Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows

O. Seidou[1,2] and T. B. M. J. Ouarda[1,2]

Received 8 March 2006; revised 4 January 2007; accepted 14 February 2007; published 4 July 2007.

[1] A large number of models in hydrology and climate sciences rely on multiple linear regression to explain the link between key variables. The relationship in the physical world may experiment sudden changes because of climatic, environmental, or anthropogenic perturbations. To deal with this issue, a Bayesian method of multiple changepoint detection in multiple linear regression is proposed in this paper. It is an adaptation of the recursion-based multiple changepoint method of Fearnhead (2005, 2006) to the classical multiple linear model. A new class of priors for the parameters of the multiple linear model is introduced, and useful formulas are derived that permit straightforward computation of the posterior distribution of the changepoints. The proposed method is numerically efficient and does not involve time consuming Monte-Carlo Markov Chain simulation as opposed to other Bayesian changepoint methods. It allows fast and straightforward simulation of the probability of each possible number of changepoints as well as the posterior probability distribution of each changepoint conditional on the number of changes. The approach is validated on simulated data sets and then compared to the methodology of Seidou et al. (2006) on two practical problems, as follows: (1) the changepoint detection in the multiple linear relationship between mean basin scale precipitation at different periods of the year and the summer-autumn flood peaks of the Broadback River located in Northern Quebec, Canada; and (b) the detection of trend variations in the streamflows of the Ogoki River located in the province of Ontario, Canada.

Citation: 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.

1. Introduction

[2] An increasing number of papers point out shifts or trends in hydrologic time series [e.g., *Burn and Hag Elnur*, 2002; *Woo and Thorne*, 2003; *Salinger*, 2005]. A change of mentality is taking place in the whole scientific community, and it is probable that hydrologic time series models which do not hold account of a possible change in the statistical distribution of the data will no longer be regarded as credible. Detection of eventual changes in collected data sets is thus obviously an important step before performing any descriptive or predictive analysis.

[3] Changepoint analysis is addressed both in Classical and Bayesian statistics. Methods in classical statistics usually consist of performing several kinds of tests to confirm or reject the hypothesis of change. Most of them address slope or intercept change in linear regression models [*Solow*, 1987; *Easterling and Peterson*, 1992; *Vincent*, 1998; *Lund and Reeves*, 2002; *Wang*, 2003].

[4] In Bayesian statistics, one is interested in obtaining a statistical distribution for the dates of change and eventually a distribution for the other model parameters. Bayesian changepoint analysis models are the subject of a large number of papers [e.g., *Booth and Smith*, 1982; *Bruneau and Rassam*, 1983; *Gelfand et al.*, 1990; *Barry and Hartigan*, 1992, 1993; *Stephens*, 1994; *Perreault et al.*, 2000a, 2000b, 2000c; *Rasmussen*, 2001]. Bayesian changepoint approaches were also applied to curve fitting by modeling the signal as a sequence of piecewise constant linear regression models [*Punskaya et al.*, 2002] or piecewise polynomial models [*Denison et al.*, 1998]. The inference on parameters was performed using Monte-Carlo Markov Chain algorithms (MCMC). More recently, *Seidou et al.* [2006] developed an approach to changepoint detection in multiple linear relationships, and *Fearnhead* [2005, 2006] proposed a recursion-based inference procedure on the basis of the theory of product-partition models [*Barry and Hartigan*, 1992, 1993] for multiple changepoint problems. In both papers, he provided solutions to perform direct simulation from the posterior distribution of a class of multiple changepoint models where the number of changepoints is unknown. He also provided efficient non-MCMC solutions for multiple changepoint detection in a single series in which the observations follow a normal or a Poisson distribution. As the models presented in the works of *Fearnhead* [2005, 2006] are similar, only the latter reference will be used in the remainder of the text. In the latter paper, a set of recursive relations are used to infer the posterior

[1]Hydro-Québec NSERC Chair in Statistical Hydrology, INRS-ETE, Quebec (QC), Canada.
[2]Canada Research Chair on the Estimation of Hydrological Variables, INRS-ETE, Quebec (QC), Canada.

0043-1397/07/2006WR005021$09.00

probabilities of different numbers of changepoints. A particularity of this approach is that it focuses only on the number and positions of changes.

[5] The aim of this paper is to adapt the methodology of *Fearnhead* [2006] to multiple changepoint detection in multiple linear relations. In particular, a special class of priors for the parameters of the multiple linear model is introduced, and useful formulas are derived that permit straightforward computation of the posterior distribution of the changepoints. The proposed methodology is validated on simulated data sets to prove its ability to infer the number and location of changepoints. It is then applied to two case studies. In the first case study, the summer-autumn flood peaks of the Broadback River located in the province of Quebec, Canada, are investigated for the eventual changes due to forest fires. The second case study deals with the detection of eventual trend variations in the streamflow data of the Ogoki River located in the province of Ontario, Canada.

[6] As the first case study has already been investigated with a changepoint detection approach using Gibbs sampling [*Seidou et al.*, 2006], the results obtained with the two methodologies will be compared and discussed in this paper. The approach of *Seidou et al.* [2006] will also be applied to the second case study in order to highlight the importance of having a methodology designed to handle several changepoints.

[7] The outline of the paper is as follows: section 2 is a quick survey of changepoint detection methodologies with an emphasis on Bayesian methodologies with application to hydrological problems. The two approaches that will be compared with the one proposed in this paper are described in sections 3 and 4. Recursion based changepoint inference models are introduced in section 5, and the model of *Fearnhead* [2006] is adapted to multiple linear regression. The simulation of changepoints given the conditional posterior probabilities of the dates of change is presented in section 6. The simulation-based validation methodology is presented in section 7. Section 8 presents the results of the simulation studies, and the applications on real data are carried out in section 9. A conclusion and some recommendations are finally presented in Section 10.

2. Changepoint Models

[8] Changepoint detection has received a great deal of attention in statistical literature because modification of model structure and/or parameters is commonly encountered in applied statistics (e.g., in finance, pharmacology, econometrics, hydrology, etc.). The change detection can be off-line (or retrospective) or online (or sequential) when it is important that the change be detected as soon as it occurs. Examples of online changepoint detection methods can be found in the works of *Lai* [1995], *Beibel* [1997], *Daumer and Falk* [1998], *Gut and Steinebach* [2002], and *Moreno et al.* [2005].

[9] Most applications in hydrology are used for retrospective changepoint detection, except a few ones [e.g., *Moreno et al.*, 2005]. Retrospective changepoint detection methods often use classical statistical methods to detect changes in slopes or intercepts of linear regression models [*Solow*, 1987; *Easterling and Peterson*, 1992; *Vincent*, 1998; *Rasmussen*, 2001; *Lund and Reeves*, 2002; *Wang*, 2003]. Other curve fitting methods are used in some rare cases [e.g., *Sagarin and Micheli,* 2001; *Bowman et al.*, 2004].

[10] A growing number of methodologies use Bayesian statistics. *Gelfand et al.* [1990] discussed Bayesian analysis of a variety of normal data models, including regression and analysis of variance type structures, where they allowed for unequal variances. *Barry and Hartigan* [1992, 1993] used product-partition models to develop a Bayesian analysis for a multiple changepoint problem that can be exactly solved using a finite number of operations. The multiple changepoint component was introduced by a normal random variable that can be added anytime to the mean of the series, but only with a certain probability. *Stephens* [1994] implemented Bayesian analysis of a multiple changepoint problem where the number of changepoints is assumed known, but the times of occurrence of the changepoints remain unknown. Other authors emphasized on the single changepoint problem. We cite for example *Carlin et al.* [1992] who applied a three-stage hierarchical Bayesian analysis to a simple linear changepoint model for normal data, $Y_t \sim N[a_1 + b_1 x_t, \delta_1^2]$, $t = 1,\ldots,\tau$, $Y_t \sim N[a_2 + b_2 x_t, \delta_2^2]$, $t = \tau + 1,\ldots,n$. *Perreault et al.* [2000a, 2000b] gave Bayesian analyses of several changepoint models of univariate normal data. All of these authors implemented their analyses using Gibbs sampling. *Rasmussen* [2001] considered a single changepoint in a simple linear regression model with noninformative priors and derived the exact analytical posterior distribution of the regression parameters. His model assumes that the changepoint occurred with certainty and does not allow a clear diagnosis of the existence of the change. *Perreault et al.* [2000c] developed an exact analytical Bayesian analysis of a changepoint in the mean of a series of multivariate normal random variables.

[11] More recently, *Seidou et al.* [2006] developed a practical and general approach to the single changepoint inference problem relying on Bayesian multivariate regression analysis. Their model can handle multivariate data and/or missing values and can be used with both informative and noninformative priors on the regression parameters. It was shown to be more performing than other approaches recently published in the hydrological literature [*Seidou et al.*, 2005]. However, the approach presented in the work of *Seidou et al.* [2006] considers only one possible changepoint and involves relatively long MCMC simulations. The method presented in this paper is expected to handle theses two issues.

3. The Changepoint Model of *Seidou et al.* [2006]

[12] The model developed in the work of *Seidou et al.* [2006] is designed to infer the position of a change in the parameters of a multivariate regression equation. They assume that the $(r \times 1)$ response vector $\mathbf{Y}_t$ is related to the $(r \times d^*)$ matrix $\mathbf{X}_t$ by

$$\mathbf{Y}_t = \mathbf{X}_t \boldsymbol{\theta}_t^{(\tau_c)} + \boldsymbol{\upsilon}_t \tag{1a}$$

where

$$\boldsymbol{\theta}_t^{(\tau_c)} = \begin{cases} \boldsymbol{\beta}_1^*, & 1 \le t < \tau_c, \\ \boldsymbol{\beta}_2^*, & \tau_c \le t \le n,. \end{cases} \tag{1b}$$

under the constraints

$$\boldsymbol{\beta}_1^* = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{\beta}_2^* = (\boldsymbol{\beta}_2, \boldsymbol{\beta}_0)^{\mathrm{T}}. \tag{1c}$$

[13] In these equations, as well as in the remainder of the paper, bold letters indicate vectors and matrices while the superscript "T" indicates the transpose. Note that in equation (1b), τ_c is the first point of the segment after the changepoint, and $\tau_c = (n + 1)$ means that there is no change in the data series. The notation is different from the one in the work of *Seidou et al.* [2006] where τ was defined as the last point before the change (that is, $\tau_c = \tau + 1$). It has been assumed that only part of the regression parameters change at τ_c. $\boldsymbol{\beta}_0$ is the vector of regression parameters that do not change, while $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ contain the values of the remaining parameters before and after the changepoint. $\boldsymbol{\beta}_1^*$ and $\boldsymbol{\beta}_2^*$ are the vectors of regression parameters before and after the change [cf. equation (1c)]. The dimensions of the vectors $\boldsymbol{\theta}_t^{(\tau_c)}$, $\boldsymbol{\beta}_1^*$, $\boldsymbol{\beta}_2^*$, $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ are $(d^* \times 1)$, $(d^* \times 1)$, $(d^* \times 1)$, $(d_0^* \times 1)$, $(d_1^* \times 1)$, and $(d_1^* \times 1)$, respectively. Equation (1c) implies that $d^* = d_0^* + d_1^*$. It is also assumed that error terms $\{\boldsymbol{\upsilon}_t\}$ are independent and identically distributed following $N[0, \Sigma_y]$.

[14] The model assumes a changepoint in the $(d^* \times 1)$ vector $\boldsymbol{\theta}_t^{(\tau_c)}$ from the $(d_1^* \times 1)$ subvector $\boldsymbol{\beta}_1$ to the $(d_1^* \times 1)$ subvector $\boldsymbol{\beta}_2$. The $(d_0^* \times 1)$ subvector $\boldsymbol{\beta}_0$ is assumed to remain part of $\boldsymbol{\theta}_t^{(\tau_c)}$ throughout the observation series.

[15] In the work of *Seidou et al.* [2006], some algebraic transformations allowed to apply some known results on Bayesian piecewise linear regression to equation (1a) and to infer its parameters. The MCMC algorithm was also designed to account for missing data in the observations record and/or in the explanatory variables. Finally, they considered a general prior specification for regression parameters as well as for the variance structure and used Gibbs sampling to obtain empirical posterior distributions for each parameter. For extensive details on prior specification and MCMC inference for equation (1a), we refer the reader to the original paper.

[16] Although the method is designed to detect only one changepoint, it can be readily applied to detect multiple changepoints using a segmentation approach. This approach consists in a recursive application of the changepoint detection methods to segments obtained by splitting the series at detected changepoints. The procedure is stopped all segments are found homogeneous or too short to be searched for changepoints.

4. The Changepoint Model of *Vincent* [1998]

[17] *Vincent* [1998] proposed a technique based on the Durbin-Watson test and the classical F test to detect a single changepoint in a multiple regression model. The original method uses a significance level p and is applied as follows:

[18] 1. For each year i in the observation period $[a, b]$

[19] a. Fit two linear regression models to segments $[a, i - 1]$ and $[i, b]$ and compute the series of residuals.

[20] b. Test the autocorrelation of the residuals using the Durbin-Watson test at p% significance level.

[21] c. If the test is positive, use the classical F test at p% significance level to compare the model with a changepoint at year i with the model without changepoint (see the work of *Vincent* [1998] for details on how to compute the test statistics and the associated critical values).

[22] 2. In case the two tests are positive for several years, consider the year with the higher F-test statistics as the date of change.

[23] To detect multiple changepoints, the segmentation method described in section 3 can be used.

5. Recursion-Based Changepoint Inference

[24] Although recursions have been used to make inference on the number of changepoints [*Yao*, 1984; *Barry and Hartigan*, 1992, 1993], this kind of approach has been less widely used than MCMC based inference. *Yao* [1984] was the first to show that Bayesian inference for a single shift in a normally distributed sample can be performed in a finite number of recursive operations. As the number of operations grows quickly when the length of the data series increases, he also proposed an approximate inference for which the number of operations is reduced to the order of sample size. *Barry and Hartigan* [1992, 1993] showed that the changepoint problem can be elegantly handled using product-partition models and generalized the results of *Yao* [1984] to multiple changepoints and more general prior assumptions. Product partition models assume that observations in a random partition of the data are independent and allow the data to weight the partitions that hold. The methodologies presented in these papers under this approach allow for an efficient computation of the posterior probability of different number of changepoints using recursive relations. *Fearnhead* [2006] used this kind of recursive relations to develop a general inference procedure for the number and positions of the changepoints.

5.1. General Inference Procedure for the Number and Positions of the Changepoints

[25] *Fearnhead* [2006] considered a class of multiple changepoint models for which the number of changes is unknown. Let $\{y_1, y_2, \ldots, y_n\}$ be the sample, n the sample size, m the number of changepoints, $\tau_0 = 0, \tau_1, \ldots, \tau_{m+1} = n$ the changepoints and $\mathbf{Y}_{i:j}$ the observations from time i to time j. We also denote $g(.)$ the probability distribution of the time interval between consecutive changepoints and $g_0(.)$ the probability distribution of the first changepoint. The jth segment is then $\mathbf{Y}_{(\tau\{j-1\}+1):\tau_j}$ with parameter $\boldsymbol{\Phi}_j$.

[26] Assuming that the observations are independents conditional on the changepoints and parameter values, *Fearnhead* [2006] derived the posterior probability of the changepoints:

$$\begin{cases} \Pr(\tau_1 | \mathbf{Y}_{1:n}) = P(1, \tau_1) Q(\tau_1 + 1) g_0(\tau_1) / Q(1) \\ \Pr(\tau_j | \tau_{j-1}, \mathbf{Y}_{1:n}) = P(\tau_{j-1} + 1, \tau_1) Q(\tau_j + 1) g(\tau_j - \tau_{j-1}) / Q(\tau_{j-1} + 1) \end{cases} \tag{2}$$

where $P(t, s)$, $s \geq t$ is the probability that t and s be in the same segment:

$$P(t,s) = \Pr(\mathbf{Y}_{t:s}; t, s \text{ in the same segment}) = \int \prod_{i=t}^{s} f(y_i|\Phi)\pi(\Phi)\mathrm{d}\Phi \quad (3)$$

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

$$\begin{cases} Q(1) = \sum_{s=1}^{n-1} P(1,s)Q(s+1)g_0(s) + P(1,n)(1 - G_0(n-1)) \\ Q(t) = \sum_{s=t}^{n-1} P(t,s)Q(s+1)g_0(s+1-t) + P(t,n)(1 - G(n-t)). \end{cases} \quad (4)$$

where $G(t) = \sum_{i=1}^{t} g(i)$ and $G_0(t) = \sum_{i=1}^{t} g_0(i)$.

5.2. Adaptation of the Changepoint Inference Procedure to Multiple Linear Regression

[27] In this section, we are going to provide analytical formulation for $P(t, s)$. Consider the $n_p + 1$ series of data y_j, $j = 1,...n$ and x_{ij}, $i = 1,...,d^*$; $j = 1,...n$ where x_{ij} is the jth value of the ith series of explanatory variables. The multiple linear relationship can be represented by

$$y_j = \sum_{k=1}^{d^*} \theta_k x_{kj} + \varepsilon_i \qquad = 1, \ldots, n. \quad (5)$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (6)$$

The parameter vector $\boldsymbol{\Phi}$ is thus given by $\boldsymbol{\Phi} = [\boldsymbol{\theta}_1 \boldsymbol{\theta}_2 ... \boldsymbol{\theta}_{d^*}; \sigma]$ and we have:

$$f(y_i|\Phi) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-0.5 \left(\frac{y_i - \sum_{j=1}^{d^*} \theta_j x_{ij}\sigma}{} \right)^2 \right) \quad (7)$$

Following *Rasmussen* [2001], we have:

$$\Pr(\mathbf{y}_{t:s}|\Phi) = \prod_{i=1}^{s} f(y_i|\Phi) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\frac{(\mathbf{Y}_{t:s} - \mathbf{X}_{t:s}\boldsymbol{\theta})^{\mathrm{T}}(\mathbf{Y}_{t:s} - \mathbf{X}_{t:s}\boldsymbol{\theta})}{2\boldsymbol{\sigma}^2} \right] \quad (8)$$

[28] The next step in the process is the prior definition for the parameters. Since there is no knowledge about the range of the regression and variance parameters, usual improper noninformative priors such as the one used in the work of *Rasmussen* [2001] seem to be appropriate. However, direct use of classical improper prior in model selection problems, especially those using Bayes Factors, is problematic because of the well known difficulty that when the models or hypothesis have parameters of differing dimensions, one cannot directly use improper noninformative priors for computing Bayes factors; improper priors are unaffected by multiplication by an arbitrary positive constant, but such arbitrary constant directly affect Bayes factors [*Berger and Pericchi*, 1998]. Possible solutions include the use of intrinsic Bayes factors (IBF), fractional Bayes Factors (FBF), or asymptotic methods. The principle of IBF is to use part of the data (the training sample) to turn the improper prior into a new prior applicable to the remaining part of the data. In this paper, an approach similar to the use of IBF was used. Assume that we have an eventually improper prior $\pi_1(\Phi)$ and two l_1- and l_2-long training samples $\{\mathbf{X}_{(1-l_1):0}; \mathbf{Y}_{(1-l_1):0}\}$ and $\{\mathbf{X}_{(n+1):(n+l_2)}; \mathbf{Y}_{(n+1):(n+l_2)}\}$ such as $L(\{(1 - l_1):0\}) = \int_{\mathrm{F}} \prod_{i=1-l_1}^{0} f(y_i|\Phi)\pi_1(\Phi)\mathrm{d}\Phi$ and $L(\{(n+1):(n+l_2)\} = \int_{\mathrm{F}} \prod_{i=n+1}^{n+l_2} f(y_i|\Phi)\pi_1(\Phi)\mathrm{d}\Phi$ are finite.Then prior distribution $\pi(\Phi) = \frac{\prod_{i=1-l_1}^{0} f(y_i|\Phi)\pi_1(\Phi)}{2L(\{(1-l_1):0\})} + \frac{\prod_{i=n+1}^{n+l_2} f(y_i|\Phi)\pi_1(\Phi)}{2L(\{(n+1):(n+l_2)\})}$ is proper since $\int_{\mathrm{F}} \pi(\Phi)\mathrm{d}\Phi = 1$. It can thus be used for selecting the model for the remaining of the data. We have:

$$\begin{aligned} P(t,s) &= \int_{\Phi} \prod_{i=1}^{n} f(y_i|\Phi)\pi(\Phi)\mathrm{d}\Phi \\ P(t,s) &= \frac{\int_{\Phi} \left(\prod_{i=1-l1}^{0} f(y_i|\Phi) \right) \left(\prod_{i=t}^{s} f(y_i|\Phi) \right) \pi_1(\Phi)\mathrm{d}\Phi}{2L(\{(1-l_1):0\})} \\ &\quad + \frac{\int_{\Phi} \left(\prod_{i=t}^{s} f(y_i|\Phi) \right) \left(\prod_{i=n+1}^{n+l_2} f(y_i|\Phi) \right) \pi_1(\Phi)\mathrm{d}\Phi}{2L(\{(n+1):(n+l_2)\})} \\ P(t,s) &= \frac{P_1(\{(1-l_1):0\} \cup \{t:s\})}{2P_1(\{(1-l_1):0\})} \\ &\quad + \frac{P_1(\{t:s\} \cup \{(n+1):(n+l_2)\})}{2P_1(\{(n+1):(n+l_2)\})} \end{aligned} \quad (9)$$

With $P_1(\{i_1, i_2, ..., i_v\}) = \int_{\Phi} \prod_{i \in \{i_1, i_2, ..., i_v\}} f(y_i|\Phi)\pi_1(\Phi)\mathrm{d}\Phi$

Let's assume first that the prior depends only on σ and has this particular form:

$$\pi_1(\Phi) = \pi_1(\sigma) = p(\sigma|a, C) = \frac{\sigma^{-a} \exp\left(-\frac{c}{2\sigma^2}\right)}{2^{\frac{a-3}{2}} c^{-\frac{a-1}{2}} \mathrm{G}\left(\frac{a-1}{2}\right)}, a > 1, c > 0 \quad (10)$$

In equation (10), the denominator $2^{\frac{a-3}{2}} c^{-\frac{a-1}{2}} \Gamma\left(\frac{a-1}{2}\right)$ is only a normalizing constant that ensures that $\int_0^{+\infty} \pi_1(\sigma) d\sigma = 1$. Note that when σ is very large, $p(\sigma)$ tends toward a multiple of σ^a. Jeffrey's noninformative prior for linear regression is $p(\boldsymbol{\theta}, \sigma) \propto \sigma^{-\frac{d^*}{2}}$ (P. Minka, Bayesian linear regression, unpublished paper, 2001, online https://research.microsoft.com/~minka/papers/minka-linear.ps.gz), and it is sometimes assumed in Bayesian linear regression that $p(\sigma) \propto \sigma^{-1}$ [e.g., *Rasmussen*, 2001]. Unfortunately, these kinds of priors are improper contrarily to the one proposed in equation (10). Basic properties of $p(\sigma|a, c)$ are derived in Appendix A.

[29] Using the above definition of the prior π_1, it is shown in Appendix B that

$$P_1(\{i_1, i_2, \ldots, i_{n_1}\}) = (2\pi)^{\frac{d^*}{2}} \cdot \frac{\left(\pi\left(\varepsilon^{\mathrm{T}}_{\{i_1,i_2,\ldots,i_{n_1}\}} \varepsilon_{\{i_1,i_2,\ldots,i_{n_1}\}} + C\right)\right)^{-\frac{(n_1+a-1)}{2}}}{(C\pi)^{-\frac{a-1}{2}} \left|\mathbf{X}^{\mathrm{T}}_{\{i_1,i_2,\ldots,i_{n_1}\}} \mathbf{X}_{\{i_1,i_2,\ldots,i_{n_1}\}}\right|^{1/2}} \cdot \frac{\Gamma\left(\frac{n_1+a-d^*}{2}\right)}{\Gamma\left(\frac{a-1}{2}\right)} \tag{11}$$

In equation (11), d^* is the number of explanatory variables (including the intercept if any), $\varepsilon_{\{i1,i2,\ldots,i_{n1}\}}$ is the vector of residuals of the linear relationship between $\mathbf{X}_{\{i1,i2,\ldots,i_{n1}\}}$ and $\mathbf{Y}_{\{i1,i2,\ldots,i_{n1}\}}$.

6. Simulation of Changepoints Given the Conditional Posterior Probabilities of the Changepoints

[30] The relations presented in section 4 give only the posterior probability mass of the first changepoint and the conditional probability mass of subsequent changepoints. To make inference on the positions of changepoints, we simulate a set $E = \{S_k, k = 1{:}M\}$ of M possible scatter schemes of the changepoints on the segment using the posterior probability mass of the first changepoint and the conditional probability mass of subsequent changepoints. Indeed, M should be large enough to obtain a reliable distribution for the positions of the changepoints. The kth element of E is a set of m_k changepoints $S_k = \{\tilde{t}_1^k, \tilde{t}_2^k, \ldots, \tilde{t}_{\tilde{m}_k}^k\}$. An efficient simulation algorithm for E is given by *Fearnhead* [2006]:

[31] 1. For a sample of size M, initiate M samples with a changepoint at $t = 0$.

[32] 2. For $t = 0, \ldots, n-2$, repeat the following steps:

[33] a. Compute the number n_t of samples for which the last changepoint was at time t;

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

[35] c. 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 changepoints which have a changepoint at time t;

[36] What is done at each iteration in this algorithm is basically the recursive generation of new changepoints using the posterior conditional distribution $\Pr(\tau|\tau_{j-1} = t, \mathbf{y}_{1:n})$. At the end of the process, a set of M samples representative of the joint posterior probability of the number and positions of the changepoints is obtained. For more details on this algorithm, the reader is referred to the work of *Fearnhead* [2006]. 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 changepoints is readily carried out using the M samples. For instance, the probability of having i changepoints is approximated by:

$$\Pr(m = i) \approx \mathrm{card}(\{k|\mathrm{card}(S_k) = i\})/M \tag{12}$$

[37] The posterior probability of having the kth changepoint at position t given m changepoints can be approximated by:

$$\Pr(\tau_i = t|m) \approx \frac{\mathrm{card}\left(\left\{k|(\mathrm{card}(S_k) = m) \& \left(\tilde{t}_i^k = t\right)\right\}\right)}{\mathrm{card}(\{k|\mathrm{card}(S) = m\})} \tag{13}$$

where card(S) stands for the number of elements of the set S. The estimators of the number and positions of changepoints are the modes of their posterior distributions, i.e.:

$$\hat{m} = \mathrm{Max}_t\ \mathrm{card}(\{k|\mathrm{card}_k(S) = t\})/M\} \tag{14}$$

$$\hat{\tau}_i = \underset{t}{\mathrm{Max}} \left\{\frac{\mathrm{card}\left(\left\{k|(\mathrm{card}(S_k) = \hat{m}) \& \left(\tilde{t}_i^k = t\right)\right\}\right)}{\mathrm{card}(\{k|\mathrm{card}(S) = \hat{m}\})}\right\} \tag{15}$$

Other estimators can be defined using the posterior distributions. However, in Bayesian analysis, the mode of the posterior is generally preferred to its mean when an estimator is needed.

7. Validation Methodology

[38] The validation of the proposed method requires large data sets in which all the characteristics of the changepoints are known. These data sets were obtained by simulation using a procedure that mimics the ranges of shifts and trends that are usually observed in streamflow data. The ability of the proposed method to correctly detect the number and position of changes was assessed using four performance measures that are described further in the text.

7.1. Simulated Data Sets

[39] Artificial shifts and trends with random magnitudes and positions were inserted in three sets of simulated normal series. The number of changes is constrained to be less than $m_{\max}$. The first set contains series which only display shifts in the mean. The series in the second set contain abrupt changes of trend, while the changepoints in the third set can be either shifts or changes in trend.

[40] The series in the first data set were simulated in the following manner:

[41] 1. Set the number of series to generate (N) the minimum number of points between changepoints ($l_{\min}$) and the maximum magnitude of the shift $\delta_{\max}$;

[42] 2. Set u to 1;

[43] 3. Simulate a set $\mathbf{Y}_u = \{y_i, i = 1,\ldots,n\}$ of n random numbers from the normal distribution with mean 0 and standard deviation 1;

[44] 4. Simulate the number of changes by uniformly drawing a number m in $\{0, 1,\ldots,m_{\max}\}$;

[45] 5. For each i in $\{1,\ldots,m\}$, if $n - l_{\min} - \tau_{i-1} > 0$, uniformly draw a changepoint position τ_i in $\{\tau_{i-1} + l_{\min},\ldots,n\}$. Repeat this step until τ_m is sampled;

[46] 6. For each i in $\{1,\ldots,m\}$, if $n - l_{\min} - \tau_{i-1} > 0$, uniformly draw a shift magnitude δ_i in $[-\delta_{\max}, \delta_{\max}]$;

[47] 7. For each i in $\{1,\ldots,m\}$, set $y_k = y_k + \delta_i$, $k = \tau_i + 1,\ldots,n$;

[48] 8. If $u < N$, increment u and return to step 3, otherwise end the simulation procedure.

[49] The second data set is generated in the same manner except that trend changes rather than shifts are introduced in the series. In that case, if we denote tr_i the trend in the $(i + 1)$th first segment, all the above listed steps hold, except the seventh step that should be replaced by this one:

[50] 7a. For each i in $\{0,\ldots,m\}$, set, $y_v = y_v + \mathrm{tr}_i$ $(x_k - x_{\tau_i}+1), v = \tau_i + 1,\ldots,n$.

[51] In the third data set, the changes can either be a shift in the mean or a change of trend. The type of change is randomly selected using a binomial distribution with parameter 0.5.

7.2. Performance Measures

[52] Let's denote m_u the number of changepoints in the uth generated sample $\{\mathbf{Y}\}_u$ and $\{t_i^k, i = 1{:}m_u\}$ their positions. Let $\tilde{m}_u$ be the estimate of m_u, and $\{\hat{t}_i^k, i = 1{:}\hat{m}_u\}$ the estimates of the positions of the $\hat{m}_u$ detected changepoints. Two simple measures of the ability of the proposed approach to detect the number of changepoints are the Percentage of Correct Detections of the Number of changepoints (PCDN) and the Root Mean Square Error (RMSE) of the estimations of the number of changepoints defined as follow:

$$\mathrm{PCDN} = \frac{1}{M}\sum_{u=1}^{M} 1_{\{\tilde{m}_u = m_u\}} \tag{16}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{u=1}^{M}(\tilde{m}_u - m_u)^2} \tag{17}$$

Another measure of the capability of the method to correctly estimate the number of changepoints is the Ranked Probability Score (RPS), if F_u denotes the empirical cumulative probability distribution of m_u obtained with the application of the changepoint detection method, the RPS can be defined as follows:

$$\mathrm{RPS} = \frac{1}{M}\sum_{u=1}^{M}\sum_{i=1}^{n}(F_u(i) - 1_{i\geq m_u})^2 \quad \text{where} \quad 1_{i\geq m_u} = \begin{cases} 1 & \text{if } i \geq m_u \\ 0 & \text{if } i < m_u \end{cases} \tag{18}$$

[53] The RPS is usually used to rate ensemble forecasts [e.g *Buizza and Palmer*, 1998; *Hamil*, 2001]. The RPS values are within $[0, n - 1]$, and a value of zero is obtained for perfect forecasts.

[54] Unfortunately, the RPS is designed to rate the prediction for a single variable and cannot be easily applied to the estimators of the positions of changepoints, as the number of detected changepoints may be different from the real number of changepoints. A new performance measure was thus developed as follows: let $\{\mathbf{Y}\}_u$ be a series generated as described in section 6 with m_u changepoints $\{t_j^u, j = 1{:}m_u\}$. The application of the changepoint detection approach to $\{\mathbf{Y}\}_u$ will give a set $E = \{S_k, k = 1{:}M\}$ of M possible scatter schemes where $S_k = \tilde{t}_1^k, \tilde{t}_2^k,\ldots,\tilde{t}_{mk}^k\}$ has $\tilde{m}_k$ elements. $\tilde{m}_k$ may be different from the real number of changes m_u in $\{\mathbf{Y}\}_u$. Given k and u, consider $\{a_i, i = 1,\ldots,\min(\tilde{m}_k, m_u)\}$ and $b_i, i = 1,\ldots,\min(\tilde{m}_k, m_u)\}$ such that $i \neq j \rightarrow a_i \neq a_j, i \neq j \rightarrow b_i \neq b_j$ and $\sum_{i=1}^{\min(\tilde{m}_k, m_u)}\left(t_{a_i^u} - \tilde{t}_{b_i^k}\right)^2$ is minimal. The performance of the changepoint detection method when applied to the generated series $\{\mathbf{Y}\}_u$ can be measured with the Multiple Change Detection Performance Index (MCDPI) defined as

$$\mathrm{MCDPI}_k = \begin{cases} \dfrac{1}{\tilde{m}_k}\displaystyle\sum_{i=1}^{\tilde{m}_k}\left(t_{a_i}^u - \tilde{t}_{b_i}^k\right)^2, & \tilde{m}_k = m_u \\ \dfrac{1}{\tilde{m}_k}\left(\displaystyle\sum_{i=1}^{m_u}\left(t_{a_i}^u - \tilde{t}_{b_i}^k\right)^2 + \sum_{j\neq b_i, i=1,\ldots,m_u}\tilde{t}_j^k\left(n - \tilde{t}_j^k\right)\right), & \tilde{m}_k > m_u \\ \dfrac{1}{m_u}\left(\displaystyle\sum_{i=1}^{\tilde{m}_k}\left(t_{a_i}^u - \tilde{t}_{b_i}^k\right)^2 + \sum_{j\neq a_i, i=1,\ldots,\tilde{m}_k} t_j^u\left(n - t_j^k\right)\right), & \tilde{m}_k m_u \end{cases} \tag{19}$$

The introduction of a_i and b_i is motivated by the need to associate as much as possible each element of the set of real changepoints to an element of the set of detected changepoints. Note that $\{a_i, i = 1,\ldots,\min(\tilde{m}_k, m_u)\}$ and $\{b_i, i = 1,\ldots,\min(\tilde{m}_k, m_u)\}$ are different for each pair (u, k). This association is performed using a minimum square distance criterion. The penalty term for the false detection of a change $\tilde{t}_j^k$ is $\tilde{t}_j^k(n - \tilde{t}_j^k)$; the penalty for the nondetection of the change t_j^u is $t_j^u(n - t_j^u)$. These penalty terms have the interesting property of not over-penalizing false detections at the beginning and at the end of the series. They are consistent with the practice of discarding detected changes that are close to the end or the beginning of the series [*Beaulieu et al.*, 2005].The overall performance is the mean of the criterion over the set of generated series

$$\mathrm{MCDPI} = \frac{1}{N}\sum_{k=1}^{N}\mathrm{MCDPI}_k \tag{20}$$

8. Settings and Results of the Simulation Studies

[55] The prior for σ and the parameters for the data generation algorithms were first chosen to have a noninformative prior. Three data sets were generated according to the procedure described in section 6, and changepoints are identified with the proposed procedure. A two-column vector of explanatory variables was considered,

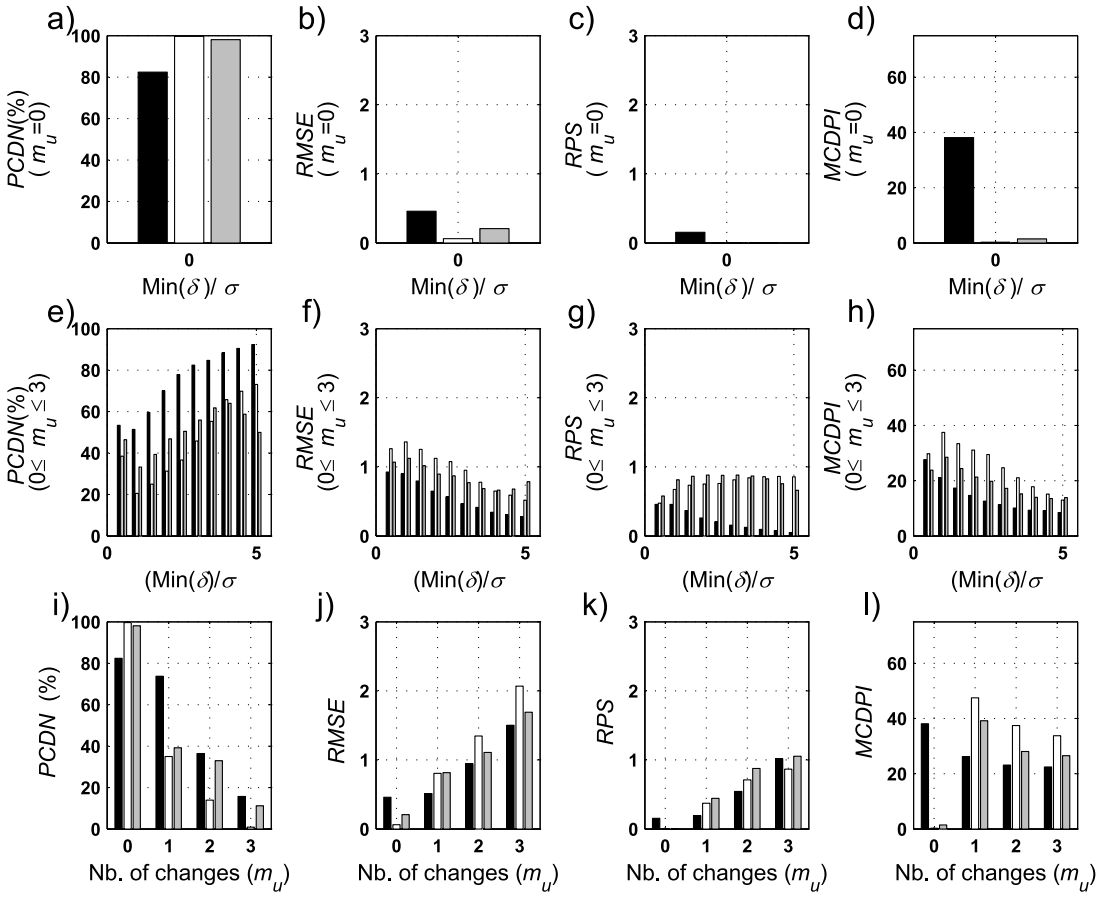

Figure 1. Performance of the changepoint detection procedure as function of the number of real changepoints and the minimum magnitude of the shift for the first simulated set, proposed method with training samples of length (black), method of *Vincent* [1998] with p = 5% (white bars), and method of *Vincent* [1998] with p = 1% (grey bars).

the first one containing only ones and the second containing the date of the observation.

8.1. Prior Specification for σ

[56] It can easily be shown that the prior variance of σ (equation A5 of Appendix A) is infinite when $a < 3$. Any value lower than 3 is thus a relatively noninformative prior. We chose $a = 2$ to be consistent with the classical $p(\sigma) \propto \sigma^{-2}$ usually used in Bayesian linear regression. As in equation (11), c has the dimension of a variance; it was set to the variance obtained by least squares estimates of the linear regression equations between the response variable and the explanatory variables in the stacked training samples, i.e.:

$$\begin{aligned} c &= \varepsilon^{\mathrm{T}}_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}}\varepsilon_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}} \\ &= Y^{\mathrm{T}}_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}}Y_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}} \\ &\quad - X_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}} \\ &\quad \times \left(X^{\mathrm{T}}_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}}X_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}}\right)^{-1} \\ &\quad \cdot Y^{\mathrm{T}}_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}}Y_{\{(1-l_1):0\}\cup\{(n+1):(n+l_2)\}}. \end{aligned} \tag{21}$$

8.2. Parameters of the Simulations

[57] The number of series in each of the three simulated data sets was set to 1000. The length of the series was fixed to 75. The number of changepoints varies from zero to three with at least 10 epochs between changepoints, and the shifts were assumed to have a magnitude ranging between zero and five times the standard deviation of the data series. The magnitudes of the trends are assumed inferior to 3 standard deviations per 10 epochs. These values are consistent with the authors experience with changes observed in streamflows data series.

8.3. Performance of the Proposed Method on Simulated Data Sets

[58] The changepoint detection method along with the method of *Vincent* [1998] using significance values p = 5% and p = 1% were applied to each simulated data set with a two-column vector of explanatory variables. The first column of this vector contains only ones while the second column contains the dates of the observations. Including the dates of observations in the vector of explanatory variables allows the detection of changes in trend in the data series. The length of the training samples were set to $l_1 = l_2 = 2$

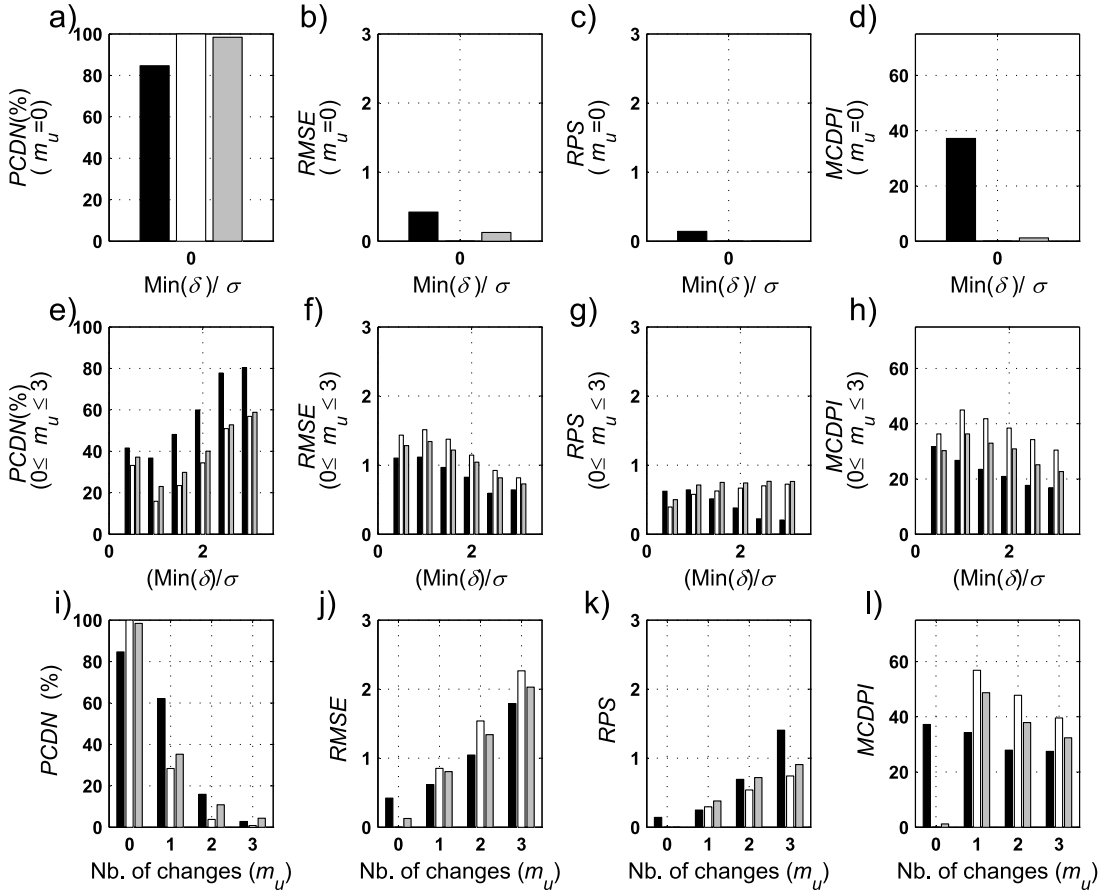

Figure 2. Performance of the changepoint detection procedure as function of the number of real changepoints and the minimum magnitude of the shift for the second simulated set, proposed method with training samples of length (black), method of *Vincent* [1998] with p = 5% (white bars), and method of *Vincent* [1998] with p = 1% (grey bars).

whish are the minimal values to have $L(\{(1 - l_1):0\})$ and $L(\{(n + 1):(n + l_2)\})$ well defined. The performances of the changepoint detection methods on the first two simulated data sets were compiled as a function of the number of real changepoints and the minimum magnitude of the change in a given series. Similar results were compiled for the third simulated data set but only using the number of changepoints since the series contained two kinds of changes with different definitions of the magnitude. These results are plotted in Figure 1 and Figure 2 for the first and second simulated data set, respectively. The same results are presented in Figure 3 for the third simulated data set. Analysis of these results allows drawing the following conclusions:

[59] 1. For the two first data sets, the rate of false detection (Figures 1a–1d; Figures 2a–2d) is lower for the proposed method as compared with the method of *Vincent* [1998]. Overall, the proposed method is the best for all data sets when $0 \leq m_u \leq 3$ (Figures 1e–1h; Figures 2e–2h). As expected, the PCDN increases when the minimum magnitude of the change increases. The same conclusions can be drawn from all the other performance measures considering that a good forecast means small RMSE, RPS, and MCDPI values.

[60] 2. The performance indices (except the MCDPI) decrease with the number of changepoints (cf. Figures 1i–1l; Figures 2i–2l; Figures 3a–3d);

[61] 3. It seems easier for the method to detect shifts than changes in trend (Figure 1 versus Figure 2), although the relative performance depends on the range of change of magnitude in each set. This conclusion holds only if we consider that the range of magnitudes that were generated is representative of the real world.

[62] The reasons for which the changepoint method may miss a changepoint are multiple and cannot be illustrated with a single figure. They include the magnitude, nature, relative position, and sign of the changepoints. A thorough study of the influence of these factors can be found in the work of *Beaulieu et al.* [2005], and the results are too extensive to be presented in this paper. For instance, it is easier for the method to detect changepoints that are far apart than changepoints that are grouped. Also, low magnitude shifts are often ignored. Low magnitude changepoints are even less likely to be detected when they are close to higher magnitude changepoints.Results suggest that in this particular case (series of 75 years), the method can be trusted if the shifts in the data set have the order of magnitude of the standard deviation and if the number of

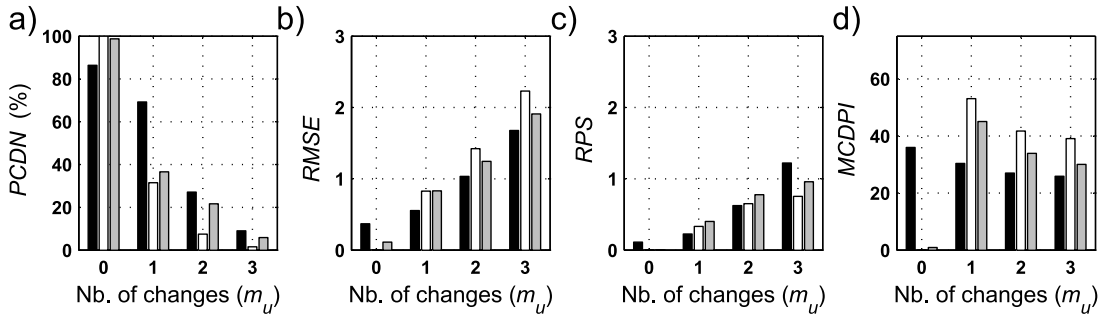

Figure 3. Performance of the changepoint detection procedure as function of number of real changepoints and the minimum magnitude of the shift for the third simulated set, proposed method with training samples of length (black), method of *Vincent* [1998] with $p = 5\%$ (white bars) and *Vincent* [1998] with $p = 1\%$ (grey bars).

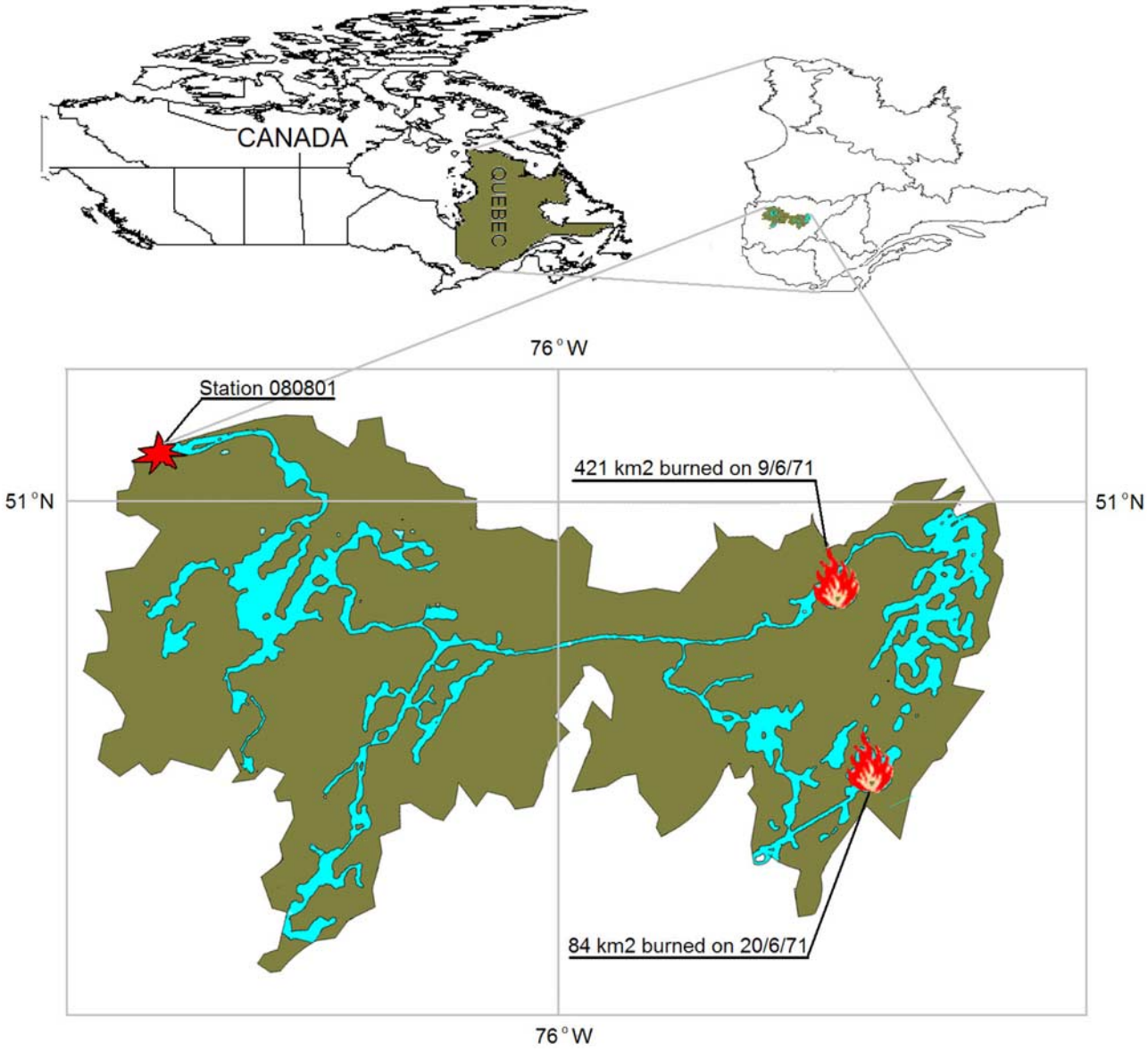

Figure 4. Location map of station 080801 (Broadback River).

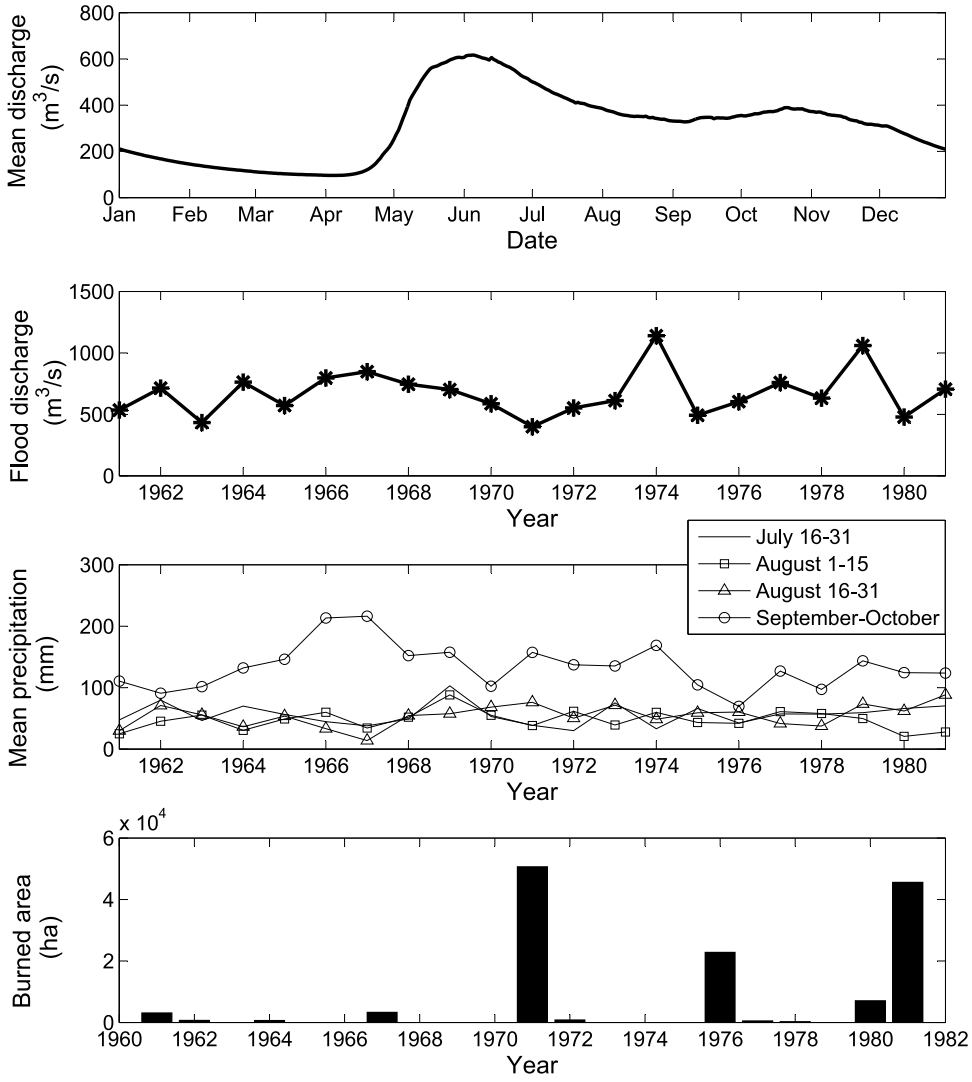

Figure 5. Data for changepoint detection in summer-autumn flood peaks for the Broadback River, (a) mean hydrograph, (b) summer-autumn flood peak time series, (c) precipitation time series, (d) burned catchment area time series.

changes is known to be inferior to three. Indeed, the performance should not be the same for other data sets with different lengths and different statistical characteristics. However, since the data sets were generated to cover the range of magnitudes generally encountered in streamflow records, the method proposed in this paper will be useful for detecting changes in river discharges. It can also be used in several other problems involving multiple linear regression, such as data homogenization or signal processing.

9. Application to Cases Studies

[63] The methodology is applied herein to two case studies to illustrate its behavior on real data and to compare it to the approaches of *Seidou et al.* [2006] and *Vincent* [1998]. The first case study deals with change detection in the linear regression describing the relationship between summer-autumn flood peaks and precipitations on the Broadback River basin. *Seidou et al.* [2006] studied this data set using the Bayesian single changepoint detection and found that the relation has significantly changed after 1972 (τ = 1972 $\Longleftrightarrow$ τ_c = 1973). The expected value of τ_c with the approach proposed in this paper should thus be 1973.

[64] The second case study is an example drawn from the Canadian Reference Hydrometric Basin Network (RBHN) database [*Brimley et al.*, 1999]. The case was selected because it displayed a relatively large number of changes.

9.1. Changepoint Detection in the Linear Regression Describing the Relationship Between Summer-Autumn Flood Peaks and Precipitations on the Broadback River Basin

9.1.1. Data

[65] The Broadback River has a catchment of 17,100 km^2 and experiences forest fire bursts from time to time (Figure 4). According to the Canadian Large Fire Database [Natural Resources Canada, Canadian large fires database, online document http://fire.cfs.nrcan.gc.ca/Downloads/LFDB/LFD_5999_e.ZIP, downloaded on August 2005; *Stocks et al.*, 2002], major forest fires occurred during the summer

Table 1. Basin Scale Precipitation and Summer-Autumn Flood Peak Time Series for the Broadback River Basin

Year	Total Precipitation for the 16th–31st July Period, mm	Total Precipitation for the 1st–15th August Period, mm	Total Precipitation for the 16th–31st August Period, mm	Total Precipitation for the September–October Period, mm	Summer-Autumn Maximum Flood Peak, m^3/s
1961	47.60	24.99	29.85	110.71	535
1962	79.61	45.34	70.96	90.98	714
1963	46.52	55.41	55.76	101.69	433
1964	69.96	30.52	36.23	132.00	762
1965	56.37	49.07	53.60	146.21	572
1966	44.56	59.93	33.27	213.33	796
1967	37.91	34.25	13.84	216.20	847
1968	49.04	52.02	54.45	152.14	745
1969	102.94	88.15	57.50	157.51	702
1970	53.04	55.06	68.32	102.24	586
1971	38.67	38.29	76.19	157.44	399
1972	29.98	61.48	50.26	137.10	552
1973	75.31	39.16	71.57	135.31	612
1974	33.14	59.81	48.58	168.72	1140
1975	66.11	43.33	59.15	104.56	493
1976	42.46	41.89	60.29	69.45	603
1977	57.16	61.02	41.64	126.90	759
1978	56.95	57.92	37.51	97.12	632
1979	59.22	49.73	73.62	143.59	1060
1980	66.02	20.74	61.98	124.47	478
1981	70.38	27.73	88.40	123.76	705

of 1971, burning 506 km^2 in the upper parts of the catchment (1/34 of the total basin area). It can be hypothesized that the deforestation because of these fires can change the basin response function to meteorological inputs. In order to perform the analysis, the 1961–1981 daily maximum flood discharges at station 80801 were obtained from Quebec Ministry of the Environment. The Broadback River is subject to two types of floods, spring floods, which are dominated by snowmelt, and summer-autumn floods which are caused by direct liquid precipitations. Figure 5a presents the mean daily discharge at this station for the 1961–1981 period. It appears that the summer-autumn maximum flood peak is generally observed at the end of October (Figure 5a). Daily precipitations for

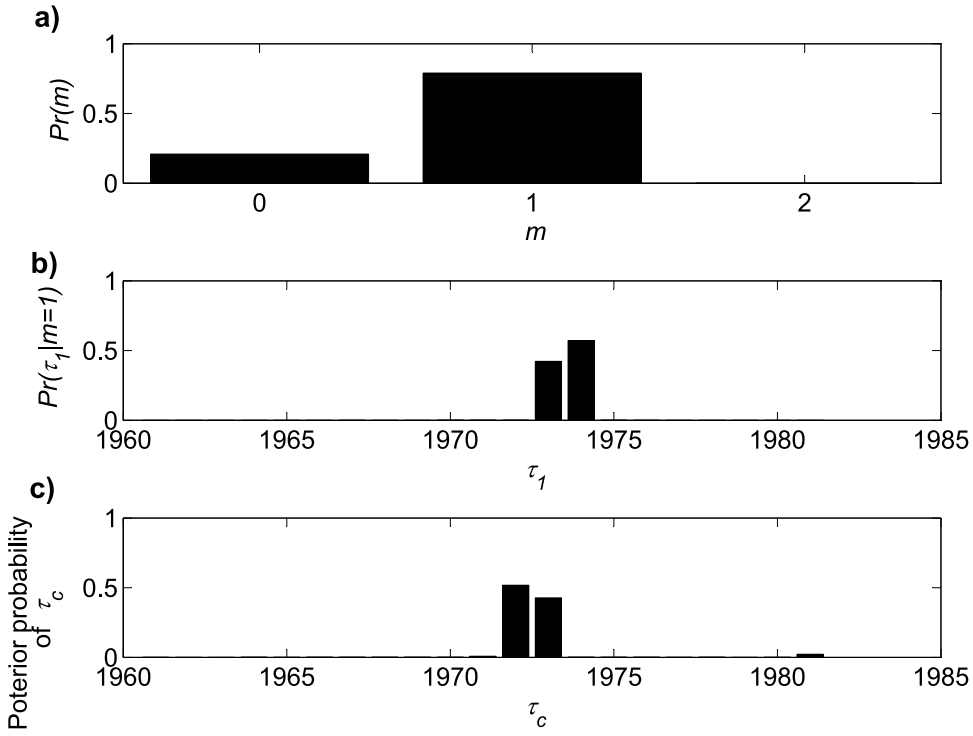

Figure 6. Changepoint detection in summer-autumn flood peaks of the Broadback river, (a) posterior probability of the number of changepoints, (b) posterior probability of the first point of the segment after the changepoint obtained with the proposed methodology, (c) posterior probability of the last point of the segment before the change obtained with a recursive application of the methodology of *Seidou et al.* [2006].

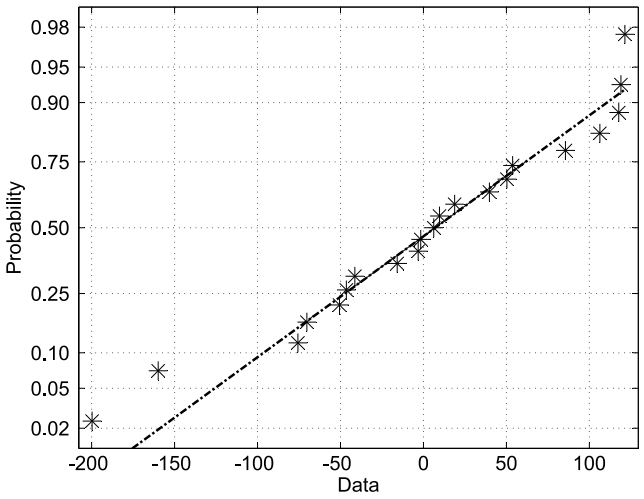

Figure 7. Normal probability plot of the residuals given (Broadback river).

the July–October period from 1961 to 1981 were obtained by interpolation from the neighboring weather stations on a regularly spaced grid of 100 × 100 points and averaged to have a time series representing precipitation at the catchment scale. This time series was then used to obtain the sum of precipitations on the Broadback river catchment for every half-month period from July to October. Exploratory analysis of the linear relationship between observed flood discharge and the obtained precipitation series led to the choice of four explanatory variables for the flood peak values, (1) the sum of precipitations of the 16th–31st of July period, (2) the sum of precipitations of the 1st–15th of August period, (3) the sum of precipitations of the 16th–31st of August period, and (4) total precipitations for the September–October period. The values of the 1961–1981 summer-autumn flood peaks are presented in Figure 5b, and those of the chosen explanatory variables are presented in Figure 5c. Figure 5d presents the burned areas on the catchment for each year of the period of study. The series of explanatory variables as well as the maximum flood peaks are summarized in Table 1.

9.1.2. Results

[66] The application of the proposed changepoint detection method leads to a probability of nearly 0.3 for the absence of changepoints, 0.7 for the existence of a unique changepoint (Figure 6a). A small weight (<0.1) is attributed to the existence of two changes. The posterior probability distribution of the changepoint τ is illustrated in Figure 6b. The posterior probability distribution of τ_c obtained with the same data set by *Seidou et al.* [2006], using another Bayesian method, is also presented in Figure 6c. The two Bayesian methods agree that the changepoint occurred probably between 1973 and 1974, with however different weights for these two dates. The method of *Vincent* [1998] with p = 5% also gave one changepoint in 1973 (potential dates of changes were 1973 and 1974, with a maximum F statistics for 1973). Hence the three methods agree that a change occurred in 1973 or 1974. The differences in posterior probabilities displayed by the two Bayesian methods may be due to the differences in the prior specifications of the two methods. However, this point cannot be proven since it is not possible to interchange the prior formulations of the two Bayesian models.

[67] All three methods are based on the assumption of normal residuals. It is hard to check the normality of residuals in a Bayesian context, but it can be tested for classical methods. Figure 7 presents the normal probability plots of the residuals in the case of a changepoint in 1973 as found with the method of *Vincent* [1998]. It can be seen that the residuals are approximately normal and that the normality assumption is not violated.

[68] This example was also an occasion to compare the relative computational cost of the three methods, on a Pentium M with a 2.13Ghz processor and 1Go ram, the method of *Vincent* [1998] executed in 1.75s, the proposed methodology in 14.2s and the method of *Seidou et al.* [2006] in 101.72s.

9.2. Shifts and Trend Change Detection in the Flood Peaks of the Ogoki River

9.2.1. Data

[69] The Ogoki River is a 480-km-long river located in the province of Ontario, Canada. It flows northeast from lakes west of Lake Nipigon to join the Albany River which ends into the James Bay. Station 04GB004 (Ogoki River above Whiteclay Lake) is part of the Canadian Reference Hydrometric Basin Network (RHBN) which comprises stations that have been carefully selected for climate change detection and assessment studies [*Brimley et al.*, 1999]. The RHBN network comprises stations that are pristine. Station 04GB004 was selected because it displays a relatively large number of changepoints. The location of this station is given in Figure 8.

9.2.2. Results

[70] The results of the changepoint analysis of the Ogoki River streamflows with the method proposed in this paper are presented in Figure 9. The results obtained with the approach of *Seidou et al.* [2006] are provided in Figure 10. The posterior probability distribution of the number of changepoints obtained with the proposed method is plotted in Figure 10a, while the simulated mean before and after the changepoint is given in Figure 10b. Up to four changepoints

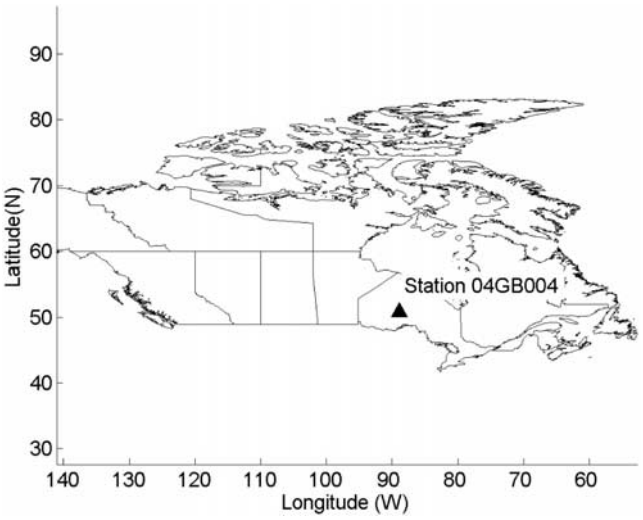

Figure 8. Location of station 04GB004 (Ogoki River above Whiteclay Lake).

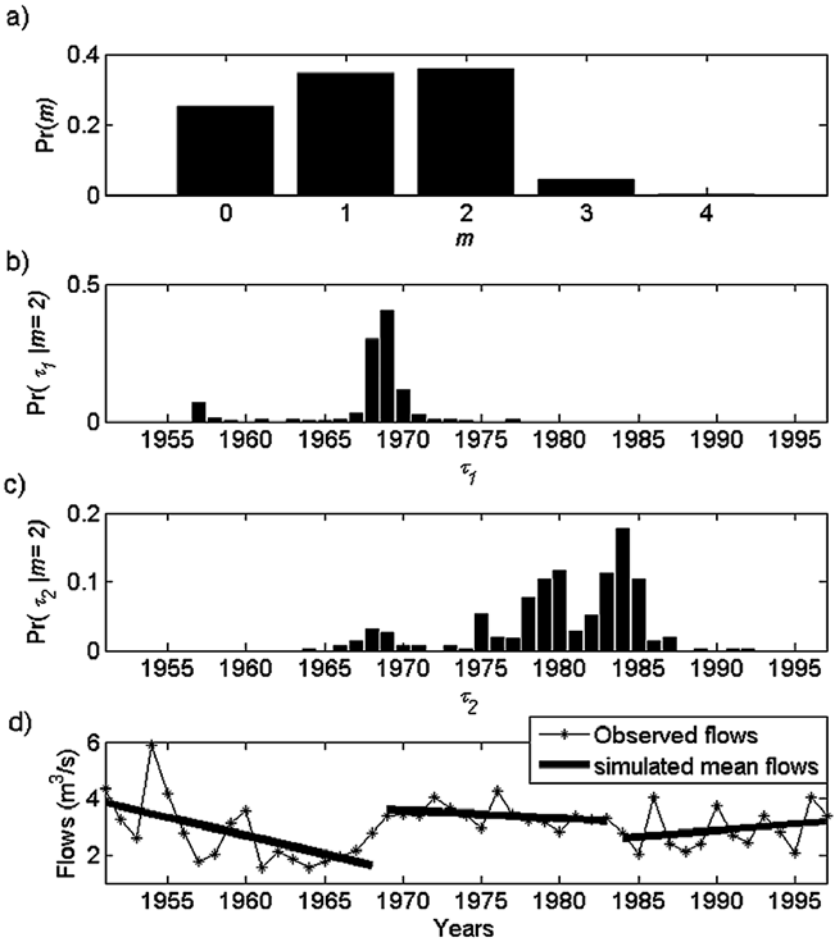

Figure 9. Detection of trend changes at station 04GB004 (Ogoki River above Whiteclay Lake) with the proposed methodology.

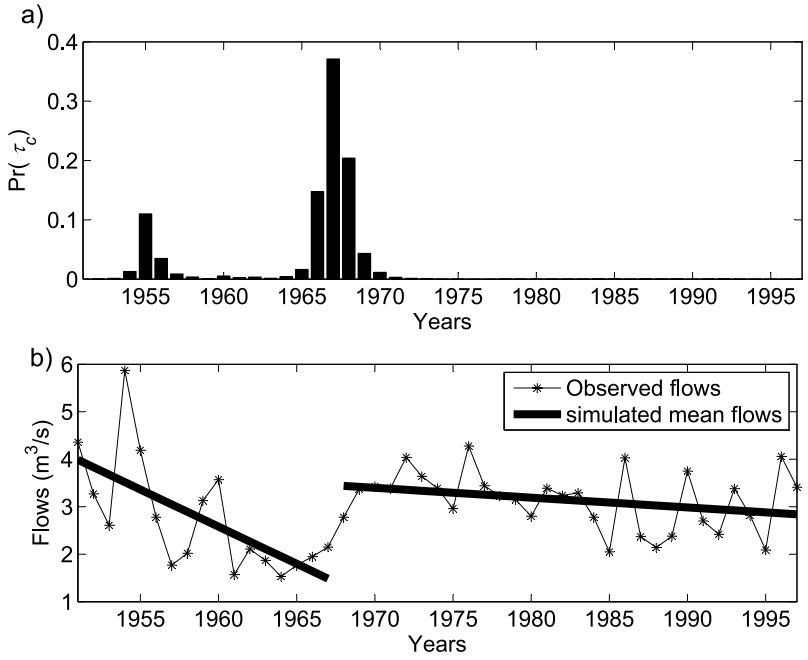

Figure 10. Detection of trend changes at station 04GB004 (Ogoki River above Whiteclay Lake) with a recursive application of the methodology of *Seidou et al.* [2006].

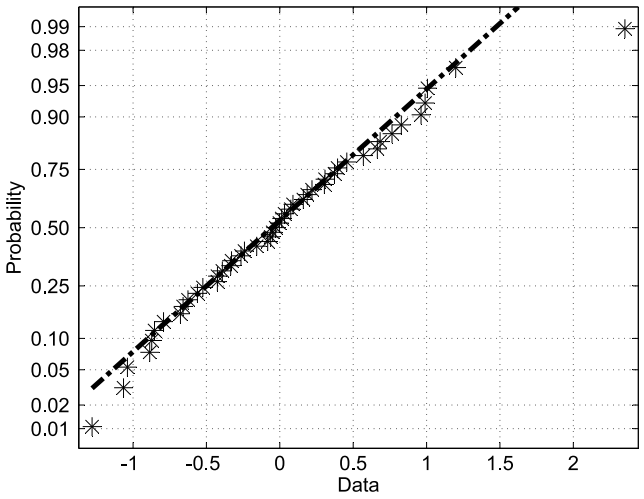

Figure 11. Normal probability plot of the residuals given (Ogoki river).

are plausible (Pr ($m = 4$) > 0), but the most probable number of changepoints is two. Figures 9b and 9c provide the posterior probability distributions of the first and second changepoints, conditional to $m = 2$. The position of each of these changepoints is chosen to be the mode of the posterior distribution, 1961 for the first changepoint and 1971 for the second changepoint. Given these positions, the posterior means of the three segments in the data series are readily computed (Figure 9d). According to the analysis, the flows of the Ogoki River displayed a negative downward trend from 1951 to 1967. There was an abrupt shift in 1968, with another smaller downward trend up to 1983. From 1984 to the present date, the streamflow record displayed a small upward trend.

[71] Figure 9a illustrates the posterior probability distribution of the changepoint obtained with the methodology of *Seidou et al.* [2006]. This method gives less than 0.01 probability of no change (with this method, the probability of no change is equal to the probability that the changepoint is at the end of the data series). The mode of the posterior distribution of the date of change corresponds to 1968. The recursive application of the method of *Vincent* [1998] gave only one changepoint, located at position 1968 (potential changepoints are 1969 and 1970, with a maximum F statistics for 1969). Figure 11 presents the normal probability plot of the residuals given a change in 1968. Once again, the residuals are approximately normal, with a slight departure from normality in the upper tail. At this stage of the research, the sensitivity of the proposed method to moderate violation of the normality assumption is not known, but its extension to nonnormal distributions is (in theory) straightforward and will be discussed at the end of this section.

[72] These dates (1968 and 1969) correspond grossly to the first of the two changepoints detected with the methodology presented in this paper. This indicates that the results of the three methods are consistent. The method of *Seidou et al.* [2006] displayed a multimodal posterior distribution which is sometimes the sign of the existence of more than one changepoint. In this example, the fact that the posterior distribution is bimodal suggests that there may be another changepoint in 1955. However, this seems to have been caused rather by the high discharge observed in 1954 than by a real change of trend in the data series.

[73] Since the causes of trend change in the streamflow record are not known, it is impossible to decide whether the results of one or the other of the two methods correspond to the reality. However, two major advantages of the proposed approach (1) are the smaller computational burden and (2) the fact that it has fewer constraints and gives a larger chance for the data to influence the posterior distributions. The proposed approach is thus preferable in cases where there is only one response variable, where no data is missing, and where more than one change is plausible. The results presented in this work are also easier to interpret than those of the approach proposed by *Seidou et al.* [2006].

[74] As most hydrological variables display nonnormal distributions, the extension of the methodology presented in this paper to more general models is highly desirable. Such extension is straightforward since the most important equations were obtained without assumptions on model structure. The remaining task is to derive an expression of the cohesion $P(t, s)$ [cf. equations (3) and (9)] or at least find an efficient numerical estimation of $P(t, s)$ for the candidate model. However, obtaining an analytical expression for $P(t, s)$ given a data model is not an easy task. *Fearnhead* [2005, 2006] provides such expressions for single series with normally distributed or Poisson-distributed observations. Numerical approximation seems attractive, but truncation errors and the computational burden would become delicate issues to deal with.

[75] Another alternative to handle nonnormality is to develop a similar approach to the hidden Markov chain models. Complex changepoint problems can be handled in the framework of hidden Markov chain models, such as problems involving nonnormal data and those which display a serial dependence structure in the observations [e.g., *Thyer and Kuczera*, 2003a, 2003b].

10. Conclusions

[76] A Bayesian method of multiple changepoint detection in multiple linear regression is developed and validated with both simulated data and real data sets. The paper also proposes a new class of priors for the parameters of the multiple linear model, as well as useful formulas that permit straightforward computation of the posterior distribution of the positions of changepoints. Results suggest that, in the particular case of series with 75 observations, the proposed method can be trusted if the shifts in the data set have the order of magnitude of the standard deviation and if the number of changes is known to be inferior to three. It is also shown that the proposed method is computationally more efficient than its MCMC-based counterpart. Hence, in cases where there is only one response variable, where no data is missing, and where more than one change is plausible, it is better to use the proposed methodology instead of the work of *Seidou et al.* [2006].

[77] As the most important equations were obtained without assumptions on model structure, its extension to more complex data models using nonnormal distributions is straightforward using minor analytical developments or

numerical approaches. Extension to hidden Markov Models is briefly discussed.

Appendix A: Properties of $p(\sigma|a, c) \propto \sigma^{-a}$ exp $(-\frac{C}{2\sigma^2})$, $a > 1$, $C > 0$

[78]

$$\text{Let } I(a) = \int_{\sigma=0}^{\infty} \sigma^{-a} \exp\left(-\frac{C}{2\sigma^2}\right) \mathrm{d}\sigma \tag{A1}$$

$$\begin{aligned} &\text{Let } t = \frac{C}{2\sigma^2} \Rightarrow \sigma = \left(\frac{C}{2t}\right)^{1/2} \Rightarrow \mathrm{d}\sigma = -\frac{1}{2} t^{-\frac{3}{2}} \left(\frac{C}{2}\right)^{1/2} \\ &\mathrm{d}t = -2^{-\frac{3}{2}} C^{\frac{1}{2}} t^{-\frac{3}{2}} \mathrm{d}t \\ &I(a) = \int_{\sigma=0}^{\infty} \left(\frac{C}{2t}\right)^{-a/2} \exp(-t)\left(-2^{-\frac{3}{2}} C^{\frac{1}{2}} t^{-\frac{3}{2}}\right) \mathrm{d}t \end{aligned} \tag{A2}$$

$$I(a) = 2^{\frac{a-3}{2}} C^{\frac{1-a}{2}} \int_{\sigma=0}^{\infty} t^{(a-3)/2} \exp(-t) \mathrm{d}t = 2^{-\frac{a-3}{2}} C^{-\frac{a-1}{2}} \Gamma\left(\frac{a-1}{2}\right) \tag{A3}$$

If $a > 2$, the expectation of σ is finite:

$$E(\sigma) = \frac{I(a-1)}{I(a)} = \frac{2^{-\frac{a-4}{2}} C^{-\frac{a-2}{2}} \Gamma\left(\frac{a-2}{2}\right)}{2^{-\frac{a-3}{2}} C^{-\frac{a-1}{2}} \Gamma\left(\frac{a-1}{2}\right)} = \frac{(2C)^{-\frac{1}{2}} \Gamma\left(\frac{a-2}{2}\right)}{\Gamma\left(\frac{a-1}{2}\right)} \tag{A4}$$

if $a > 3$, the variance of σ is finite:

$$E(\sigma^2) = \frac{I(a-2)}{I(a)} = \frac{2^{-\frac{a-5}{2}} C^{-\frac{a-3}{2}} \mathrm{G}\left(\frac{a-3}{2}\right)}{2^{-\frac{a-3}{2}} C^{-\frac{a-1}{2}} \Gamma\left(\frac{a-1}{2}\right)} = \frac{\Gamma\left(\frac{a-2}{2}\right)}{(2C)\Gamma\left(\frac{a-1}{2}\right)} \tag{A5}$$

$$\mathrm{Var}(\sigma) = E(\sigma^2) - (E(\sigma))^2 = \frac{\Gamma\left(\frac{a-3}{2}\right)\Gamma\left(\frac{a-1}{2}\right) - \left(\Gamma\left(\frac{a-2}{2}\right)\right)^2}{\left(2C\Gamma\left(\frac{a-1}{2}\right)^2\right)} \tag{A6}$$

The case $a < 3$ leads to an infinite variance for σ, that is, $\lim_{x\to+\infty} \int_0^x p(\sigma)d\sigma = +\infty$. Any value of a less than 3 can thus be used as a noninformative prior. Note that when σ is very large, $p(\sigma) \propto \sigma^{-a}$.

Appendix B: Derivation of $L(\{i_1, i_2, \ldots, i_{n1}\})$

[79] In this section, we will derive the expression for $L(\{i_1, i_2, \ldots, i_{n1}\})$. To simplify the notation, let $idx = \{i_1, i_2, \ldots, i_{n1}\}$. Let $\hat{\theta}$ be the ordinary least squares solution of the equation $\mathbf{Y}_{idx} = \mathbf{X}_{idx}\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}_{idx} = \mathbf{Y}_{idx} - \mathbf{X}_{idx} > \hat{\theta}\}$. Note that $\boldsymbol{\varepsilon}_{idx}$ does not depend on $\boldsymbol{\theta}$ or σ. It is well known from linear algebra that $\hat{\theta} = (\mathbf{X}_{idx}^{\mathrm{T}}\ \mathbf{X}_{idx})^{-1}\ \mathbf{X}_{idx}^{\mathrm{T}}\ \mathbf{Y}_{idx}$. We also suppose that $p(\sigma|a, c) = \frac{\sigma^{-a} \exp\left(-\frac{c}{2\sigma^2}\right)}{2^{\frac{a-3}{2}} c^{-\frac{a-1}{2}} \Gamma\left(\frac{a-1}{2}\right)}$, a > 1,c > 0. We have:

$$\begin{aligned} L(idx) = &\int_{\sigma} \left(2\pi\sigma^2\right)^{-(\mathrm{card}(idx))/2} \pi(\sigma) \\ &\cdot \int_{\mathbf{b}} \exp\left[-\frac{(\mathbf{Y}_{idx} - \mathbf{X}_{idx}\boldsymbol{\theta})^{\mathrm{T}}(\mathbf{Y}_{idx} - \mathbf{X}_{idx}\boldsymbol{\theta})}{2\boldsymbol{\sigma}^2}\right] \mathrm{d}\boldsymbol{\sigma}\mathrm{d}\boldsymbol{\theta} \end{aligned} \tag{B1}$$

Equation (B1) can be simplified since

$$\begin{aligned} &\frac{(\mathbf{Y}_{idx} - \mathbf{X}_{idx}\theta)^{\mathrm{T}}(\mathbf{Y}_{idx} - \mathbf{X}_{idx}\boldsymbol{\theta})}{2\boldsymbol{\sigma}^2} \\ &= \frac{\left(\boldsymbol{\varepsilon}_{idx} - \mathbf{X}_{idx}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)\right)^{\mathrm{T}}\left(\boldsymbol{\varepsilon}_{idx} - \mathbf{X}_{idx}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)\right)}{2\boldsymbol{\sigma}^2} \\ &= \frac{1}{2\boldsymbol{\sigma}^2}\left(\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx} - \mathbf{e}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^{\mathrm{T}}\mathbf{X}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx}\right. \\ &\quad \left. + \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^{\mathrm{T}}(\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx})\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)\right) \end{aligned} \tag{B2}$$

and

$$\mathbf{X}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx} = \left(\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right)^{\mathrm{T}} = \mathbf{X}_{idx}^{\mathrm{T}}\left(\mathbf{Y}_{idx} - \mathbf{X}_{idx}\left(\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right)^{-1}\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{Y}_{idx}\right) = 0, \tag{B3}$$

thus:

$$\begin{aligned} L(idx) = &\int \left(2\pi\boldsymbol{\sigma}^2\right)^{-(\mathrm{card}(idx))/2} p(\boldsymbol{\sigma}) \\ &\cdot \exp\left[-\frac{\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx} + \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^{\mathrm{T}}\left(\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)}{2\boldsymbol{\sigma}^2}\right] \mathrm{d}\boldsymbol{\sigma}\mathrm{d}\boldsymbol{\theta} \end{aligned} \tag{B4}$$

$$\begin{aligned} L(idx) = &\int \left(2\pi\boldsymbol{\sigma}^2\right)^{-\mathrm{card}(idx)/2} \boldsymbol{\pi}(\boldsymbol{\sigma}) \exp\left(-\frac{\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx}}{2\boldsymbol{\sigma}}\right) \\ &\cdot \int -\frac{\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^{\mathrm{T}}\left(\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)}{2\boldsymbol{\sigma}} \mathrm{d}\boldsymbol{\sigma}\mathrm{d}\boldsymbol{\theta} \end{aligned} \tag{B5}$$

let $\boldsymbol{\Sigma}$ be $\boldsymbol{\sigma}^2\ (\mathbf{X}_{idx}^{\mathrm{T}}\ \mathbf{X}_{idx})^{-1} \Rightarrow |\Sigma| = \frac{\delta^{2d*}}{|\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}|}$;

$$\begin{aligned} L(idx) = &\int_{\sigma} \left(2\pi\sigma^2\right)^{-(\mathrm{card}(idx))/2} p(\sigma) \exp\left(-\frac{\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx}}{2\boldsymbol{\sigma}^2}\right) (2\boldsymbol{\pi})^{d*/2} |\boldsymbol{\Sigma}|^{1/2} \\ &\times \int_{b} \underbrace{\frac{(2\boldsymbol{\pi})^{-d*/2}}{|\mathrm{S}|^{1/2}} \exp\left(-\frac{\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^{T} \mathrm{S}^{-1}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)}{2}\right)}_{\int_{\theta} N_{\mathrm{card}(idx)}\left(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathrm{S}\right) = 1} \mathrm{d}\sigma\mathrm{d}\boldsymbol{\theta} \end{aligned} \tag{B6}$$

In equation B6, $d*$ is the length of $\boldsymbol{\theta}$, i.e., the number of explanatory variables (including the intercept if any).

$$L(idx) = (2\pi)^{-(\mathrm{card}(idx)-d*)/2}\left|\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right|^{-1/2} \cdot \int_{\sigma} \sigma^{d*-\mathrm{card}(idx)} \exp\left(-\frac{\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx}}{2\boldsymbol{\sigma}^2}\right) p(\boldsymbol{\sigma}) d\boldsymbol{\sigma} \tag{B7}$$

$$L(idx) = \frac{(2\boldsymbol{\pi})^{-\left(\mathrm{card}(idx)-d^*\right)/2}\left|\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right|^{1/2}}{2^{\frac{a-3}{2}}c^{-\frac{a-1}{2}}\boldsymbol{\Gamma}\left(\frac{a-1}{2}\right)} \cdot \int_{\sigma} \sigma^{d*-\mathrm{card}(idx)} \exp\left(-\frac{\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx}+C}{2\boldsymbol{\sigma}^2}\right) p(\boldsymbol{\sigma}) \mathrm{d}\boldsymbol{\sigma} \tag{B8}$$

$$L(idx) = \frac{(2\boldsymbol{\pi})^{-\left(\mathrm{card}(idx)-d^*\right)/2}\left|\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right|^{-1/2}}{2^{\frac{a-3}{2}}c^{-\frac{a-1}{2}}\boldsymbol{\Gamma}\left(\frac{a-1}{2}\right)} 2^{\frac{\mathrm{card}(idx)+a-3}{2}} \cdot \left(\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx}+C\right)^{-\frac{\left(\mathrm{card}(idx)+a-d*\right)}{2}} \boldsymbol{\Gamma}\left(\frac{\mathrm{card}(idx)+a-d*}{2}\right) \tag{B9}$$

$$L(idx) = (2\pi)^{\frac{d*}{2}} \frac{\left(\pi\left(\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx}+C\right)\right)^{-\frac{(\mathrm{card}(idx)+a-1)}{2}}}{(C\boldsymbol{\pi})^{-\frac{a-1}{2}}\left|\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right|^{1/2}} \cdot \frac{\boldsymbol{\Gamma}\left(\frac{\mathrm{card}(idx)+a-d*}{2}\right)}{\Gamma\left(\frac{a-1}{2}\right)} \tag{B10}$$

as $\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\ \boldsymbol{\varepsilon}_{idx} = Y_{idx}^{\mathrm{T}}\ Y_{idx} - X_{idx}\ (X_{idx}^{\mathrm{T}}\ X_{idx})^{-1}\ Y_{idx}^{\mathrm{T}}\ Y_{idx})$ we finally obtain the expression for P_1 (idx):

$$L(idx) = (2\pi)^{\frac{d*}{2}} \frac{\left(\pi\left(\boldsymbol{\varepsilon}_{idx}^{\mathrm{T}}\boldsymbol{\varepsilon}_{idx}+C\right)\right)^{-\frac{(\mathrm{card}(idx)+a-d*)}{2}}}{(C\boldsymbol{\pi})^{-\frac{a-1}{2}}\left|\mathbf{X}_{idx}^{\mathrm{T}}\mathbf{X}_{idx}\right|^{1/2}} \cdot \frac{\boldsymbol{\Gamma}\left(\frac{\mathrm{card}(idx)+a-d*}{2}\right)}{\Gamma\left(\frac{a-1}{2}\right)} \tag{B11}$$

List of symbols

$\boldsymbol{\beta}_1^*$ Regression parameters before the changepoint in the methodology of *Seidou et al.* [2006]

$\boldsymbol{\beta}_2^*$ Regression parameters after the changepoint in the methodology of *Seidou et al.* [2006]

$\boldsymbol{\beta}_0$ Component of the vector of regression parameter that does not change in the methodology of *Seidou et al.* [2006]

$\boldsymbol{\beta}_1$ Component of the vector of regression parameter that change to $\boldsymbol{\beta}_2$ after τ_c t in the methodology of *Seidou et al.* [2006]

$\boldsymbol{\beta}_2$ Component of the vector of regression parameter that replaces $\boldsymbol{\beta}_1$ after τ_c in the methodology of *Seidou et al.* [2006]

$\boldsymbol{\varepsilon}$ Vector of random errors in the linear regression equation (one response variable)

$\boldsymbol{\varepsilon}_{s:t}$ Part of the vector of random errors between s and t

$\boldsymbol{\upsilon}_t$ Vector of random errors in the linear regression equation (several response variables)

Φ Parameters of the linear regression equation

Σ_y Variance-covariance matrix of the distribution of $\boldsymbol{\upsilon}_t$

τ_c Last point of the segment before the change (methodology of *Seidou et al.* [2006])

τ_k kth changepoint in the proposed methodology

$\boldsymbol{\theta}$ Vector of regression parameters

$\boldsymbol{\theta}_t^{(\tau_c)}$ Vector of regression parameters at date t given τ_c (methodology of *Seidou et al.* [2006])

a Parameter of the prior distribution of Φ

c Parameter of the prior distribution of Φ

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

d_0^* Number of explanatory variables for which the regression coefficients do not change

d_1^* Number of explanatory variables for which the regression coefficients display a change (methodology of *Seidou et al.* [2006])

E Set of generated scatter schemes

$G(t)$ Cumulative probability distribution of the time interval between consecutive changepoints

$g(t)$ Probability distribution of the time interval between consecutive changepoints

$G_0(t)$ Cumulative probability distribution of the first changepoint

$g_0(t)$ Probability distribution of the first changepoint

k Number of generated scatter schemes $S_k = \{\tilde{t}_1^k, \tilde{t}_2^k, \ldots, \tilde{t}_{mk}^k\}$ in the inference procedure

$L(\{i_1, i_2, \ldots, i_n\})$ $\int_{\mathrm{F}} (f(y_{i_1}|\Phi)\, f(y_{i2}|\Phi) \ldots f(y_{i_n}|\Phi))\pi_1(\Phi) d\Phi$

M Number of scatter schemes to generate with the posterior distributions of the positions of changepoints

MCDPI Multiple change detection performance index

m_u Number of changes in the uth generated series

$\hat{m}_u$	Estimate of the number of changes in the uth generated series
$\tilde{m}_k$	Number of changes in the kth generated scatter scheme during the simulation of the changepoints
n	Length of the data series
N	Number of sets to generate
$P(t, s), s \geq t$	Probability that t and s be in the same segment
PCDN	Percentage of correct detections of the number of changepoints
$Q(t)$	Likelihood of the segment $\mathbf{Y}_{t:n}$ given a changepoint at $t - 1$
r	Number of response variables (methodology of *Seidou et al.* [2006])
RMSE	Root mean square error
RPS	Ranked probability score
$S_k = \{\tilde{t}_1^k, \tilde{t}_2^k, \ldots, \tilde{t}_{m_k}^k\}$	kth scatter scheme generated with the posterior distributions of the positions of changepoints
t	Time
$\tilde{t}_i^k$	Estimate of the ith change in the kth generated scatter scheme
t_i^k	ith change in the kth generated scatter scheme
u	Number of the generated series $\{\mathbf{Y}\}_u$ in the validation procedure
$\mathbf{X}$	Vector of explanatory variables
$\mathbf{X}_t$	tth row of the vector of explanatory variables
$\mathbf{X}_{t:s}$	Rows t to s of the vector of explanatory variables
$\mathbf{Y}_t$	Rows t to s of the vector of response variables
$\{\mathbf{Y}\}_u$	uth generated series in the validation procedure

[80] **Acknowledgments.** The financial support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chair Program, and the Ouranos consortium is gratefully acknowledged. The authors are also grateful to the Quebec Ministry of the Environment for having provided the data sets used in the case studies. The authors thank the associate editor and the three anonymous reviewers for their comments which greatly helped improve the quality of the paper.

References

Barry, D., and J. A. Hartigan (1992), Product partition models for change point models, *Ann. Stat.*, *20*, 260–279.

Barry, D., and J. A. Hartigan (1993), A Bayesian analysis for change point problems, *J. Am. Stat. Assoc.*, *88*, 309–319.

Beaulieu, C., T. B. M. J. Ouarda, and O. Seidou (2005), Comparative study of homogenization techniques for precipitation data series (*in French*), Progress report no. 3 (Project on the homogenization of precipitation data), Ouranos Consortium, Montreal.

Beibel, M. (1997), Sequential change-point detection in continuous time when the post-change drift is unknown, *Bernoulli J. Math. Stat. Probab.*, *3*(4), 457–478.

Berger, J. O., and L. R. Pericchi (1998), On criticisms and comparisons of default Bayes factors for model selection and hypothesis testing (with discussion), In *Proceedings of the Workshop on Model Selection*, Rassegna di Metodi Statistici ed Applicazioni, Pitagora Editrice, Bologn.

Booth, N. B., and A. F. M. Smith (1982), A Bayesian approach to retrospective identification of change-points, *J. Econometrics*, *19*, 7–22.

Bowman, A. W., A. Pope, and B. Ismail (2004), Detecting discontinuities in nonparametric regression curves and surfaces, *Research Report*, Department of Statistics, University of Glasgow. [http://www.stats.gla.ac.uk/%7Eadrian/research-reports/discontinuity-paper.ps]

Brimley, B., et al. (1999), *Establishment of the reference hydrometric basin network* (RHBN), 41 p., Research report, Environment Canada.

Bruneau, P., and J.-C. Rassam (1983), Application d'un modèle Bayésien de détection de changements de moyennes dans une série, *J. Sci. Hydrol.*, *28*, 341–354.

Buizza, R., and T. N. Palmer (1998), Impact of ensemble size on ensemble prediction, *Mon. Weather Rev.*, *126*(9), 2503–2518.

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

Carlin, B. P., A. E. Gelfand, and A. F. M. Smith (1992), Hierarchical Bayesian analysis of changepoint problems, *Appl. Stat.*, *41*, 389–405.

Daumer, M., and M. Falk (1998), On-line change-point detection for state space models using multi-process Kalman filters, *Linear Algebra Appl.*, *284*, 125–135.

Denison, D. G. T., B. K. Mallick, and A. F. M. Smith (1998), Automatic Bayesian curve fitting, *J. R. Stat. Soc., Ser. B*, *60*(2), 333–350.

Easterling, D. R., and T. C. Peterson (1992), Techniques for detecting and adjusting for artificial discontinuities in climatological time series: A review, *Proc. of the Fifth International Meeting on Statistical Climatology*, 22–26 June 1996, Toronto, Ontario, Canada.

Fearnhead, P. (2005), Exact Bayesian curve fitting and signal segmentation, *IEEE Trans. Signal Process.*, *53*(6), 2160–2166.

Fearnhead, P. (2006), Exact and efficient Bayesian inference for multiple changepoint problems, *Stat. Comput.*, *16*, 203–213.

Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. M. Smith (1990), Illustration of Bayesian inference in normal data models using Gibbs sampling, *J. Am. Stat. Assoc.*, *85*, 972–985.

Gut, A., and J. Steinebach (2002), Truncated sequential change-point detection based on renewal counting processes, *Scand. J. Stat.*, *29*(4), 693–719.

Hamil, T. M. (2001), Interpretation of rank histograms for verifying ensemble forecasts, *Mon. Weather Rev.*, *129*(3), 550–560.

Lai, T. L. (1995), Sequential change point detection in quality control and dynamical systems, *J. R. Stat. Soc., Ser. B*, *57*, 613–658.

Lund, R., and J. Reeves (2002), Detection of undocumented changepoints: A revision of the two-phase regression model, *J. Clim.*, *15*, 2547–2554.

Moreno, E., G. Casella, and A. Garcia-Ferrer (2005), An objective Bayesian analysis of the change point problem, *Stochastic Environ. Res. Risk Assess. (SERRA)*, *19*(3), 191–204.

Perreault, L., J. Bernier, B. Bobée, and E. Parent (2000a), Bayesian changepoint analysis in hydrometeorological time series 1. Part 1. The normal model revisited, *J. Hydrol.*, *235*, 221–241.

Perreault, L., J. Bernier, B. Bobée, and E. Parent (2000b), Bayesian change-point analysis in hydrometeorological time series 2. Part 2. Comparison of change-point models and forecasting, *J. Hydrol.*, *235*, 242–263.

Perreault, L., É. Parent, J. Bernier, and B. Bobée (2000c), Retrospective multivariate Bayesian change-point analysis: A simultaneous single change in the mean of several hydrological sequences, *Stochastic Environ. Res. Risk Assess.*, *14*, 243–261.

Punskaya, E., C. Andrieu, A. Doucet, and W. J. Fitzgerald (2002), Bayesian curve fitting using MCMC with applications to signalsegmentation, *IEEE Trans. Signal Process.*, *50*(4), 747–758.

Rasmussen, P. (2001), Bayesian estimation of change points using the general linear model, *Water Resour. Res.*, *37*, 2723–2731.

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

Salinger, M. (2005), Climate variability and change: Past, present and future—An overview, *Clim. Change*, *70*, 9–29.

Seidou, O., J. J. Asselin, and T. B. M. J. Ouarda (2006), Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables, *Water Resour. Res.*, doi:10.1029/2005WR004835, in press.

Stephens, D. A. (1994), Bayesian retrospective multiple-changepoint identification, *Appl. Stat.*, *43*, 159–178.

Solow, A. R. (1987), Testing for climate change: An application of the two-phase regression model, *J. Appl. Meteorol.*, *26*, 1401–1405.

Stocks, B. J., et al. (2002), Large forest fires in Canada, 1959–1997, *J. Geophys. Res.*, *107*, 8149, doi:10.1029/2001JD000484. [printed 108(D1), 2003]

Thyer, M., and G. Kuczera (2003a), A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach, *J. Hydrol.*, *275*, 12–26.

Thyer, M., and G. Kuczera (2003b), A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. 2. Real data analysis, *J. Hydrol.*, *275*, 27–48.

Vincent, L. A. (1998), A technique for the identification of inhomogeneities in Canadian temperature series, *J. Clim.*, *11*, 1094–1105.

Wang, X. L. (2003), Comments on 'Detection of Undocumented Changepoints: A revision of the two-phase regression model', *J. Clim.*, *16*, 3383–3385.

Woo, M., and R. Thorne (2003), Comment on 'Detection of hydrologic trends and variability' by D. H. Burn and M. A. Hag Elnur, 2002, Journal of Hydrology 255, 107–122, *J. Hydrol.*, *277*, 150–160.

Yao, Y. (1984), Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches, *Ann. Stat.*, *12*, 1434–1447.

T. B. M. J. Ouarda and O. Seidou, Hydro-Québec NSERC Chair in Statistical Hydrology, INRS-ETE, 490 rue dela Couronne, Quebec (QC), G1K949 Quebec, Canada. (ousman_seidou@ete.irnrs.ca)