}_{1:n})$ and use the values to update the n_t samples of change points which have a change point at time t ;

[54] This algorithm is very efficient since $\Pr(\tau|\tau_{j-1} = t, \mathbf{y}_{1:n})$ has to be computed only one time regardless of the number of samples required from it. Inference on the number and positions of the change points is readily carried out using the M samples. For instance, the probability of having i change points is approximated by

$$\Pr(m = i) \approx \text{card}(\{k | \text{card}(S_k) = i\}) / M. \quad (\text{A11})$$

The posterior probability of having the k th change point at position t given m change points can be approximated by

$$\Pr(\tau_i = t | m) \approx \frac{\text{card}(\{k | (\text{card}(S_k) = m) \& (\tilde{t}_i^k = t)\})}{\text{card}(\{k | \text{card}(S) = m\})}, \quad (\text{A12})$$

where $\text{card}(S)$ stands for the number of elements of the set S . The estimators of the number and positions of change points are the modes of their posterior distributions.

Notation

β_1^* regression parameters before the change point (BARE).

β_2^* regression parameters after the change point (BARE).

β_0 component of the vector of regression parameters that is constant before and after the change point (BARE).

β_1 component of the vector of regression parameters that changes to β_2 after the change point (BARE).

β_2 component of the vector of regression parameters that replaces β_1 after the change point (BARE).

Γ Gamma function.

ε vector of random errors in the linear regression equation (BAMS).

$\varepsilon_{t:s}$ part of the vector of random errors between s and t (BAMS).

θ vector of regression parameters (BAMS).

$\theta_t^{(p)}$ vector of regression parameters at time t given p (BARE).

v_t vector of residuals of the linear regression (BARE).

π (Φ) prior of Φ (BAMS).

σ^2 variance (BAMS).

Σ_y variance-covariance matrix (BARE).

τ_k position of the k th change point (BAMS).

Φ vector of parameters of the linear regression equation (BAMS).

a parameter of the prior distribution of Φ (BAMS).

c parameter of the prior distribution of Φ (BAMS).

C positioning criterion.

d^* number of explanatory variables (including the intercept if any).

d_0^* number of explanatory variables for which the regression coefficients are constant before and after the change (BARE).

d_1^* number of explanatory variables for which the regression coefficients change (BARE).

E set of generated scatter schemes (BAMS).

$f(y_i | \Phi)$ conditional distribution of y given Φ (BAMS).

$g(t)$ probability distribution of the time interval between consecutive change points (BAMS).

$g_0(t)$ probability distribution of the first change point (BAMS).

k number of generated scatter schemes (BAMS).

M number of scatter schemes to generate with the posterior distributions of the positions of change points (BAMS).

m_u number of changes in the u th generated series (BAMS).

\hat{m}_u estimate of the number of changes in the u th generated series (BAMS).

\tilde{m}_k number of changes in the k th generated scatter scheme during the simulation of the change points.

n number of observations.

N Normal distribution.

nd number of shifts detected in the series.

nr number of true shifts in the series.

p position of the change point (BARE).

p_i positions of the true shifts ($i = 1, \dots, nr$).

p_i^d positions of the detected shifts ($i = 1, \dots, nd$).

$P(t, s)$ probability that t and s are in the same segment $s \geq t$ (BAMS).

$Q(t)$ likelihood of the segment $\mathbf{Y}_{t:n}$ given a change point at $t - 1$ (BAMS).

r number of response variables (BARE).

S_k k th scatter scheme generated with the posterior distributions of the positions of change points (BAMS).

t time.

\tilde{t}_i^k estimate of the i th change in the k th generated scatter scheme (BAMS).

t_i^k i th change in the k th generated scatter scheme (BAMS).

u number of the generated series $\{\mathbf{Y}\}_u$ in the validation procedure (BAMS).

\mathbf{X} matrix of explanatory variables (BAMS).

\mathbf{X}_t t th row of the matrix of explanatory variables (BARE).

$\mathbf{X}_{t:s}$ rows t to s of the matrix of explanatory variables (BAMS).

\mathbf{Y} vector of the response variable (BAMS).

\mathbf{Y}_t t th row of the vector of the response variable (BARE).

$\mathbf{Y}_{t:s}$ rows t to s of the vector of the response variable (BAMS).

$\{\mathbf{Y}\}_u$ u th generated series in the validation procedure (BAMS).

[55] **Acknowledgments.** The authors wish to thank the Fonds Québécois de Recherche sur la Nature et les Technologies, the National Sciences and Engineering Research Council of Canada (NSERC), the OURANOS Consortium, and the Canada Research Chair Program for funding this research. The authors would also like to thank L. Vincent of the Meteorological Service of Canada for providing the metadata and G. Boulet of Environment Quebec and A. Yagouti of Transport Canada for their helpful comments. Finally, the authors thank the Associate Editor and three anonymous reviewers for their valuable comments and suggestions.

References

- Alexandersson, H. (1986), A homogeneity test applied to precipitation data, *J. Climatol.*, *6*, 661–675, doi:10.1002/joc.3370060607.
- Bacon, D. W., and D. G. Watts (1971), Estimating the transition between two intersecting straight lines, *Biometrika Trust*, *58*(3), 525–534, doi:10.1093/biomet/58.3.525.
- Barry, D., and J. A. Hartigan (1992), Product partition models for change point problems, *Ann. Stat.*, *20*(1), 260–279, doi:10.1214/aos/1176348521.
- Barry, D., and J. A. Hartigan (1993), A Bayesian analysis for change point problems, *J. Am. Stat. Assoc.*, *88*, 309–319, doi:10.2307/2290726.
- Beaulieu, C., T. B. M. J. Ouarda, and O. Seidou (2007), A review of homogenization techniques for precipitation data and their applicability to precipitation series (in French), *Hydrol. Sci. J.*, *52*(1), 18–37, doi:10.1623/hysj.52.1.18.
- 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 Resour. Res.*, *44*, W02425, doi:10.1029/2006WR005615.
- Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., 617 pp., Springer, New York.
- Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*, 211–246.
- Caussinus, H., and F. Lyazrhî (1997), Choosing a linear model with a random number of change-points and outliers, *Ann. Inst. Stat. Math.*, *49*, 761–775, doi:10.1023/A:1003230713770.
- Caussinus, H., and O. Mestre (2004), Detection and correction of artificial shifts in climate series, *J. R. Stat. Soc., Ser. C*, *53*, 405–425.
- Chen, J., and A. K. Gupta (2001), On change-point detection and estimation, *Commun. Stat. Simul. Comput.*, *30*(3), 665–697, doi:10.1081/SAC-100105085.
- Chib, S. (1998), Estimation and comparison of multiple change-point models, *J. Econometrics*, *86*, 221–241, doi:10.1016/S0304-4076(97)00115-2.
- DeGaetano, A. T. (2006), Attributes of several methods for detecting discontinuities in mean temperature series, *J. Clim.*, *19*, 838–853, doi:10.1175/JCLI3662.1.
- Ducré-Robitaille, J. F., G. Boulet, and L. A. Vincent (2003), Comparison of techniques for detection of discontinuities in temperature series, *Int. J. Climatol.*, *23*, 1087–1101, doi:10.1002/joc.924.
- Easterling, D. R., and T. C. Peterson (1995), A new method for detecting undocumented discontinuities in climatological time series, *Int. J. Climatol.*, *15*, 369–377, doi:10.1002/joc.3370150403.
- Fearnhead, P. (2006), Exact and efficient Bayesian inference for multiple changepoint, *Stat. Comput.*, *16*, 203–213, doi:10.1007/s11222-006-8450-8.
- Gelfand, A. E., and A. F. M. Smith (1990), Sampling-based approaches to calculating marginal densities, *J. Am. Stat. Assoc.*, *85*, 398–409, doi:10.2307/2289776.
- Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, *82*(4), 711–732, doi:10.1093/biomet/82.4.711.
- Heino, R. (1997), Metadata and their role in homogenization, paper presented at the First Seminar for Homogenization of Surface Climate Data, Hung. Meteorol. Serv., Budapest.
- Holbert, D. (1982), A Bayesian analysis of a switching linear model, *J. Econometrics*, *19*, 77–87, doi:10.1016/0304-4076(82)90051-3.
- Jaruskova, D. (1996), Change-point detection in meteorological measurement, *Mon. Weather Rev.*, *124*, 1535–1543, doi:10.1175/1520-0493(1996)124 < 1535:CPDIMM > 2.0.CO;2.
- Lavielle, M., and E. Lebarbier (2001), An application of MCMC methods for the multiple change-points problem, *Signal Process.*, *81*, 39–53, doi:10.1016/S0165-1684(00)00189-4.
- Lee, A. S. F., and S. M. Heghinian (1977), A shift of the mean level in a sequence of independent normal random variables, *Technometrics*, *19*, 503–506, doi:10.2307/1267892.
- Lund, R., and J. Reeves (2002), Detection of undocumented changepoints: A revision of the two-phase regression model, *J. Clim.*, *15*, 2547–2554, doi:10.1175/1520-0442(2002)015 < 2547:DOUCAR > 2.0.CO;2.
- Lund, R., X. L. Wang, Q. Lu, J. Reeves, C. Gallagher, and Y. Feng (2007), Changepoint detection in periodic and autocorrelated time series, *J. Clim.*, *20*, 5178–5190, doi:10.1175/JCLI4291.1.
- Maronna, R., and V. J. Yohai (1978), A bivariate test for the detection of a systematic change in mean, *J. Am. Stat. Assoc.*, *73*, 640–645, doi:10.2307/2286616.
- Menne, M. J., and C. N. J. Williams (2005), Detection of undocumented changepoints using multiple test statistics and composite reference series, *J. Clim.*, *18*, 4271–4286, doi:10.1175/JCLI13524.1.
- Perreault, L., M. Haché, M. Slivitzky, and B. Bobée (1999), Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach, *Stochastic Environ. Res. Risk Assess.*, *13*, 201–216, doi:10.1007/s004770050039.
- Perreault, L., J. Bernier, B. Bobée, and E. Parent (2000), Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited, *J. Hydrol.*, *235*, 221–241, doi:10.1016/S0022-1694(00)00270-5.
- Peterson, T. C., et al. (1998), Homogeneity adjustments of in situ atmospheric climate data: A review, *Int. J. Climatol.*, *18*, 1493–1517, doi:10.1002/(SICI)1097-0088(199811)18:13 < 1493::AID-JOC329 > 3.0.CO;2-T.
- Potter, K. W. (1981), Illustration of a new test for detecting a shift in mean in precipitation series, *Mon. Weather Rev.*, *109*, 2040–2045, doi:10.1175/1520-0493(1981)109 < 2040:IOANTF > 2.0.CO;2.
- Rasmussen, P. (2001), Bayesian estimation of change points using the general linear model, *Water Resour. Res.*, *37*, 2723–2731, doi:10.1029/2001WR000311.
- Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Lu (2007), A review and comparison of changepoint detection techniques for climate data, *J. Appl. Meteorol. Climatol.*, *46*, 900–915, doi:10.1175/JAM2493.1.
- Seidou, O., and T. B. M. J. Ouarda (2007), Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows, *Water Resour. Res.*, *43*, W07404, doi:10.1029/2006WR005021.
- Seidou, O., J. J. Asselin, and T. B. M. J. Ouarda (2007), Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables, *Water Resour. Res.*, *43*, W08401, doi:10.1029/2005WR004835.
- Slonosky, V. C., P. D. Jones, and T. D. Davies (1999), Homogenization techniques for European monthly mean surface pressure series, *J. Clim.*, *12*, 2658–2672, doi:10.1175/1520-0442(1999)012 < 2658:HTFEMM > 2.0.CO;2.
- Smith, A. F. M., and D. G. Cook (1980), Straight lines with a change-point: A Bayesian analysis of some renal transplant data, *Appl. Stat.*, *29*(2), 180–189, doi:10.2307/2986304.
- Solow, A. R. (1988), A Bayesian approach to statistical inference about climate change, *J. Clim.*, *1*, 512–521, doi:10.1175/1520-0442(1988)001 < 0512:ABATSI > 2.0.CO;2.
- Stephens, D. A. (1994), Bayesian retrospective multiple-changepoint identification, *Appl. Stat.*, *43*(1), 159–178, doi:10.2307/2986119.
- Szentimrey, T. (1999), Multiple analysis of series for homogenization (MASH), in *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, WMO-TD Rep. 962, World Meteorol. Organ., Geneva, Switzerland.
- Vincent, L. A. (1998), A technique for the identification of inhomogeneities in Canadian temperature series, *J. Clim.*, *11*, 1094–1105, doi:10.1175/1520-0442(1998)011 < 1094:ATFTIO > 2.0.CO;2.
- Wang, X. L. (2003), Comments on ‘Detection of undocumented changepoints: A revision of the two-phase regression model,’ *J. Clim.*, *16*, 3383–3385, doi:10.1175/1520-0442(2003)016 < 3383:CODOUC > 2.0.CO;2.
- Wang, X. L., Q. H. Wen, and Y. Wu (2007), Penalized maximal t test for detecting undocumented mean change in climate data series, *J. Appl. Meteorol. Climatol.*, *46*, 916–931, doi:10.1175/JAM2504.1.
- Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 627 pp., Elsevier, New York.
- World Meteorological Organization (2003), Guidelines on climate metadata and homogenization, *Rep. WMO-TD 1186*, 50 pp., Geneva, Switzerland.

C. Beaulieu, Program in Atmospheric and Oceanic Sciences, Princeton University, 300 Forrester Road, Princeton, NJ 08540, United States. (beaulieu@princeton.edu)

T. B. M. J. Ouarda, Centre Eau, Terre et Environnement, Institut National de la Recherche Scientifique, University of Quebec, 490 de la Couronne, Quebec, QC G1K-9A9, Canada.

O. Seidou, Department of Civil Engineering, University of Ottawa, 161 Louis Pasteur, Ottawa, ON K1N 6N5, Canada.

X. Zhang, Climate Research Division, Science and Technology Branch, Environment Canada, 4905 Dufferin Street, Downsview, ON M5H-5T4, Canada.