INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **30**: 2342–2357 (2010)
Published online 24 December 2009 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.2056

RMetS
Royal Meteorological Society

A Bayesian normal homogeneity test for the detection of artificial discontinuities in climatic series

Claudie Beaulieu,[a*†] Taha B.M.J. Ouarda[b] and Ousmane Seidou[c]
[a] *Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA*
[b] *INRS-ETE, University of Quebec, Quebec, QC, Canada*
[c] *Department of Civil Engineering, University of Ottawa, Ottawa, ON, Canada*

ABSTRACT: A Bayesian Normal Homogeneity Test (BNHT) for the detection of artificial discontinuities in climatic series is presented. The test is simple to use and allows the integration of prior knowledge on the date of change from various sources of information (e.g. metadata or expert belief) in the analysis. The performance of the new test was evaluated on synthetic series with similar statistical properties as observed total annual precipitation in the southern and central parts of the province of Quebec, Canada. Different priors were used to investigate the sensitivity of the test to the choice of priors. It was found that (1) high-prior probability of no change yields low false detection rates on the homogeneous series; (2) the test has a very high power of detection on series with a single shift (the best power of detection if compared with previous methods applied to the same synthetic series); (3) shifts having a small magnitude are detectable with a low prior probability of no change and (4) when applied to series with multiple shifts with a segmentation procedure and a high probability of no change, the test proved to be performing well in detecting multiple shifts (as performing as the best techniques previously applied to the same synthetic series). An example of application to total annual precipitation in Quebec City, Canada is also presented to illustrate (1) a case for which the results are not affected by the choice of the prior parameters and (2) a case for which information about potential changes found in the metadata was integrated in the analysis and allowed the detection of a change that would not have been detected with a non-informative prior.

KEY WORDS homogenisation; climatic series; Bayesian technique; changepoint; exponential family; normal distribution; metadata

Received 19 December 2008; Revised 18 September 2009; Accepted 29 October 2009

1. Introduction

Raw climatic series are likely to exhibit artificial shifts due to modifications in the measurement procedures of the stations (e.g. relocation of a station, change of observer, modification in the immediate environment of the site, and so on). If left uncorrected, these artificial shifts may introduce a bias in climate studies such as trend or extreme value analysis. The problems induced by these artificial shifts in climatic data are the motivation for developing homogenisation techniques that are able to detect and correct these artificial shifts. For a comprehensive review of all these techniques, the reader is referred to Peterson *et al*. (1998), Aguilar *et al*. (2003), Beaulieu *et al*. (2007) and Reeves *et al*. (2007).

Homogenisation techniques can detect artificial changes by looking at metadata (history of the station), at relative variations with respect to neighbour series or at both of them. Other approaches such as side by side comparisons of instruments, statistical studies of instrument changes or use of a single station can also be used (Peterson *et al*., 1998; Aguilar *et al*., 2003), but are not discussed in this paper. Specific information contained in the metadata can provide the researcher with valuable knowledge of when a discontinuity is likely to have occurred and what may have caused it. Unfortunately, metadata is not always available and sometimes incomplete. Therefore, techniques that do not rely entirely on the metadata but can incorporate it in the analysis when available are desirable.

Neighbour series are often used in homogenisation to represent the regional climate. They are used to isolate the real regional climate change from the inhomogeneities that are present in the base series (i.e. series to be tested for homogeneity). Homogenisation techniques based on a comparison of the base series with neighbour series are presented, for example, in Potter (1981), Alexandersson (1986), Easterling and Peterson (1995), Vincent (1998), Szentimrey (1999) and Caussinus and Mestre (2004). When using such methods, metadata may be consulted independently of the analysis to identify the potential causes of inhomogeneities. However, an ideal method would simultaneously use information from

* Correspondence to: Claudie Beaulieu, Program in Atmospheric and Oceanic Sciences, Princeton University, 300 Forrestal Road, Princeton, NJ, USA. E-mail: beaulieu@princeton.edu
† This work was done while the corresponding author was at INRS-ETE, University of Quebec, Quebec, Canada.

both metadata and neighbour series in the analysis to minimise the risk of false detection and increase the power of detection of inhomogeneities. Bayesian approaches provide a straightforward framework to combine all available sources of information (metadata, expert knowledge) or lack of information (metadata missing or incomplete) through a prior distribution about the phenomenon being studied. The information provided by the prior distribution is updated with the observations to give a posterior distribution, which is used to make inference about the parameters of the data model (in the present work: the changepoint location).

Some Bayesian techniques were applied to detect inhomogeneities in temperature and precipitation series in previous comparative studies (Ducré-Robitaille *et al.*, 2003; DeGaetano, 2006; Beaulieu *et al.*, 2008, 2009). The technique presented in Lee and Heghinian (1977) and Perreault *et al.* (1999) was used to detect inhomogeneities in synthetic temperature series (Ducré-Robitaille *et al.*, 2003; DeGaetano, 2006). The technique developed by Rasmussen (2001) was applied to detect inhomogeneities in synthetic precipitation series (Beaulieu *et al.*, 2008). The techniques developed in Seidou *et al.* (2007) and Seidou and Ouarda (2007) were also applied to synthetic precipitation series. None of these approaches led to a good overall performance which can be defined by a small false detection rates in homogeneous series and a high power of detection in series with one or multiple shifts (Beaulieu *et al.*, 2009). These techniques were applied with non-informative priors because they were compared with classical homogenisation techniques. This can disadvantage the Bayesian approach, especially when additional information is available.

The general objective of this work is to present a new Bayesian Normal Homogeneity Test (BNHT) to detect inhomogeneities in climatic series. The specific objectives of this work are (1) to verify if the proposed Bayesian technique seems promising for the homogenisation of precipitation series, (2) to study its sensitivity to the choice of the prior distribution and (3) to present an application of the technique to real series. The proposed technique was originally developed by Lee (1998) to detect changes in the parameters of a distribution belonging to the exponential family.

The remainder of this paper is organised as follows: the BNHT is described in Section 2, the simulation study is presented in Section 3, the results are presented in Section 4, the real data series is analysed in Section 5 and finally, a discussion and some conclusions are, respectively, presented in Sections 6 and 7.

2. Description of the technique

2.1. Bayesian approach

Given a prior distribution π which summarises the information about the parameter θ and a vector of observations $\mathbf{x}$ having a probability density $f(\mathbf{x}|\theta)$, the Bayes theorem allows to actualise $\pi(\theta)$ with the observations:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta)d\theta} \tag{1}$$

The choice of the prior distribution can be either informative or non-informative. The knowledge about the phenomenon can be expressed through an informative prior. The lack of information and the associated uncertainty can be included by using a non-informative prior distribution. Parameters estimation can be performed with different Bayesian parametric estimators such as the posterior mean, median or mode. However, the posterior mode is preferred for hypothesis testing (Robert, 1992).

2.2. Bayesian normal homogeneity test

The technique used in this work is the normal case of a general changepoint technique which allows the detection of a change in the parameter of a distribution belonging to the natural exponential family (Lee, 1998). This problem has also been addressed in the Bayesian framework by Kander and Zacks (1966), Smith (1975) and Ghorbanzadeh and Lounes (2001). However, the approach of Lee (1998) was used because it is the most recent presenting the special case of a changepoint detection in the mean of a normally distributed time series, which is the case of interest in this paper. The general approach of Lee (1998) is presented in the Appendix.

The BNHT test, presented herein, enables the detection of a change in the mean of a single normally distributed time series. In order to account for neighbour series, the BNHT may be applied to a series of ratios or differences between the base series and neighbour series as proposed in Alexandersson (1986). The changepoint model can be represented by:

$$x_i \sim \begin{cases} N(\mu_1, \sigma^2) & i = 1, \ldots, k \\ N(\mu_2, \sigma^2) & i = k+1, \ldots, n \end{cases} \tag{2}$$

where x_i represents the ith observation in the series of ratios/differences, μ_1 and μ_2 are the mean of the series before and after the shift, σ^2 is the variance of the series that is assumed to be known, k represents the position of the change and n is the number of observations. The observations must be independent. The hypothesis that there is no change in the mean is tested against the hypothesis that there is a change in the mean at time k:

$$\begin{aligned} &H_0 : k = n \\ &H_1 : 1 \le k \le n-1 \end{aligned} \tag{3}$$

Hence, $k = n$ represents the case where no change has occurred in the series, while $k \neq n$ represents the case where a change has occurred at position k. The prior probability distribution for k is given by:

$$g_0(k) = \begin{cases} p & k = n \\ \dfrac{1-p}{n-1} & 1 \le k \le n-1 \end{cases} \tag{4}$$

The following prior probability distribution for k, $g_1(k)$ is proposed instead of the original one presented in Equation (4):

$$g_1(k) = \begin{cases} p & k = n \\ \dfrac{1-p}{n-a-b-1} & a < k < n-b \\ 0 & otherwise \end{cases} \quad (5)$$

where p is the prior probability of no change, a and b represent the support of the distribution. When a and b are set to 0, Equation (5) is equivalent to Equation (4). When a and b are different from 0, the position of the shift is constrained between positions a and $n-b$. The latter type of prior distribution is often used in the literature because it allows the user to set the range where a change has to be detected. It also allows reducing the problem of high-false detection rates at the beginning or end of the time series, as pointed out in Wang *et al.* (2007). A similar prior distribution was used by Holbert (1982) for changepoint detection in a multiple linear regression model.

Conjugate priors are used for the means before and after the changepoint: μ_1 and μ_2 were assumed to have $N(c/\sigma^2, 1/m\sigma^2)$ distribution. Lee (1998) found that the ML-II hyperparameters c and m are $\hat{c}_{1k} = \overline{x}_n$ and $\hat{m}_{1k} = n$ for the case $k = n$. When $k \neq n$, the ML-II hyperparameters are, respectively, $\hat{c}_{1k} = \overline{x}_k$, $\hat{m}_{1k} = k$, $\hat{c}_{2k} = \overline{x}_k^*$ and $\hat{m}_{2k} = n-k$ for the means before and after the shift. The posterior distribution of k given the vector of observations $\mathbf{x}$ is:

$$\pi(k|\mathbf{x}) \propto g_1(k) \cdot L(\mathbf{x}|k) \quad (6)$$

$$L(\mathbf{x}|k) = \begin{cases} \exp\left\{-\sum_{i=1}^{n} \dfrac{(x_i - \overline{x}_n)^2}{2\sigma^2}\right\} \\ \qquad k = n \\ \exp\left\{\dfrac{-\left[\sum_{i=1}^{k}(x_i - \overline{x}_k)^2 + \sum_{i=k+1}^{n}(x_i - \overline{x}_k^*)^2\right]}{2\sigma^2}\right\} \\ \qquad k \neq n \end{cases} \quad (7)$$

The variance can be estimated by the pooled sample variances (Lee, 1998):

$$\hat{\sigma}_p^2 = \min_{1 \leq k < n} \left[\sum_{i=1}^{k}(x_i - \overline{x}_k) + \sum_{i=k+1}^{n}\left(x_i - \overline{x}_k^*\right)\right] \Big/ (n-2) \quad (8)$$

The use of the pooled sample variance has also been suggested by Reeves *et al.* (2007) to improve the Standard Normal Homogeneity Test (Alexandersson, 1986). The motivation of using the pooled variance is that in the presence of a changepoint, the overall sample variance $\hat{\sigma}^2$ is biased and is an inconsistent estimator of σ^2, and in which case, the variance should be estimated by $\hat{\sigma}_p^2$.

Another prior distribution that could have been used for the position of the change is the triangular distribution:

$$g_2(k) = \begin{cases} p & k = n \\ \dfrac{2(1-p)(k-a)}{(e-a)(d-a)} & a \leq k \leq d \\ \dfrac{2(1-p)(e-k)}{(e-a)(e-d)} & d \leq k \leq e \\ 0 & otherwise \end{cases} \quad (9)$$

where d represents the most probable position of the shift (the mode of the distribution) and a and e represent the support of the distribution ($a < d < e < n$). This uniform distribution could be used instead of the uniform in the case where the date of change is known. This would be more informative about the position of the change. In this case, the posterior distribution of k given $\mathbf{x}$ would be:

$$\pi(k|\mathbf{x}) \propto g_2(k) \cdot L(\mathbf{x}|k) \quad (10)$$

The above prior distributions for the changepoint position were proposed because they lead to analytical expression for the posterior distribution and they can easily be used to integrate metadata in the analysis. Some examples of alternative prior distributions can be found in Ghorbanzadeh and Lounes (2001).

More generally, prior distributions can be chosen according to expert's belief and/or careful review of available metadata. For example, if a potential changepoint causing event is documented in the metadata, the prior probability of no change (p) should be smaller than 0.5. If the metadata is complete and no change is documented, the probability that a change occurred is small and then, p should be set higher than 0.5. If the user does not have enough information about the probability of change or no change, a common practice is to set p to 0.5, giving the same prior probability for the two possibilities. Similarly, if the metadata indicates that a change occurred during a given period, parameters a, b and e can be chosen to restrict the detection to this specific period. Finally, if the metadata indicates the specific date of change, parameter d can be chosen to attribute a higher probability of change to this date.

The absence/presence of a shift and its position can be inferred from the posterior distribution as the position corresponding to the highest posterior probability (posterior mode). If the mode of the posterior distribution is n, it means that there is no change in the series. If the mode of the posterior distribution is $k (k \neq n)$, it seems that there is a shift in the mean at position k. A credibility interval for the position can be computed as well. Figure 1 illustrates how a credibility interval is computed. In this example, the most probable position of the shift is 30. It corresponds to the maximum *a posteriori* (MAP). The 95% credibility interval is [23, 38]. This means that there is a probability of 0.95 that the shift occurs between positions 23 and 38. A multimodal posterior probability

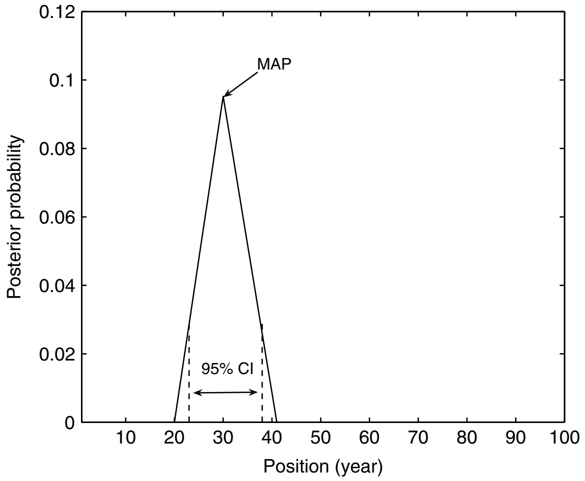

Figure 1. Illustration of the concept of credibility interval.

distribution and a discontinuous credibility interval would indicate that there might be multiple shifts.

If an inhomogeneity is detected, the base series can be corrected using the same approach as presented in Alexandersson (1986):

(1) for ratios: multiply the observations in the series before the change by the correction factor $\overline{x}_k^* / \overline{x}_k$, where $\overline{x}_k$ and $\overline{x}_k^*$ are, respectively, the mean of the series of ratios before and after the change at position k;
(2) for differences: add to the observations in the series before the change the value $\left(\overline{x}_k^* - \overline{x}_k\right)$.

3. Simulation study

3.1. Synthetic series

For comparison purposes, the BNHT was applied to the same synthetic series of precipitation that was generated for previous comparative studies of several homogenisation techniques (Beaulieu *et al.*, 2008, 2009). To estimate the statistical properties to reproduce with the model, a set of stations located in central Quebec and surroundings having long time series with few missing data was selected and their average characteristics were computed. The statistical properties reproduced in these series are (1) a mean total annual precipitation of 1089 mm, (2) a standard deviation (SD) of 142 mm and (3) a lag one autocorrelation of 0.02. Four different data sets were generated to represent the different cases that could occur in the reality: homogeneous series, series with a single shift and series with two and three shifts. The magnitude and positions were generated randomly. The positions are drawn from a truncated discrete uniform distribution $[11, n-10]$. The magnitudes are generated from a Beta distribution and lie between -3 and 3 SDs. For each base series, three correlated neighbour series were 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. More details about the generation scheme of the synthetic series are presented in Beaulieu *et al.* (2008).

3.2. Application of the technique

The BNHT was applied to the series of ratios between the base series and the mean of the neighbour series. As all neighbour series have the same correlation with the base series, they all have the same weight in the series of ratios. The BNHT was applied to all data sets by varying the parameters of the prior distribution of the position to study the sensitivity of the technique to the choice of prior. As it is a simulation study, non-informative priors for the position (Equation (5)) were used. For all data sets, the posterior distributions were computed with different prior probabilities of no change (p): 0.01, 0.05, 0.10, 0.25, 0.5, 0.75, 0.90, 0.95 and 0.99. The parameters a and b, used to set the interval of detection, were set to different values according to the type of data set. For the homogenous series, a and b were set to 0, 5, 10, 15 and 25, to allow the detection of a change in at least half of the series length. For the other series, a and b were set to 0, 5 and 10 only, as no shifts were introduced in the first ten or last ten observations in the series. To detect multiple shifts, the BNHT was applied recursively using the same approach as used in Beaulieu *et al.* (2008, 2009): 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. In Beaulieu *et al.* (2008, 2009), the techniques were applied by ignoring the shifts detected in the first or last ten observations. For the series with multiple shifts, the techniques were applied with a minimum length between two consecutive shifts set to ten observations. To be able to compare the results obtained with the BNHT with those obtained in Beaulieu *et al.* (2008, 2009), the BNHT was applied using a and b set to 0.

3.3. Performance evaluation

To make the results comparable with those obtained in Beaulieu *et al.* (2008, 2009), the same performance criteria were used. For the homogeneous series, the performance is assessed by computing the false detection rate, which is the percentage of cases for which the homogeneity is rejected, while it is true. For the series with a single shift, the number of well-positioned shifts (i.e. located within ± 2 years of the true position) is computed. For the series with multiple shifts, the performance is evaluated by penalising omissions, false detections and shifts detected at a position different than the true position by a criterion proposed by Beaulieu *et al.* (2008):

$$C = \begin{cases} \frac{1}{nd}\sum_{i=1}^{nd}(p_i^d - p_i)^2, nr = nd \\ \frac{1}{nr}[\sum_{i=1}^{nd}(p_i^d - p_i)^2 + |nr - nd|(n-1)^2], nr > nd \\ \frac{1}{nd}[\sum_{j=1}^{nr}(p_j^d - p_j)^2 + |nr - nd|(n-1)^2], nr < nd \end{cases} \quad (11)$$

where p_i^d, $i = 1, \ldots, nd$ and p_j, $j = 1, \ldots, nr$ represent, respectively, the positions of the detected and real shifts, 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 minimise the criterion. When the exact number of shifts is detected ($nr = nd$), the criterion penalises for the distance between the position of the real shift and the detected position. When $nr < nd$ or $nr > nd$, the criterion penalised each shift omitted or falsely detected by adding $(n-1)^2$. The best performance is obtained when C is equal to zero (all shifts are correctly positioned or close to zero and the detected shifts are located near the true positions). When C is high, some shifts were not detected or falsely detected. Following Beaulieu *et al.* (2008, 2009), the 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.

4. Results

4.1. False detection rates in the homogeneous series

Table I presents the false detection rates obtained with different prior parameters. It can be seen that the technique is very sensitive to the prior probability of no change (p). Hence, with a very low p (0.01), a false detection occurs in about 100% of the synthetic series. With a very high p (0.99), a false detection rarely occurs (0.1–0.2% of false detection). With a less informative p (0.5), the false detection rates lie between 6% and 8%. The technique seems less sensitive to the interval of detection. When the shifts detected at the beginning or end of the series are ignored, the percentage of false detections is under 5% with a prior probability of no change between 0.5 and 0.99. Figure 2 presents the histograms of the positions of the falsely detected shifts when the prior probability of no change is fixed to 0.25. This figure shows that false detection rates display a typical U-shape. When a and b are set to 0, false detections at the beginning or end of the series increase. This effect is reduced when the interval of detection is decreased (when a and b are increased). Therefore, restricting the detection at the beginning or end of the series is useful in reducing the number of false detections, but does not completely remove the U-shape effect.

Table I. Percentage (%) of false detections in the homogeneous series, with different prior probabilities of no change (p), different prior parameters of the interval of detection (a,b) and with a minimum segment length (segmin).

p	a,b						Segmin
	0	5	10	15	20	25	10
0.01	100	100	100	100	100	100	55.3
0.05	67.7	64.9	62.1	61.7	63.0	65.3	40.0
0.10	42.6	38.3	36.1	35.1	35.0	36.6	25.5
0.25	19.4	16.9	15.7	15.0	14.5	14.7	11.8
0.50	7.8	6.9	6.6	6.4	6.2	6.1	5.2[a]
0.75	3.2[a]	2.8[a]	2.8[a]	2.7[a]	2.7[a]	2.6[a]	2.1[a]
0.90	1.4[a]	1.2[a]	1.1[a]	1.0[a]	1.1[a]	1.0[a]	0.9[a]
0.95	0.7[a]	0.6[a]	0.5[a]	0.5[a]	0.5[a]	0.5[a]	0.4[a]
0.99	0.2[a]	0.1[a]	0.1[a]	0.1[a]	0.2[a]	0.1[a]	0.1[a]

[a] Significantly smaller than 5% (5% critical level).

4.2. Series with a single shift

Table II presents the percentage of well-positioned shifts in the series with a single shift. With a low value of p, the percentage of well-positioned shifts is around 93%. With a very high value of p, the percentage of well-positioned shifts is around 85%. When the shifts detected at the beginning or end of the series are ignored, the percentage of well-positioned shifts slightly diminishes. Figure 3 presents the percentage of well-positioned shifts according to their position and magnitude when the parameters a and b are set to 0. As the results are very similar for other values of a and b, it was not judged necessary to present these figures. The percentage of detection increases along with the magnitude of the shift. Shifts having a high magnitude (more than 1 SD) are almost always detected, even with high values of p. The percentage of detection starts to decrease around a magnitude of about 1 SD. Figure 3 also shows that a low value of p (0.01) allows the detection of shifts having a small magnitude that would not be detected with a high value of p (0.99). The position of the shift does not seem to have a high impact on the power of detection.

4.3. Series with multiple shifts

Tables III and IV present the descriptive statistics of the positioning criterion (C) computed from Equation (11) for the series with two and three shifts. Figure 4 presents the mean positioning criterion (C) for the series with two and three shifts. The positioning criterion seems to depend on the prior probability of no change and on the choice of parameters to handle the detections at the beginning or end of the series (Friedman's test, 5% critical level). With a minimum segment length of ten

Table II. Percentage (%) of well-positioned shifts in the series, with a single shift with different prior probabilities of no change (p), different prior parameters of the interval of detection (a,b) and with a minimum segment length (segmin).

p	a,b			Segmin
	0	5	10	10
0.01	92.5	92.8	93.1	91.5
0.05	92.1	92.3	92.6	91.3
0.10	91.5	91.6	91.9	90.8
0.25	90.5	90.6	90.8	90.0
0.50	89.6	89.7	89.9	89.2
0.75	88.5	88.6	88.8	88.2
0.90	87.4	87.5	87.7	87.1
0.95	86.6	86.8	86.9	86.4
0.99	84.9	85.0	85.1	84.7

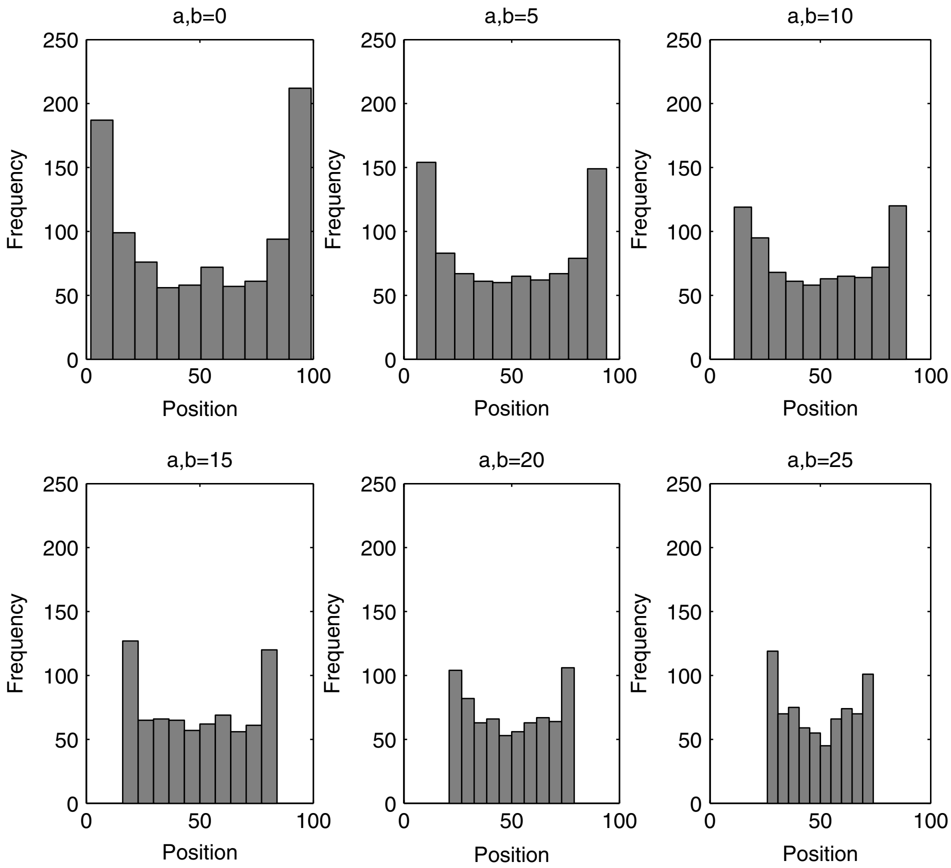

Figure 2. Histograms of the false detections in the homogeneous series when p is set to 0.25.

observations, the mean positioning criterion reaches its minimum. When the prior probability of no change is small ($p < 0.5$), the mean positioning criterion is high. This is due to a large number of false detections. Hence, when p is small, the mean number of shifts detected is higher than the real number of shifts (Tables V and VI). Furthermore, the risk of false detection increases as the test is reapplied several times to the same series. When the prior probability of no change is around 0.5–0.75, the mean positioning criterion is the smallest (Figure 3). With a high prior probability of no change ($p > 0.9$), the mean positioning criterion seems to increase as p increases. This is due to undetected shifts. Hence, when p has a high value, the mean number of detected shifts is smaller than the real number of shifts (Tables V and VI). The prior probability of no change has less impact when the BNHT is applied with a minimum length of ten observations between two shifts. Finally, it seems that it is better to set either a high value of p (0.5–0.95) and a and b to ten or to set a minimum length of ten observations between two shifts to detect multiple shifts.

4.4. Previous results

Previous comparative studies of several homogenisation techniques on the same synthetic series have been presented in Beaulieu *et al.* (2008, 2009). In these two previous studies, the shifts detected at the beginning or end of the series were ignored and a minimum of ten observations was imposed between two consecutive shifts. To be able to compare the BNHT with the other techniques, it was applied using the same procedure. Furthermore, only the results obtained with a p of 0.5 are compared, as it corresponds to the non-informative case. The other Bayesian techniques compared previously were also applied with non-informative priors.

For the homogeneous series, the false detection rate is significantly smaller than 5%, as with the majority of the techniques presented in Beaulieu *et al.* (2008, 2009). For the series with a single shift, the percentage of shifts positioned is 89.2% with a p of 0.5 and by ignoring the shifts detected at the beginning or end of the series. The highest percentage of well-positioned shifts in Beaulieu *et al.* (2008, 2009) was 85.2%. By testing the equality of these two proportions, it is found that the percentage of well-positioned shifts is significantly higher with the BNHT (rank-sum test, 5% critical level). In Beaulieu *et al.* (2008, 2009), the best results in the series with multiple shifts were obtained with the Bayesian approach for multiple shifts (BAMS) developed by Seidou and Ouarda (2007). These results were a mean C of 1702 and a median C of 3 in the series with two shifts, and in the series with three shifts, a mean C of 2056 and a median C of 2453. The median C obtained with the BNHT are not significantly different than those obtained with BAMS (rank-sum test, 5% critical level).

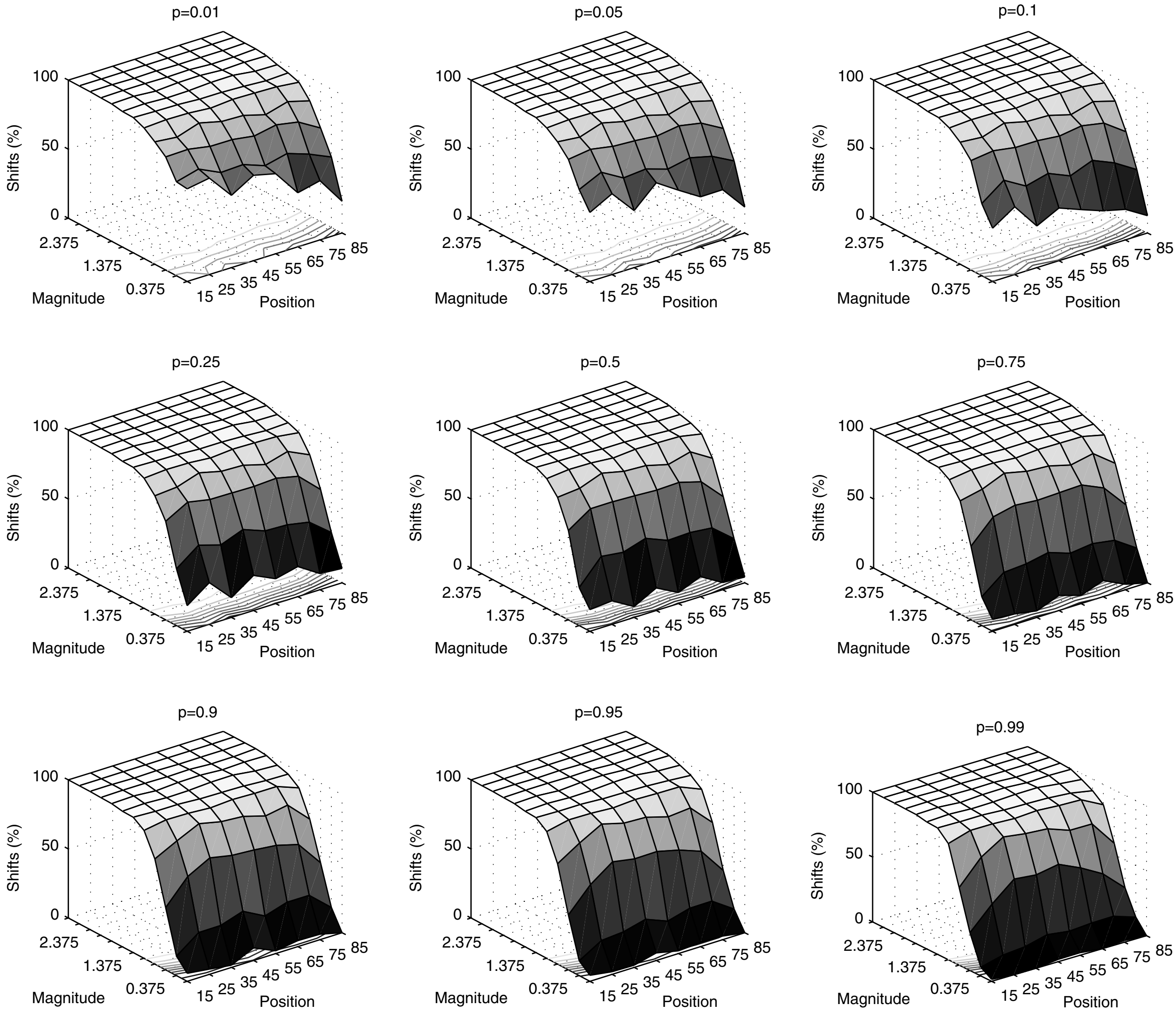

Figure 3. Percentage of well-positioned shifts according to their position and magnitude when a and b are set to 0.

Table III. Descriptive statistics of the positioning criteria obtained in the series with two shifts, with different prior probabilities of no change (p), different prior parameters of the interval of detection (a,b) and with a minimum segment length (segmin)[a].

p	a,b									Segmin		
	0			5			10			10		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.01	7666.1	8167.5	1425.9	8199.0	8167.9	108.5	6551.6	6534.2	332.6	2692.2	3267.0	2146.1
0.05	7213.2	7840.8	1784.5	7529.7	8167.5	1447.2	5847.4	6534.0	1795.3	2533.7	3267.0	2164.5
0.10	5969.8	6534.7	2283.0	6184.5	6535.5	1948.8	4357.4	4900.8	2321.7	2080.2	1120.5	2148.4
0.25	3523.6	3547.3	2505.2	3867.6	4900.5	2262.9	2671.8	3267.0	2312.5	1675.1	6.5	2145.0
0.50	2326.5	3267.0	2387.6	2481.1	3267.0	2322.1	1928.3	22.5	2297.1	1610.3	2.0	2214.5
0.75	2041.8	13.0	2468.4	1982.8	12.5	2438.3	1754.6	2.0	2445.0	1710.6	2.0	2303.4
0.90	2071.9	4.5	2606.0	2000.6	4.5	2591.0	1882.0	2.0	2601.7	1871.0	2.0	2376.1
0.95	2158.1	4.5	2696.9	2100.8	4.5	2685.8	2018.1	2.0	2693.6	1979.8	2.0	2410.3
0.99	2422.6	8.0	2868.6	2381.2	4.5	2863.2	2325.7	2.5	2866.3	2215.8	4.5	2463.6

[a] The prior probability of no change and the choice of parameters to handle the detections at the beginning or end of the series both influence the positioning criteria (Friedman's test, 5% critical level).

Table IV. Descriptive statistics of the positioning criteria obtained in the series with three shifts, with different prior probabilities of no change (p), different prior parameters of the interval of detection (a,b) and with a minimum segment length (segmin)[a].

p	a,b									Segmin		
	0			5			10			10		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.01	7041.7	7539.2	1390.1	7395.6	7351.5	165.6	4973.4	4902.0	482.6	2012.2	2450.5	1766.6
0.05	6719.5	7151.7	1639.3	6893.5	7351.1	1308.1	4632.8	4901.2	1254.9	2002.7	2450.5	1799.8
0.10	5720.1	6128.8	1949.9	5655.4	6125.9	1794.4	3469.1	3922.0	1931.2	1888.9	2450.3	1885.1
0.25	3459.0	3379.3	2135.5	3603.4	3920.4	1946.8	2233.0	2450.8	1944.1	1894.4	2450.3	2056.6
0.50	2510.6	2466.5	2142.6	2576.3	2457.5	2031.4	1989.3	2450.3	2138.8	2103.8	2451.3	2221.6
0.75	2458.4	2470.6	2359.4	2374.8	2452.3	2335.0	2147.4	2450.3	2409.8	2365.1	3267.0	2342.3
0.90	2694.1	3267.0	2581.9	2610.2	3267.0	2587.2	2473.5	3267.0	2615.5	2652.5	3267.0	2425.3
0.95	2894.2	3267.0	2698.7	2818.7	3267.0	2703.4	2719.0	3267.0	2714.8	2839.7	3267.0	2464.1
0.99	3328.8	3267.0	2842.9	3269.8	3267.0	2849.3	3206.5	3267.0	2863.8	3213.3	3267.0	2490.0

[a] The prior probability of no change and the choice of parameters to handle the detections at the beginning or end of the series both influence the positioning criteria (Friedman's test, 5% critical level).

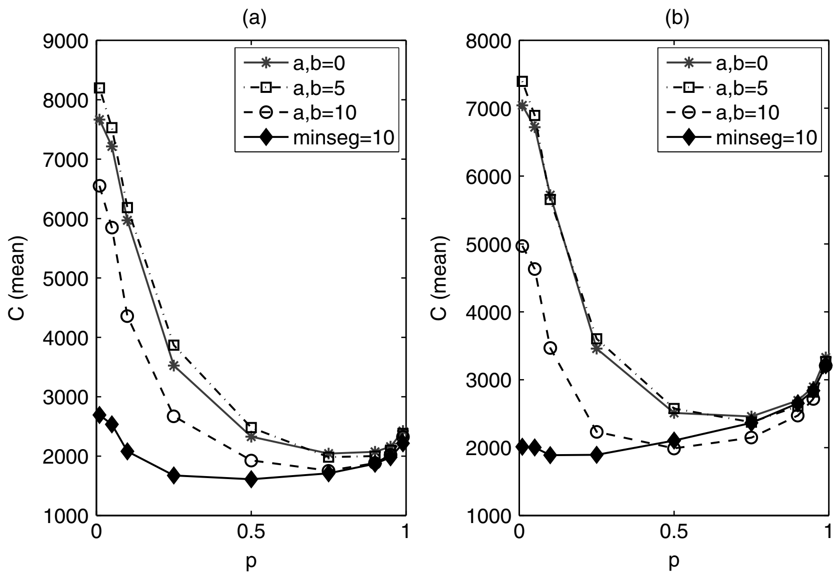

Figure 4. Mean positioning criteria obtained in the series with (a) two shifts (b) three shifts.

Table V. Number of detected shifts in the series with two shifts, with different prior probabilities of no change (p), different prior parameters of the interval of detection (a,b) and with a minimum segment length (segmin).

p	a,b						Segmin	
	0		5		10		10	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0.01	11.7	4.9	12.3	0.8	6.1	0.6	2.7	1.1
0.05	10.2	5.1	10.5	3.3	5.5	1.5	2.6	1.0
0.10	6.8	4.0	6.8	3.3	4.0	1.8	2.3	0.9
0.25	3.3	1.9	3.4	1.8	2.6	1.2	2.0	0.7
0.50	2.2	1.0	2.3	1.0	2.1	0.8	1.8	0.6
0.75	1.9	0.8	1.9	0.7	1.8	0.6	1.7	0.5
0.90	1.7	0.7	1.7	0.6	1.7	0.6	1.7	0.5
0.95	1.6	0.6	1.6	0.6	1.6	0.6	1.6	0.5
0.99	1.5	0.6	1.5	0.6	1.5	0.6	1.6	0.5

Table VI. Number of detected shifts in the series with three shifts, with different prior probabilities of no change (p), different prior parameters of the interval of detection (a,b) and with a minimum segment length (segmin).

p	a,b						Segmin	
	0		5		10		10	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0.01	12.6	4.6	12.3	0.8	6.1	0.6	3.3	1.0
0.05	11.6	4.8	11.2	2.7	5.8	1.2	3.2	1.0
0.10	8.6	4.1	8.2	3.2	4.8	1.6	3.0	1.0
0.25	4.6	2.2	4.7	2.0	3.4	1.2	2.6	0.8
0.50	3.1	1.4	3.3	1.4	2.8	1.0	2.5	0.8
0.75	2.6	1.1	2.6	1.0	2.5	0.9	2.4	0.8
0.90	2.4	1.0	2.3	0.9	2.3	0.9	2.3	0.8
0.95	2.2	0.9	2.2	0.9	2.2	0.9	2.2	0.8
0.99	2.0	0.9	2.0	0.9	2.0	0.9	2.0	0.8

5. Case study

An application of the BNHT to a real case study is presented in this section. The homogeneity of the total annual precipitation records of Quebec City (station 7016294) is tested. The observations and the metadata were provided by Environment Canada.

The Quebec City station was chosen because it has no missing values and a long series of observations and also because neighbour series and metadata are available for this station. The same case was studied previously in Beaulieu *et al*. (2009) with other techniques. The base station is located at latitude 46.8, longitude −71.38 and at an altitude of 70 m. Two neighbour stations, Shawinigan Falls (7018000) and La Pocatiere (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. The analysis was performed for the common period of observations: 1944–1982. For this period, the correlations between Quebec-Shawinigan Falls and Quebec-La Pocatiere are, respectively, 0.72 and 0.53. However, the correlations computed from the first difference series are higher, 0.89 and 0.88 for Quebec-Shawinigan Falls and Quebec-La Pocatiere, respectively. This indicates that there may be a shift in the base series which affects the correlation (Vincent, 1998). Figure 5 presents the data series of the base and the two neighbour stations. For the Quebec station, there are many documented relocations; two of which represent changes in elevation (1958 and 1977). For precipitation series, a small change in the instrument height can induce a very important shift in a series (Heino, 1997). Hence, these two relocations could have introduced artificial shifts in the Quebec precipitation series.

The BNHT was applied to the series of ratios between the base series and the mean of the two neighbour series. As two relocations could have introduced a shift in the series, the probability that there is a change in this series is high and hence, p should be set low. However, different values of p were used to verify the sensitivity of the technique to the choice of the prior parameters. A uniform prior distribution was used for the position (Equation (4)). The interval of detection was set to the entire series, because the second potential shift is located at the end of the series. The triangular prior distribution is not a good choice in this case because there are two potential positions of change (1958 and 1977) and no indications about which one is the most probable. Figure 6 presents the obtained posterior probability distributions. A change is detected in 1958 in all cases. The 95% Bayesian credibility interval is given by the years (1958,1959,1960,1961,1962). This indicates that there is a probability of 0.95 that a change occurred during 1958 and 1962, with 1958 being the most probable year of change. The equality of all posterior densities indicates that the detection of this shift is robust to the choice of prior. Even with a very high-prior probability of no change (0.99), the shift is detected. The shift has a large magnitude (>1 SD) and then, the detection is less influenced by the choice of the prior. This effect

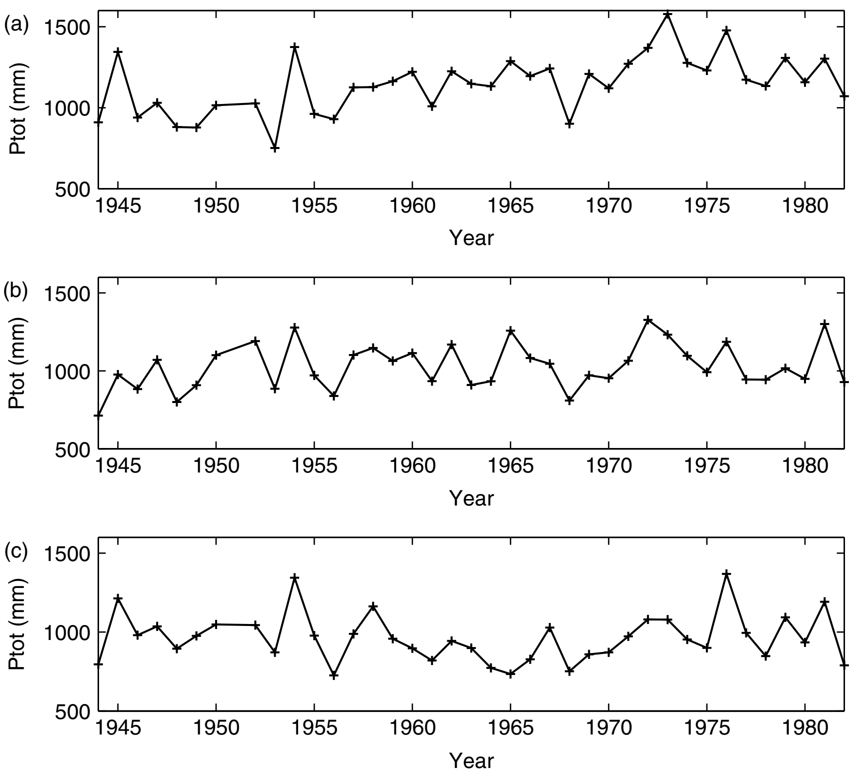

Figure 5. Total annual precipitation (Ptot) of the base station and the two neighbour stations: (a) Quebec City (base), (b) Shawinigan Falls (neighbour), (c) La Pocatiere (neighbour).

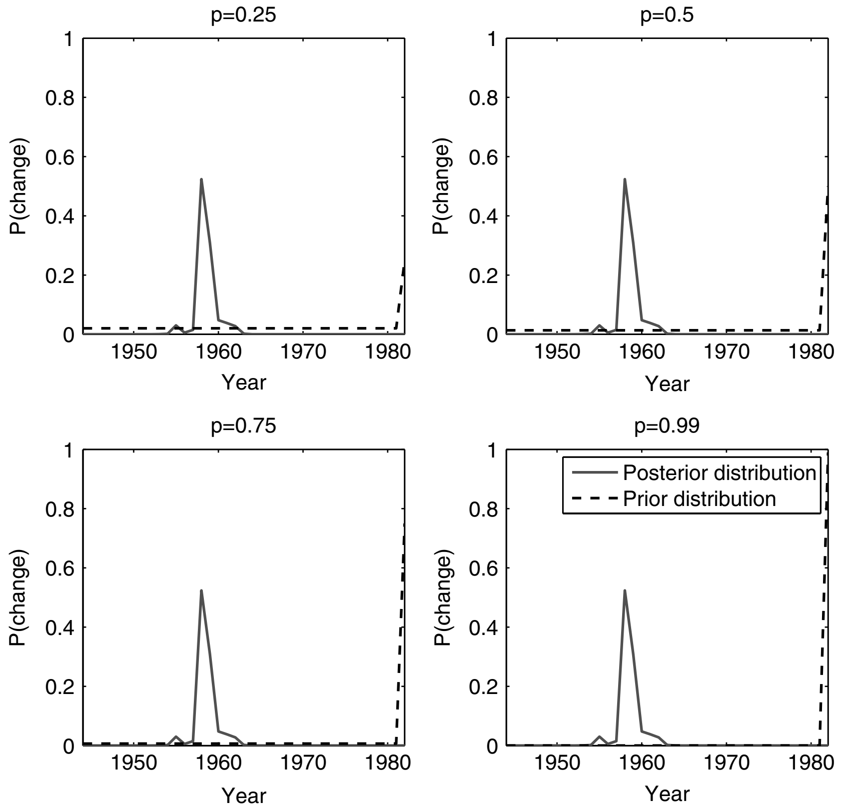

Figure 6. Posterior probability distributions for the position of the shift in the series of ratios of the base station (Quebec City) and the two neighbour stations (Shawinigan Falls and La Pocatiere).

was observed in the simulation study on the series with a single shift in Section 4.2: a shift having a small magnitude is more easily detected using an informative prior, but a shift having a large magnitude is detected most of the time and less influenced by the choice of the prior. In Beaulieu *et al*. (2009), different homogenisation techniques were applied to the same case study and a single significant shift was found in 1958.

The Shapiro–Wilk and Wald–Wolfowitz tests were applied to the series of ratios to verify their normality and independence, as the technique relies on these hypotheses. Both tests are rejected at the 5% critical level. The normality test is probably rejected due to the presence of a shift in the mean of the series, which affects the symmetry of the distribution (Figure 7). The Wald–Wolfowitz statistic checks the randomness hypothesis for a two-valued data sequence. In the presence of an important shift, the sequence is not random, as the smallest values are concentrated in one segment and the highest values are concentrated in the other segment. The usual statistical tests can only be performed as an indicator on both sides of the detected changepoint, after the changepoint test is applied (Perreault *et al*., 1999). These two tests were reapplied to the two segments of the series of ratios, separated after 1958, and the normality and independence hypotheses are respected (5% critical level). Furthermore, the equality of the variances before and after the shift was tested. The two variances do not seem different (Fisher test, 5% critical level).

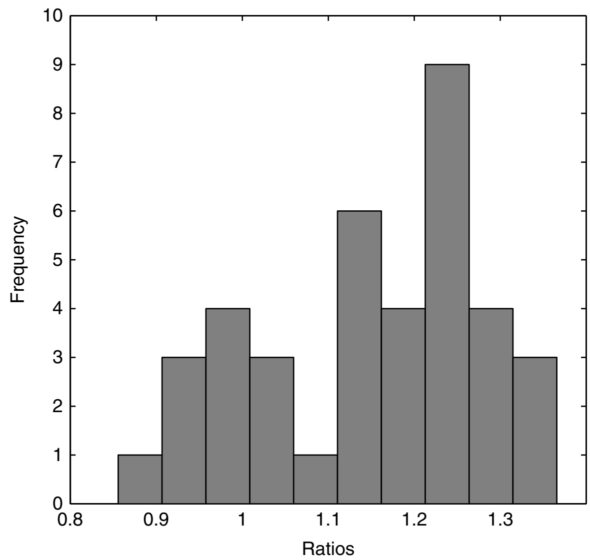

Figure 7. Histogram of the ratios of the base (Quebec City) and the mean of the two neighbour series (Shawinigan Falls and La Pocatiere).

The other documented change in 1977 could also have affected the homogeneity of the series. Furthermore, an undocumented change could have introduced an artificial shift in the series. Hence, the technique was reapplied to the two segments of the series. In the first segment, there is no indication about a potential change in the metadata. Then, a uniform prior distribution for the position was used and the prior probability of no change was set high. In a first attempt, a shift was detected at position 2, which seemed to be a false detection at the beginning of the series. Then, the interval of detection was restrained to years 1947–1956. The analysis was performed with p taking the values 0.5, 0.75 and 0.99 in order to study the sensitivity. Figure 8 presents the prior and posterior distributions for the analysis performed on the period 1944–1958. With $p = 0.5$, a change is detected in 1947. With $p = 0.75$ or $p = 0.99$, no change is detected. As the metadata does not indicate any change, $p = 0.75$ seems to be a more reasonable choice. Therefore, the series seems homogeneous for the period 1944–1958.

In the second segment, the metadata indicates that an inhomogeneity could have occurred in 1977. The triangular prior distribution was used by setting the interval of detection between 1972 and 1981, the mode at 1977 and a low prior probability of no change. Once again, $p = 0.5$ was used to study the sensitivity of the technique. Figure 9 presents the posterior probabilities for the position of the change. With p taking the values 0.01 or 0.25, the most probable year of change is 1980. The 95% credibility interval for the year of change is given by (1974,1975,1976,1977,1978,1979,1980) (computed with $p = 0.01$). The shift detected could be due to the relocation in 1977, since this year is in the credibility interval. It could also be a false detection occurring at the end of the series. With $p = 0.5$, no shift is detected. This could be because either the segment is not long enough or that the magnitude of the change is too small and is only detectable with a small p.

6. Discussion

6.1. Use of metadata and neighbour series

Homogenisation techniques presented in the literature rely on either the metadata or/and on a comparison with neighbour series to perform the analysis. When using relative homogenisation techniques (based on neighbour series), the metadata is often consulted after a change is detected to determine its cause. Some techniques use metadata to identify which segment of the series will be tested (Peterson *et al*., 1998). The BNHT allows a straightforward use of the metadata by incorporating it into the prior distribution. This is an advantage of the BNHT over the other homogenisation techniques which may not allow easy integration of non-uniform prior distribution. Furthermore, the BNHT will help to detect changes having a small magnitude which would not be detected with other techniques or with a non-informative prior distribution. This feature is especially important for precipitation series, because they exhibit a high variability which reduces the power of detection of changepoint techniques. Furthermore, the power of detection of changepoint techniques generally increases with the length of the series. Once again, the use of informative priors can help to increase the power of

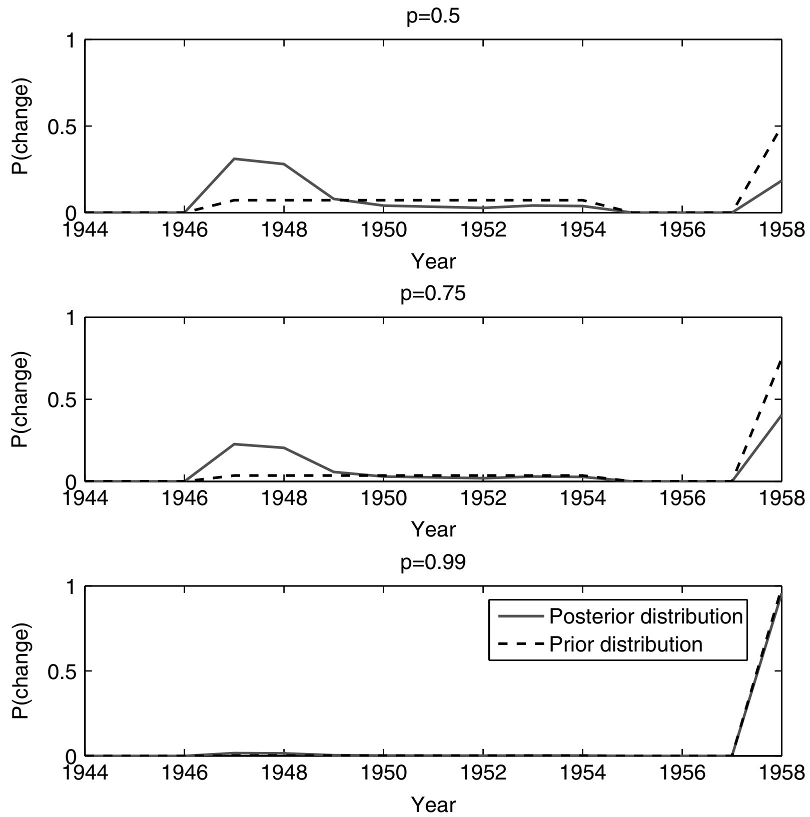

Figure 8. Posterior probability distributions for the position of the shift (1944–1958) in the ratios of the base (Quebec City) and the mean of the two neighbour series (Shawinigan Falls and La Pocatiere).

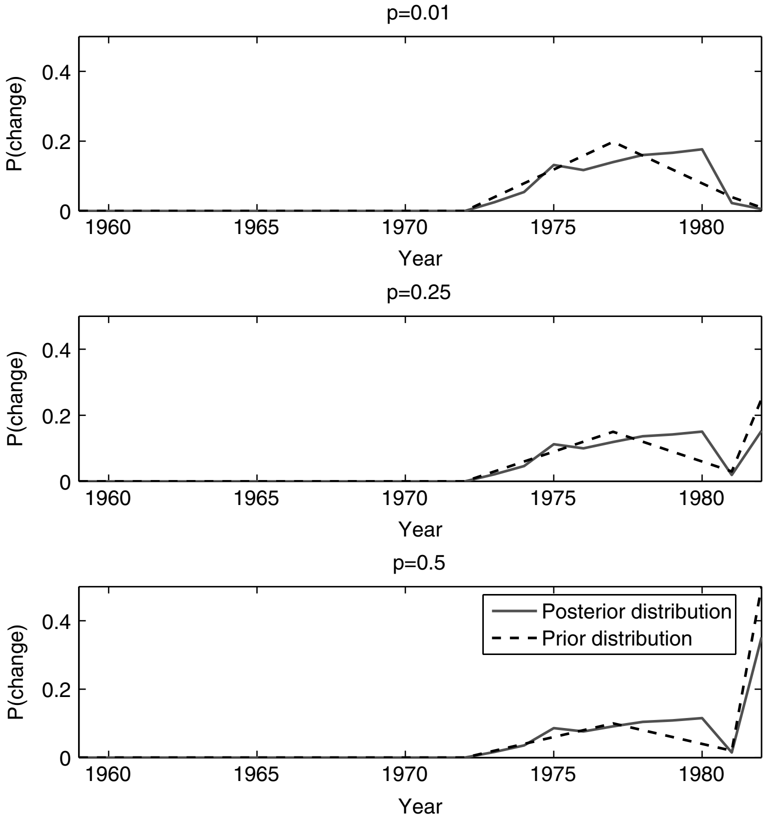

Figure 9. Posterior probability distributions for the position of the shift (1959–1982) in the ratios of the base (Quebec City) and the mean of the two neighbour series (Shawinigan Falls and La Pocatiere).

detection in short series. It is also possible to apply the BNHT without neighbour stations. However, it should be acknowledged that the use of a neighbour series that is inhomogeneous could alter the detection and correction of a shift (Menne and Williams, 2005; Reeves *et al.*, 2007).

6.2. Priors and subjectivity

The prior distribution for the position of the shift used in the simulation study was uniform. The advantage is that it can accommodate both the case for which the position of the potential shift is known and the case for which it is not. However, it was shown in the case study that other prior distributions could be relevant. For example, a triangular distribution could be used to give the highest prior probability to the position for which a change is recorded in the metadata. Furthermore, the distributions for the means are conjugate. The hyperparameters are chosen according to the maximum likelihood type-II approach for prior selection. Once again, other approaches for prior selection could be used and could lead to different results. This is a common criticism of the Bayesian approach: it contains a certain amount of subjectivity. Hence, the specification of the prior distribution and of the likelihood function entails a succession of subjective judgements (Goldstein, 2006). However, even a classical use of statistics can also end up being subjective as the choice of the technique to use for a given problem, as well as the choice of the neighbour series, heavily depends on the analyst and can lead to different results.

6.3. Multiple shifts

The BNHT was developed for at most one changepoint, as it is the case for most homogenisation techniques. Nevertheless, it was applied to multiple shifts using a segmentation procedure, as it is often done in practice. The test could be extended to detect multiple shifts in a more fashionable way. For example, a hierarchy of sub-hypotheses as presented in Chen and Gupta (2001) could be constructed. The form of the posterior distribution (unimodal *vs* multimodal) can also provide a good hint about the number of shifts. For example, if the posterior probability distribution of the position of the change has two main peaks and that the credibility interval is discontinuous, there are probably more than one shift in the series.

However, the detection of multiple inhomogeneities in precipitation series is usually more difficult than in temperature series, because of their larger spatial and temporal variability. For example, Müller-Westermeier *et al.* (2004) performed a homogeneity analysis of temperature and precipitation series in Germany for each month of the year. They rarely detected more than one inhomogeneity in precipitation series.

6.4. U-shape

In Wang *et al.* (2007), it was shown that the false detection rates and the probabilities of detection according to the shift positions generally have a U-shape. It was shown in Figure 2 that BNHT tends to detect more shifts at the beginning or end of the series. The parameters of the interval of detection help to reduce this effect, but do not eliminate it entirely. This has to be taken into consideration when applying this technique. It was also shown that false detections are reduced when the shifts detected at the beginning or end of the series are ignored. Nevertheless, other forms of prior distributions which give less weight to the positions at the end of the series could be used to correct the U-shape effect.

6.5. Normality and independence hypotheses

The BNHT was developed for normal and independent series. The normal distribution was used as it is reasonable to represent total annual precipitation and also commonly used to represent temperature series. The effects of departure from normality were not studied in this work. It is also realistic to assume that annual precipitation series will be independent. Therefore, it was not judged necessary to study the effect of the presence of autocorrelation in this work. However, it is known that homogenisation techniques developed for independent error series tend to detect more false shifts in the presence of positive autocorrelation (Lund *et al*., 2007; Wang, 2008). Tang and MacNeill (1993) proposed a straightforward technique to adjust for the presence of autocorrelation with different changepoint techniques. In presence of first-order autocorrelation (φ_1), the posterior probability distribution for the case of no change, $\pi(k = n|\mathbf{x})$, can be multiplied by the factor $\sqrt{1 + \varphi_1/1 - \varphi_1}$. For a positive autocorrelation, this correction will increase the posterior probability of no change ($k = n$), and hence, the number of false detections will be reduced.

7. Conclusions

7.1. Summary of results

The BNHT allows for the detection of a change in the mean of a normal series. The advantages of the proposed approach are its ease of use, the straightforward inclusion of metadata and the analytical form of the posterior distributions. The performance of this test and its sensitivity to the choice of the prior parameters were assessed in a simulation study. For the homogeneous series, the test was very sensitive to the choice of the prior probability of no change. The percentages of false detections were small ($<5\%$) when the prior probability of no change was set to at least 0.75 or to at least 0.5 when the shifts detected at the extremities of the series are ignored. For series with a single shift, the technique was less sensitive to the choice of the prior parameters and had a high power of detection in all cases. It gave better results than those obtained in Beaulieu *et al*. (2008, 2009). Furthermore, this technique is able to detect a shift with a small magnitude, especially when the prior probability of no change is small. For the series with multiple shifts, the technique gave positioning criteria equivalent to the best techniques previously compared. A case study was also presented to illustrate the application of the BNHT to a real data series. Different priors were used to verify their effect on the posterior density of the shift's position. A change documented in the metadata in 1958 was detected with all different priors. The same shift was also detected by other homogenisation techniques applied to the same case study in Beaulieu *et al*. (2009).

7.2. Recommendations

According to the results of the simulation study and the case study, some recommendations can be made for further use of the BNHT. If the metadata seems incomplete, the BNHT should be applied with a prior probability of no change of 0.5 and a uniform prior distribution. If there is nothing documented that could have affected the homogeneity at this station, then the BNHT should be applied with a high-prior probability of no change (≥ 0.75) to minimise the risk of false detection. The parameters of the interval of detection should then be set to avoid false detection at the beginning or end of the series. Hence, even with a high-prior probability of no change, this technique still has a high power of detection. If there is a documented change in the metadata that could have affected the homogeneity of the series, the prior probability of no change could be set low (≤ 0.25) and the parameters of the interval of detection could concentrate the probabilities around the date of the potential shift or the triangular distribution could be used. This would allow the detection of shifts having a small magnitude that would not be detected with a high-prior probability of no change. In the case study, it was shown that a changepoint can affect the results of the normality and independence tests. Then, these hypotheses should be verified in each segment after the detection of a shift.

7.3. Future work

The BNHT was applied in this paper to synthetic series of precipitation, but it could be used to detect inhomogeneities or a change in the mean in other climatic series that are normally distributed (e.g. temperature).

The general formulation of the technique as presented in Lee (1998) allows the detection of a change in the parameters of a distribution belonging to the exponential family. The Gamma distribution also belongs to the exponential family and is a more natural choice to represent variables with asymmetric distributions such as monthly or seasonal precipitation. Future work should focus on the application of the technique to detect inhomogeneities in precipitation series based on the Gamma distribution.

Such test would be useful to detect real or artificial (due to modifications in the observation procedures) changes in the parameter of the distribution (belonging to the exponential family) of a larger range of climatic series. For example, it could be used to detect changes in the intensity of tornado counts (Poisson distributed), in wind speed (Gamma), in cloudiness (Beta) or in climatic extremes (Weibull). The posterior densities for other distributions should constitute the subject of future work.

Acknowledgements

The authors wish to thank the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT), the National Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair Program for funding this research. The authors would like to thank an anonymous reviewer, the editor and I. B. MacNeill for their precious comments. The authors would also like to thank L. A. Vincent of the Meteorological Service of Canada for providing the metadata.

Appendix

Bayesian changepoint detection in the natural exponential family

A probability distribution is said to belong to the natural exponential family if its probability density function can be written as:

$$f(x|\theta) = \exp\{\theta x + \varphi(\theta) + S(x)\} \tag{A1}$$

where $\varphi(\theta)$ and $S(x)$ are the two functions which depend on the form of the distribution. Consider the following changepoint model:

$$x_i \sim \begin{cases} F_{\theta_1} & i = 1, \ldots, k \\ F_{\theta_2} & i = k+1, \ldots, n \end{cases} \tag{A2}$$

where the x_i s are independent variables, F_{θ_1} and F_{θ_2} represent the distributions before and after the shift with parameters θ_1 and θ_2, respectively, k is the unknown position of the changepoint and n is the length of the series. The joint distribution of $\mathbf{x} = (x_1, \ldots, x_k, x_{k+1}, \ldots, x_n)$ is then:

$$f(\mathbf{x}|\theta) = \exp\left\{\theta_1 \sum_{i=1}^{k} x_i + \theta_2 \sum_{i=k+1}^{n} x_i + k\varphi(\theta_1) + (n-k)\varphi(\theta_2) + \sum_{i=1}^{n} S(x_i)\right\} \tag{A3}$$

The hypothesis that there is no change in the parameter of the distribution ($\theta_1 = \theta_2$) is tested against the hypothesis that there is a change in the parameter of the distribution at time k:

$$\begin{aligned} H_0 &: k = n \\ H_1 &: 1 \le k \le n-1 \end{aligned} \tag{A4}$$

In Lee (1998), the prior probability distribution for k is given by:

$$g_0(k) = \begin{cases} p & k = n \\ \dfrac{1-p}{n-1} & k \neq n \end{cases} \tag{A5}$$

where p represents the prior probability of no change ($0 \le p \le 1$). A conjugate prior is used for the parameters θ_1 and θ_2. For the exponential family, the conjugate prior has the form:

$$\pi(\theta|m, c) \propto \exp\{mc\theta + m\varphi(\theta)\} \tag{A6}$$

where m and c are hyperparameters such that $m \ge 0$ and $c \in X$. X represents the sample space of the x_i s. Lee (1998) set the prior information about θ_1 and θ_2 to take into account the position of the changepoint:

$$\begin{aligned} \pi(\theta_1|k) &= \pi(\theta_1|m_{1k}, c_{1k}) \quad k = n \\ \pi(\theta_1, \theta_2|k) &= \prod_{i=1}^{2} \pi(\theta_i|m_{ik}, c_{ik}) \quad k \neq n \end{aligned} \tag{A7}$$

where $\pi(\theta_i|m_{ik}, c_{ik}) \propto \exp\{m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)\}$, $i = 1, 2, m_{ik} \ge 0$, $c_{ik} \in X$. The joint prior distribution of (k, θ_1, θ_2) is given by:

$$\begin{aligned} \pi(k, \theta_1) &= p \cdot \frac{\exp\{m_{1k}c_{1k}\theta_1 + m_{1k}\varphi(\theta_1)\}}{\tilde{\pi}(\Theta|m_{1k}, c_{1k})} \quad k = n \\ \pi(k, \theta_1, \theta_2) &= \frac{1-p}{n-1} \cdot \frac{\exp\left\{\sum_{i=1}^{2}[m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)]\right\}}{\prod_{i=1}^{2}\tilde{\pi}(\Theta|m_{ik}, c_{ik})} \quad k \neq n \end{aligned} \tag{A8}$$

where

$$\tilde{\pi}(\Theta|m_{ik}, c_{ik}) = \int_{\Theta} \exp\{m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)\}\mathrm{d}\theta_i \quad i = 1, 2 \tag{A9}$$

and Θ represents the parameter space of θ_1 and θ_2. According to the Bayes theorem and from the joint distribution of $\mathbf{x}$ (Equation (A3)) and the joint prior distribution (Equation (A8)), the posterior distribution of the position of the changepoint k is:

$$\pi(k|\mathbf{x}) \propto \begin{cases} p \cdot \dfrac{\tilde{\pi}\left(\Theta|m_{1k}+n, \dfrac{m_{1k}c_{1k} + \sum_{i=1}^{n} x_i}{m_{1k}+n}\right)}{\tilde{\pi}(\Theta|m_{1k}, c_{1k})}, k = n \\ \dfrac{1-p}{n-1} \cdot \dfrac{\tilde{\pi}\left(\Theta|m_{1k}+k, \dfrac{m_{1k}c_{1k} + \sum_{i=1}^{k} x_i}{m_{1k}+k}\right)}{\tilde{\pi}(\Theta|m_{1k}, c_{1k})} \cdot \dfrac{\tilde{\pi}\left(\Theta|m_{2k}+(n-k), \dfrac{m_{2k}c_{2k} + \sum_{i=k+1}^{n} x_i}{m_{2k}+(n-k)}\right)}{\tilde{\pi}(\Theta|m_{2k}, c_{2k})}, k \neq n \end{cases} \tag{A10}$$

The values of the hyperparameters m_{ik} can be thought of as the sample size of priors. Then they can be set such that $m_{1k} = k$ and $m_{2k} = n - k$, which corresponds to the size of the two samples $(x_1, \ldots, x_k)$ and $(x_{k+1}, \ldots, x_n)$ (Lee, 1998). To estimate the hyperparameters c_{ik}, Lee (1998) proposed to use the maximum likelihood type-II (ML-II) approach for prior selection, such as presented in Berger (1985). It consists in performing a maximisation of the marginal distribution of $x, m(\mathbf{x}|\pi)$, over the hyperparameters $c_{ik}, i = 1, 2, k = 1, \ldots, n$. When $m_{1k} = k$ and $m_{2k} = n - k$, to maximise $m(\mathbf{x}|\pi)$ over π is equivalent to maximising the two following functions on c_{1k} and c_{2k} (Lee, 1998):

$$\frac{\tilde{\pi}(\Theta|2k, (c_{1k} + \overline{x}_k)/2)}{\tilde{\pi}(\Theta|k, c_{1k})} \text{ or } \frac{\tilde{\pi}(\Theta|2(n-k), (c_{2k} + \overline{x}_k^*)/2}{\tilde{\pi}(\Theta|n-k, c_{2k})} \tag{A11}$$

where $\overline{x}_k = \sum_{i=1}^{k} x_i/k$ and $\overline{x}_k^* = \sum_{i=k+1}^{n} x_i/(n-k)$. The logarithm of these functions gives an expression which is similar to the function:

$$h(s) = \log(\tilde{\pi}(\Theta|2k, (c+s)/2)) - \log(\tilde{\pi}(\Theta|k, s)) \tag{A12}$$

c is $\overline{x}_k$ or $\overline{x}_k^*$ and s represents the hyperparameters to maximise. By taking the derivative function, the maximum values can be obtained. For the normal case, $h(s)$ is concave downward and the ML-II hyperparameters are easily obtained.

Notations

Θ	parameter space for θ_1 and θ_2
θ	parameter of a distribution
θ_1	parameter of a distribution before the shift
θ_2	parameter of a distribution after the shift
μ_1	mean of the series before the shift
μ_2	mean of the series after the shift
$\pi(\theta)$	prior distribution of the parameter θ
$\pi(\theta_1\|k)$	prior distribution of the parameter θ_1 given the position of the shift (under the no-change hypothesis)
$\pi(\theta_1, \theta_2\|k)$	prior distribution of the parameters θ_1 and θ_2 given the position of the shift (under the change hypothesis)
$\pi(\theta\|m, c)$	prior distribution of the parameter θ with the hyperparameters m and c
$\pi(k\|\mathbf{x})$	posterior distribution of k given the observations $\mathbf{x}$
σ^2	variance
$\hat{\sigma}^2$	sample variance
$\hat{\sigma}_p^2$	pooled sample variance
φ_1	first-order autocorrelation
$\varphi(\theta)$	function of θ in the exponential family form
a	prior parameter indicating the position of the beginning of the interval of detection
b	prior parameter indicating the position of the end of the interval of detection
C	positioning criterion
c	hyperparameter for the prior distribution of the mean
$\hat{c}_{1k}$	ML-II hyperparameter for the prior distribution of the mean μ_1
$\hat{c}_{2k}$	ML-II hyperparameter for the prior distribution of the mean μ_2
d	prior parameter indicating the mode of the prior triangular distribution
e	prior parameter indicating the end of the interval of detection (triangular distribution).
F_{θ_1}	distribution before the shift
F_{θ_2}	distribution after the shift
$f(\mathbf{x}\|\theta)$	probability density of $\mathbf{x}$ given the parameter θ
$g_0(k)$	prior probability distribution of the position of the shift in Lee (1998)
$g_1(k)$	prior probability distribution of the position of the shift uniform over $[a, n-b]$
$g_2(k)$	prior probability distribution of the position of the shift triangular with a mode of d and over the support $[a, n-b]$
k	position of the shift
m	hyperparameter for the prior distribution of the mean
$\hat{m}_{1k}$	ML-II hyperparameter for the prior distribution of the mean μ_1
$\hat{m}_{2j}$	ML-II hyperparameter for the prior distribution of the mean μ_2
N	normal distribution
n	number of observations
nd	number of shifts detected in the series
nr	number of true shifts in the series
p	prior probability of no change
p_i	positions of the true shifts ($i = 1, \ldots, nr$)
p_i^d	positions of the detected shifts ($i = 1, \ldots, nd$)
$S(x)$	function of x in the exponential family form
X	sample space for $\mathbf{x}$
$\mathbf{x}$	vector of observations
x_i	ith observation in the series of ratios/differences
$\overline{x}_k$	mean of the first k observations in the series of ratios/differences
$\overline{x}_k^*$	mean of the last $n-k$ observations in the series of ratios/differences
$\overline{x}_n$	overall mean of the series of ratios/differences

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. *Guidelines on Climate Metadata and Homogenization*. Report WMO-TD 1186, World Meteorological Organization, Geneva, Switzerland.

Alexandersson H. 1986. A homogeneity test applied to precipitation data. *Journal of Climatology* **6**: 661–675.

Beaulieu C, Ouarda TBMJ, Seidou O. 2007. A review of homogenization techniques for precipitation data and their applicability to precipitation series (in French). *Hydrological Sciences Journal* **52**: 18–37.

Beaulieu C, Seidou O, Ouarda TBMJ, Zhang X. 2009. Intercomparison of homogenization techniques for precipitation data continued: the recent comparison of two changepoint regression models. *Water Resources Research* **45**: W08410.

Beaulieu C, Seidou O, Ouarda TBMJ, Zhang X, Boulet G, Yagouti A. 2008. Intercomparison of homogenization techniques for precipitation data. *Water Resources Research* **44**: W02425.

Berger JO. 1985. *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. Springer: New York.

Caussinus H, Mestre O. 2004. Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society, Series C* **53**: 405–425.

Chen J, Gupta AK. 2001. On change point detection and estimation. *Communications in Statistics-Simulation and Computation* **30**: 665–697.

DeGaetano AT. 2006. Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate* **19**: 838–853.

Ducré-Robitaille JF, Boulet G, Vincent LA. 2003. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology* **23**: 1087–2003.

Easterling DR, Peterson TC. 1995. A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology* **15**: 369–377.

Ghorbanzadeh G, Lounes R. 2001. Bayesian analysis for detecting a change in exponential family. *Applied Mathematics and Computation* **124**: 1–15.

Goldstein JO. 2006. Subjective Bayesian analysis: principles and practice. *Bayesian Analysis* **1**: 403–420.

Heino R. 1997. Metadata and their role in homogenization. *Proceedings of the First Seminar for Homogenization of Surface Climate Data*, Hungarian Meteorological Service, Budapest, Hungary.

Holbert D. 1982. A Bayesian analysis of a switching linear model. *Journal of Econometrics* **19**: 77–87.

Kander Z, Zacks S. 1966. Tests procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *The Annals of Mathematical Statistics* **37**: 1196–1210.

Lee CB. 1998. Bayesian analysis of a change-point in exponential families with applications. *Computational Statistics and Data Analysis* **27**: 195–208.

Lee ASF, Heghinian SM. 1977. A shift of the mean level in a sequence of independent normal random variables. *Technometrics* **19**: 503–506.

Lund R, Wang XL, Lu Q, Reeves J, Gallagher C, Feng Y. 2007. Changepoint detection in periodic and autocorrelated time series. *Journal of Climate* **20**: 5178–5190.

Menne MJ, Williams CNJ. 2005. Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate* **18**: 4271–4286.

Müller-Westermeier G 2004. Statistical analysis of results of homogeneity testing and homogenization of long climatological time series in Germany. *Proceedings of the Fourth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary*, Report WCDMP 56, WMO-TD 1236, World Climata Data and Monitoring Programme. World Meteorological Organization: Geneva, Switzerland.

Perreault L, Haché M, Slivitzsky M, Bobée B. 1999. Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stochastic Environmental Research and Risk Assessment* **13**: 201–216.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Forland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D. 1998. Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology* **18**: 1493–1517.

Potter KW. 1981. Illustration of a new test for detecting a shift in mean in precipitation series. *Monthly Weather Review* **109**: 2040–2045.

Rasmussen P. 2001. Bayesian estimation of change points using the general linear model. *Water Resources Research* **37**: 2723–2731.

Reeves J, Chen J, Wang XL, Lund R, Lu Q. 2007. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology* **46**: 900–915.

Robert C. 1992. *L'analyse statistique Bayésienne*. Economica: Paris.

Seidou O, Asselin JJ, Ouarda TBMJ. 2007. Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables. *Water Resources Research* **43**: W08401.

Seidou O, Ouarda TBMJ. 2007. Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resources Research* **43**: W07404.

Smith AFM. 1975. A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62**: 407–416.

Szentimrey T. 1999. Multiple analysis of series for homogenization (MASH). *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*. Hungarian Meteorological Service: Budapest, Hungary.

Tang SM, MacNeill IB. 1993. The effect of serial correlation on tests for parameter change at unknown time. *The Annals of Statistics* **21**: 552–575.

Vincent LA. 1998. A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate* **11**: 1094–1105.

Wang XL, Wen QH, Wu Y. 2007. Penalized maximal t-test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology* **46**: 916–931.

Wang XL. 2008. Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F- test. *Journal of Applied Meteorology and Climatology* **47**: 2423–2444.

