Bayesian multivariate linear regression with application to change point models in hydrometeorological variables

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

Received 22 December 2005; revised 31 January 2007; accepted 5 March 2007; published 2 August 2007.

[1] Multivariate linear regression is one of the most popular modeling tools in hydrology and climate sciences for explaining the link between key variables. Piecewise linear regression is not always appropriate since the relationship may experiment sudden changes due to climatic, environmental, or anthropogenic perturbations. To address this issue, a practical and general approach to the Bayesian analysis of the multivariate regression model is presented. The approach allows simultaneous single change point detection in a multivariate sample and can account for missing data in the response variables and/or in the explicative variables. It also improves on recently published change point detection methodologies by allowing a more flexible and thus more realistic prior specification for the existence of a change and the date of change as well as for the regression parameters. The estimation of all unknown parameters is achieved by Monte Carlo Markov chain simulations. It is shown that the developed approach is able to reproduce the results of Rasmussen (2001) as well as those of Perreault et al. (2000a, 2000b). Furthermore, two of the examples provided in the paper show that the proposed methodology can readily be applied to some problems that cannot be addressed by any of the above-mentioned approaches because of limiting model structure and/or restrictive prior assumptions. The first of these examples deals with single change point detection in the multivariate 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. The second one addresses the problem of missing data estimation with uncertainty assessment in multisite streamflow records with a possible simultaneous shift in mean streamflow values that occurred at an unknown date.

Citation: Seidou, O., J. J. Asselin, and T. B. M. J. Ouarda (2007), Bayesian multivariate linear regression with application to change point models in hydrometeorological variables, *Water Resour. Res.*, *43*, W08401, doi:10.1029/2005WR004835.

1. Introduction

[2] Because of the growing evidence of climate change, the common assumption of stationarity of hydrologic phenomena no longer holds. Several recently published works point out shifts or trend changes in hydrologic time series [e.g., *Salinger*, 2005; *Woo and Thorne*, 2003; *Burn and Hag Elnur*, 2002]. Possible reasons of change in statistical characteristics of observed data series include natural or anthropogenic actions on the physical environment (deforestation, construction of hydraulic structures, pollution, etc.), and modifications in measurement equipment or protocol.

[3] To deal with these nonstationary data sets, change point analysis in hydrologic time series is regularly revisited using various assumptions on the data model, on the parameter that exhibits a change as well as on the type of change. However, problems in hydrology often involve missing data and interactions between several data series that cannot easily be handled by recently published methodologies. The main objective of the present paper is to develop a change point model that allows simultaneous change point analysis of several time series, each time series being modeled as a linear combination of a set of explanatory variables. The method generalizes the model of *Rasmussen* [2001] to cases where there is more than one response variable, to cases where the change point does not occur with certainty and to cases where informative priors on the regression parameters are required. It also improves on the models of *Perreault et al.* [2000a, 2000b] which are all special cases of the model presented in this paper. Unfortunately, the solution is no longer analytic and inference is performed using Monte Carlo Markov chain simulation.

[4] The developed model is also applied to three different examples to illustrate its features and flexibility: (1) detection of a single shift in univariate data, (2) detection of a change point in univariate data with several covariates, and (3) a case of shift detection and missing data estimation in a multivariate data set. The first example aims to show that the proposed methodology gives the same results than the above-mentioned approaches when applied to the same data sets with the same prior assumptions. The two last examples illustrate the additional features of the proposed approach

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

[2] Now at Department of Civil Engineering, University of Ottawa, Ottawa, Ontario, Canada.

0043-1397/07/2005WR004835$09.00

and prove that it can be applied to cases where the other published methodologies are inadequate.

[5] Aside from the introduction, the paper is organized as follows: a quick review of recently published change point methodologies is presented in section 2. Section 3 presents the development of the proposed method of change point detection in multiple linear regression using Gibbs sampling. The three case studies are described in section 4, and the findings are discussed in section 5. A conclusion is finally given in section 6.

2. Review of Recently Published Bayesian Change Point Detection Models

[6] A change point can be defined as the date at which at least one parameter of a statistical model (e.g., mean, variance, intercept, trend) undergo an abrupt change. It is a very common problem in signal processing and a large number of techniques can be found in the literature to find the date of a potential change and to check if the change is significant or not. Most of the published methodologies use classical statistical hypothesis testing to detect changes in slopes or intercept of linear regression models [*Solow*, 1987; *Easterling and Peterson*, 1995; *Vincent*, 1998; *Lund and Reeves*, 2002; *Wang*, 2003]. For instance, *Solow* [1987], *Easterling and Peterson* [1995], *Vincent* [1998], *Lund and Reeves* [2002] and *Wang* [2003] all use the Fisher test to compare the model with a change point to the model without change points. *Vincent* [1998] uses the Durbin-Watson test to check the autocorrelation of the residuals before applying the Fisher test. The Student test and the Wilcoxon test can be applied sequentially to detect change points in data series [*Beaulieu et al.*, 2007].

[7] However, not all change point approaches are based on hypothesis testing: for instance, *Wong et al.* [2006] used the grey relational method [*Moore*, 1979; *Deng*, 1989] for single change point detection in streamflow data series. In some rare cases, other curve fitting methods are used [e.g., *Sagarin and Micheli*, 2001; *Bowman et al.*, 2006]. Extensive reviews of change point detection and correction methodologies in hydrology and climate sciences can be found in work by *Peterson et al.* [1998] and *Beaulieu et al.* [2007].

[8] The change point problem was also addressed in Bayesian statistics. The advantage of Bayesian statistics over classical statistics is the comprehensive description of parameters uncertainty. While classical statistics may give the most probable position of the change point, Bayesian methods provide a full posterior probability distribution of its position. The posterior probability distribution can for instance be skewed and/or multimodal. It thus provides much more information than a simple estimation and a credibility interval as usually obtained with classical methods.

[9] The Bayesian methods were applied considering single or multiple changes, in conjunction with a known or an unknown number of change points. The case where the number of change points is known is easier to handle, while multiple change points can involve sophisticated algorithms in which several potential models with different number of change points are compared. For instance, *Green* [1995] uses reversible jump Monte Carlo Markov chains to solve a multiple change point problem, using a sampler that jumps between parameter subspaces of differing dimensionality. Examples of alternative strategies for handling an unknown number of change points can also be found in work by *Barry and Hartigan* [1993], *Fearnhead* [2005] or *Seidou and Ouarda* [2006]. *Barry and Hartigan* [1993] introduced the multiple change point component by a normal random variable that can be added anytime to the mean of the series, but only with a certain probability. In work by *Fearnhead* [2005] and *Seidou and Ouarda* [2006], the conditional distributions of the $(i + 1)^{th}$ change point given the i^{th} change points are derived in a Bayesian framework.

[10] Examples of approaches using a known number of change points include *Gelfand et al.* [1990], *Stephens* [1994], and *Rasmussen* [2001]. *Stephens* [1994] implemented a Bayesian analysis of a multiple change point problem where the number of change points is assumed known, but the times of occurrence of the change points remain unknown. *Gelfand et al.* [1990] also considered a known number of change points and discussed Bayesian analysis of a variety of normal data models, including regression and ANOVA-type structures, where they allowed for unequal variances. *Rasmussen* [2001] considered a single change point 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 change point occurred with certainty, and does not allow a clear diagnosis of the existence of the change.

[11] Other authors emphasized on problem in which the data series contain at most one change point (i.e., the authors give a criterion to decide whether there is zero or one change points in the data series). We cite, for example, *Carlin et al.* [1992], who applied a three-stage hierarchical Bayesian analysis to a simple linear change point model for normal data where a single change point occurs on the regression and variance parameters. *Perreault et al.* [2000a] gave Bayesian analyses of several change point models of univariate normal data. All of these authors implemented their analyses using Gibbs sampling. Exact analytical Bayesian analyses were proposed by *Solow* [1987] for the two-phase regression model and by *Perreault et al.* [2000b] for a change in the mean of a series of multivariate normal random variables. More recently, *Zhao and Chu* [2006] used a hierarchical Bayesian model to detect multiple change points in annual Hurricane counts.

[12] In the three following subsections, more details are provided on three well known change point models that will be compared to the proposed methodology: the model of single shift detection in univariate data developed by *Perreault et al.* [2000a], the model of single shift detection in multivariate normal data of *Perreault et al.* [2000b], and the change point detection model in the general linear model developed by *Rasmussen* [2001].

2.1. Single Change Point Detection in a Normal Univariate Random Sample

[13] The single shift in a normal random sample can be represented by the following model:

$$Y_i \begin{cases} N(\mu_1, \sigma^2), & i = 1, \ldots, \tau \\ N(\mu_2, \sigma^2), & i = \tau + 1, \ldots, n, \end{cases} \quad (1)$$

where τ is the date of change, σ^2 the variance, μ_1 and μ_2 the mean before and after the change. This problem was first addressed in a Bayesian context by *Chernoff and Zacks* [1963], followed by several others [*Smith*, 1975; *Lee and Heighinian*, 1977; *Booth and Smith*, 1982; *Bruneau and Rassam*, 1983; *Perreault et al.*, 2000a]. The difference in the above-mentioned approaches lie mainly in the prior distributions of the unknown parameters. *Perreault et al.* [2000a] derived the exact analytical expression of the posterior probability of the time and magnitude of the shift under the assumption of constant variance. Inferences are based on the analysis of posterior distributions and are conditional upon the fact that a change happened with certainty. The following additional assumptions were made about the prior distributions: The prior distribution of the date of change $p(\tau)$ is independent of that of (μ, σ^2); the prior distribution of μ_1 is normal with parameters Φ_1 and $\lambda_1\sigma^2$; the prior distribution of μ_2 is normal with parameters Φ_2 and $\lambda_2\sigma^2$; and the prior distribution of σ^2 is inverted gamma with parameters α and β:

$$\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\alpha-1}\exp\left(-\frac{\beta}{\sigma^2}\right). \tag{2}$$

[14] The posterior probability of the change point is then

$$p(\tau|\mathbf{x}) \propto p(\tau)\sqrt{\lambda_1'\lambda_2'}(|\beta'|)^{-a'}. \tag{3}$$

where

$$\lambda_1' = \lambda_1/[1+\tau\lambda_1], \lambda_2' = \lambda_2/[1+\tau\lambda_2], \alpha' = \alpha + n/2,$$
$$y_\tau = \sum_{i=1}^{\tau}\frac{y_i}{\tau}, y_{n-\tau} = \sum_{i=\tau+1}^{n}\frac{y_i}{n-\tau},$$

$$s_\tau = \sum_{i=1}^{\tau}\frac{(y_i-\bar{y}_\tau)^2}{\tau}, s_{n-\tau} = \sum_{i=\tau+1}^{n}\frac{(y_i-\bar{y}_{n-\tau})^2}{n-\tau}$$

and

$$\beta' = \frac{\tau}{2}\left[s_\tau + \left(1-\lambda_1'\tau\right)(\phi_1-\bar{x}_\tau)^2\right] + \frac{n-\tau}{2} \cdot \left[s_{n-\tau} + \left(1-\lambda_1'(n-\tau)\right)(\phi_2-\bar{x}_{n-\tau})^2\right] + \beta.$$

2.2. Single Change Point Detection in a Multivariate Random Sample

[15] *Perreault et al.* [2000b] generalized the approach presented in section 2.1 to the case of a change point in a multivariate sample. The equations are quite similar except that the parameters are now p dimensional. The multivariate normal distribution replaces the univariate one and the inverse Wishart distribution is used instead of the inverse Gamma distribution:

$$\mathbf{Y}_i \sim \begin{cases} N_p(\boldsymbol{\mu}_1, \mathbf{P}), & i = 1, \ldots, \boldsymbol{\tau} \\ N_p(\boldsymbol{\mu}_2, \mathbf{P}), & i = \boldsymbol{\tau}+1, \ldots, n, \end{cases}. \tag{4}$$

where N_p stands for the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and precision matrix $\mathbf{P}$.

[16] As in the univariate case, the following assumptions are made about the prior distributions: The prior distribution of the date of change $p(\tau)$ is independent of that of ($\boldsymbol{\mu}$, $\mathbf{P}$); the prior distribution of $\boldsymbol{\mu}_1$ is multivariate normal with parameters Φ_1 and $\lambda_1\mathbf{P}$; the prior distribution of $\boldsymbol{\mu}_2$ is multivariate normal with parameters Φ_2 and $\lambda_2\mathbf{P}$; and the prior distribution of $\mathbf{P}$ is inverse Wishart with parameters a and $\mathbf{B}$, i.e.,

$$\pi(\Sigma) = \left(2^{\upsilon k/2}p^{k(k-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{\upsilon+1-i}{2}\right)\right)^{-1}|\mathbf{S}|^{\upsilon/2}|\Sigma|^{-(\upsilon+k+1)/2} \cdot \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}\Sigma^{-1}\right)\right). \tag{5}$$

[17] Under these assumptions, the posterior probability of the change point

$$p(\tau|\mathbf{Y}) \propto p(\tau)\sqrt{\lambda_1'\lambda_2'}|\mathbf{B}'|^{-a'}, \tag{6}$$

where

$$\lambda_1' = \lambda_1 + \tau,\ \lambda_2' = \lambda_2 + n - \tau,\ a' = a + n/2,$$
$$\bar{\mathbf{y}}_\tau = \sum_{i=1}^{\tau}\frac{\mathbf{y}_i}{\tau},\ \bar{\mathbf{y}}_{n-\tau} = \sum_{i=\tau+1}^{n}\frac{\mathbf{y}_i}{n-\tau},$$

$$\mathbf{S}_\tau = \frac{1}{\tau}\sum_{i=1}^{\tau}(\mathbf{y}_i-\bar{\mathbf{y}}_\tau)(\mathbf{y}_i-\bar{\mathbf{y}}_\tau)^T,$$
$$\mathbf{S}_{n-\tau} = \frac{1}{n-\tau}\sum_{i=\tau+1}^{n}(\mathbf{y}_i-\bar{\mathbf{y}}_{n-\tau})\left(\mathbf{y}_i-\bar{\mathbf{y}}_{n-\tau}\right)^T$$

and

$$\mathbf{B}' = \tau\left[\mathbf{S}_\tau + \left(1-\frac{\tau}{\lambda_1'}\right)(\bar{x}_\tau-\phi_1)(\bar{x}_\tau-\phi_1)^T\right] + (n-\tau) \cdot \left[\mathbf{S}_{n-\tau} + \left(1-\frac{(n-\tau)}{\lambda_2'}\right)(\bar{x}_{n-\tau}-\phi_2)(\bar{x}_{n-\tau}-\phi_2)^T\right] + \mathbf{B}.$$

2.3. Single Change Point Detection in the General Linear Regression

[18] *Rasmussen* [2001] considered the Bayesian estimation of change point in the general linear model for which the mean at a given time i is a linear combination of M basis functions g_k (i), $i = k, .., M$. The basis functions g_k () are function of the observation time, and may just represent a time series of explanatory variables such as precipitation or temperature:

$$\mathbf{Y}_t \begin{cases} N\left(\sum_{k=1}^{M} b_k^1 g_k(t), \sigma^2\right), & t = 1, \ldots, \tau \\ N\left(\sum_{k=1}^{M} b_k^2 g_k(t), \sigma^2\right), t = \tau+1, \ldots, n. \end{cases}. \tag{7}$$

[19] *Rasmussen* [2001] takes advantage of the fact that for a given value of τ, equation (7) can be written in matrix form as a plain linear regression equation:

$$\mathbf{Y} = \mathbf{F}_\tau \boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{8}$$

[20] Assuming a uniform distribution for the elements of $\boldsymbol{\theta}$, for $\log(\sigma)$ and for any parameter of the basis functions, the posterior distribution of the date of change is obtained:

$$P(\tau|\mathbf{Y}) = \frac{|\mathbf{F}_{\tau-1}^T\mathbf{F}_{\tau-1}|^{-1/2}\left[\mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{F}_{\tau-1}\left(\mathbf{F}_{\tau-1}^T\mathbf{F}_{\tau-1}\right)^{-1}\mathbf{F}_{\tau-1}^T\mathbf{Y}\right]^{-(n-M)/2}}{\sum_{i=1}^{n-1}|\mathbf{F}_i^T\mathbf{F}_i|^{-1/2}\left[\mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{F}_i\left(\mathbf{F}_i^T\mathbf{F}_i\right)^{-1}\mathbf{F}_i^T\mathbf{Y}\right]^{-(n-M)/2}}. \tag{9}$$

3. General Change Point Detection in Multiple Linear Regression Using Gibbs Sampling

[21] We now present the model that is proposed in the paper. For extensive details on the model derivation, the reader is referred to *Asselin et al.* [2005].

3.1. Formulation

[22] The simplest formulation for the change point model assumes that the random vector $\mathbf{Y}_t$ has a probability density function (pdf) f for $t = 1,\ldots,\tau$ and pdf g for $t = \tau + 1,\ldots,n$. The case $\tau = n$ stands for the absence of a change point. It follows that the likelihood of τ is

$$L(\tau|\mathbf{Y}) = \prod_{t=1}^{\tau} f(\mathbf{Y}_t).\prod_{t=\tau+1}^{n} g(\mathbf{Y}_t) \tag{10}$$

with $\prod_{t=n+1}^{n} g(.) = 1$, if $\tau = n$. For known pdfs f and g, the maximum likelihood estimate (MLE) for τ can be directly obtained by calculating the likelihood (10) for each $\tau \in \{1, 2,\ldots,n\}$. When we consider parametric families $f(\cdot|\xi)$ and $g(\cdot|\zeta)$, the likelihood becomes

$$L(\tau,\xi,\zeta|\mathbf{Y}) = \prod_{t=1}^{\tau} f(\mathbf{Y}_t|\xi)\prod_{t=\tau+1}^{n} g(\mathbf{Y}_t|\zeta).$$

[23] Obtaining maximum likelihood estimates of τ, ξ and ζ can be challenging for problems involving a large number of unknown parameters, even when using numerical methods as opposed to mathematical inference. A Bayesian formulation of the change point problem gives an alternate approach to inferring on the parameters. Assuming a prior $p(\tau, \xi, \zeta)$ for the parameters, the joint distribution of data and parameters is

$$L(\tau,\xi,\zeta|\mathbf{Y})p(\tau,\xi,\zeta), \tag{11}$$

which is proportional to the joint posterior distribution of τ, ξ and ζ. Obtaining the exact posterior marginal distribution of the parameter τ requires the integration of (11) with respect to ξ and ζ. However, this might not be practical in high-dimensional problems. In such cases, we prefer to approximate the posterior distribution using Markov chain Monte Carlo methods as will be discussed later.

[24] We now present the multivariate regression model with a change point that is developed in this work. Multivariate regression is widely used in applied science to describe relationships between observation series. The change at time τ can be interpreted as a natural of anthropogenic perturbation of the system dynamics.

[25] Our strategy in this paper is to redesign the change point model into a multivariate regression model, so that normal theory can be used and applied whenever possible. This greatly simplifies analytical developments. The change point model (redesigned later in this section) assumes that the $(r \times 1)$ vector $\mathbf{Y}_t$ is related to the $(r \times m^*)$ matrix $\mathbf{X}_t$ by

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\theta}_t^{(\tau)} + \boldsymbol{\upsilon}_t, \tag{12a}$$

where

$$\boldsymbol{\theta}_t^{(\tau)} = \begin{cases} \boldsymbol{\beta}_1^*, & 1 \le t \le \boldsymbol{\tau} \\ \boldsymbol{\beta}_2^*, & \boldsymbol{\tau} < t \le n \end{cases} \tag{12b}$$

under the constraints

$$\boldsymbol{\beta}_1^* = (\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_0)^T \ \text{ and } \ \boldsymbol{\beta}_2^* = (\boldsymbol{\beta}_2, \boldsymbol{\beta}_0)^T. \tag{12c}$$

[26] The dimensions of the vectors $\boldsymbol{\theta}_t^{(\tau)}$, $\boldsymbol{\beta}_1{}^*$, $\boldsymbol{\beta}^*$, $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ are respectively $(m^* \times 1)$, $(m^* \times 1)$, $(m^* \times 1)$, $(m_0^* \times 1)$, $(m_1^* \times 1)$ and $(m_1^* \times 1)$. Equation (12c) implies that $m^* = m_0^* + m_1^*$. It is also assumed that error terms $\{\boldsymbol{\upsilon}_t\}$ are independent and identically distributed following $N[0, \Sigma_y]$.

[27] Model (12a) assumes a change point in the vector $\boldsymbol{\theta}_t^{(\tau)}$ from the subvector $\boldsymbol{\beta}_1$ to the subvector $\boldsymbol{\beta}_2$. The subvector $\boldsymbol{\beta}_0$ is assumed to remain part of $\boldsymbol{\theta}_t^{(\tau)}$ throughout the observation series. This feature allows to model, as a special case, a change point in the intercept parameter.

[28] By defining $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_0)^T$ and

$$\boldsymbol{\Delta}_t^{(\tau)} = \begin{pmatrix} \boldsymbol{\delta}_t^{(\tau)}\mathbf{I}_{m_1^*} & \left(1 - \boldsymbol{\delta}_t^{(\tau)}\right)\mathbf{I}_{m_1^*} & 0 \\ 0 & 0 & \mathbf{I}_{m_0^*} \end{pmatrix},$$

where $\mathbf{I}_{m_0^*}$ and $\mathbf{I}_{m_1^*}$ are the identity matrixes of dimension m_0^* and m_1^*, and

$$\delta_t^{(\tau)} = \begin{cases} 1, & t \le \tau \\ 0, & t > \tau, \end{cases}$$

model (12a) can be written in a simpler form as

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\Delta}_t^{(\tau)}\boldsymbol{\theta} + \boldsymbol{\upsilon}_t. \tag{13}$$

[29] Hence, with the knowledge of the time t of the change point, the change point structure can be redesigned as a single multivariate regression equation. This fact allows the use of the general and well known linear model

$$\mathbf{Y}_t = \mathbf{F}_t\boldsymbol{\theta} + \boldsymbol{\upsilon}_\mathbf{t}, \tag{14}$$

where $\mathbf{F}_t$ is any ($r \times m$) design matrix. Conditional distributions based on (14) are derived in section 3.3. Models (12a), and (13) correspond to the special case $\mathbf{F}_t = \mathbf{X}_t\boldsymbol{\Delta}_t^{(\tau)}$. Note that in practice, there may be a continuity constraint at the change point even if some parameters undergo an abrupt change. This is the case when a time series temporarily displays a trend due to a progressive increase of a causal factor. In this case the expressions of $\boldsymbol{\theta}$ and $\boldsymbol{\Delta}_t^{(\tau)}$ are different of those presented in the preceding sections. These expressions are given in Appendix A for the case of two linear relationships before and after the change point, and the case of a linear relationship followed by a constant mean, assuming that the mean is continuous at the change point.

3.2. Monte Carlo Markov Chain

[30] To make inference on a parameter of a Bayesian model, it will be necessary to integrate the joint posterior probability with respect to all the other parameters. Except in very simple cases where the solution is analytical, this integration is carried out through computer simulation. The idea of studying the stochastic properties of a random variable through computer simulation is not recent [see *Metropolis and Ulam*, 1949]. Contributions from *Metropolis et al.* [1953] and *Hastings* [1970] led to a general method nowadays referred to as the Metropolis-Hastings algorithm. When all conditional distributions are known, Gibbs sampling [*Geman and Geman*, 1984] is preferred to the Metropolis-Hastings algorithm because it leads to less numerical problems. The power of the Metropolis-Hastings algorithm and the Gibbs sampler is undeniable. They allow Bayesian analysis of highly complicated models even when exact closed-form solutions are theoretically impossible to obtain.

3.3. Conditional Distributions

[31] To simplify the developments, an approach similar to the one proposed by *Gelman et al.* [1995] is adopted. Only relevant results are presented, and the reader is referred to *Gelman et al.* [1995] for details on the derivation of conditional distributions for multivariate regression models. The extension to missing data is given in Appendix B.

[32] First, model (14) is expressed into the equivalent univariate multiple regression representation by stacking the observed $\mathbf{Y}_t$ s in a single vector $\mathbf{Y}^{\mathrm{v}}$. Hence we define

$$\begin{aligned}\mathbf{Y}^{\mathrm{v}} &= \left(\mathbf{Y}_1^T, \mathbf{Y}_2^T, \ldots, \mathbf{Y}_n^T\right)^T,\\ \mathbf{F} &= \left(\mathbf{F}_1^T, \ldots, \mathbf{F}_n^T\right),\\ \upsilon^{\mathrm{v}} &= \left(\upsilon_1^T, \upsilon_2^T, \ldots, \upsilon_n^T\right)^T,\end{aligned}$$

where $\mathbf{Y}^{\mathrm{v}}$ is the ($nr \times 1$) vector of observations, $\mathbf{F}$ is the ($nr \times m$) matrix of explanatory variables, and $\boldsymbol{\upsilon}^{\mathrm{v}}$ is the ($nr \times 1$) multivariate normal vector of residuals with zero mean. Under the assumption of normality of the residual vector υ^{v}, and considering a multivariate Normal prior for the parameter vector theta with a vector of means $\mathbf{f}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, the posterior distribution of $\mathbf{Y}^{\mathrm{v}}$ is

$$\mathbf{Y}^{\mathrm{v}}|\mathbf{F}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y \quad \mathrm{N}\left[\mathbf{F}\boldsymbol{\theta}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}_{\mathbf{y}}\right] \tag{15}$$

and

$$\begin{pmatrix}\mathbf{Y}^{\mathrm{v}}\\ \boldsymbol{\theta}\end{pmatrix}|\mathbf{F}, \boldsymbol{\Sigma}_y \quad \mathrm{N}\left[\begin{pmatrix}\mathbf{f}\\ \boldsymbol{\theta}_0\end{pmatrix}, \begin{pmatrix}\mathbf{Q} & \mathbf{S}^T\\ \mathbf{S} & \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\end{pmatrix}\right], \tag{16}$$

where

$$\begin{aligned}\mathbf{S} &= \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{F}'.\\ \mathbf{f} &= \mathbf{F}\boldsymbol{\theta}_0\\ \mathbf{Q} &= \mathbf{I}_n \otimes \boldsymbol{\Sigma}_y + \mathbf{F}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{F}^T.\end{aligned} \tag{17}$$

[33] Using a normal prior for $\boldsymbol{\theta}$, we have

$$\boldsymbol{\theta}|\mathbf{Y}, \mathbf{F}, \boldsymbol{\Sigma}_y \quad \mathrm{N}[\mathbf{m}, \mathbf{C}],$$

where

$$\mathbf{m} = \boldsymbol{\theta}_0 + \mathbf{S}\mathbf{Q}^{-1}(\mathbf{Y}^{\mathrm{v}} - \mathbf{f})$$

$$\mathbf{C} = \Sigma_\theta - \mathbf{S}\mathbf{Q}^{-1}\mathbf{S}^T \tag{18a}$$

$$= \left(\Sigma_\theta^{-1} + \mathbf{F}^T\left(\mathbf{I}_n \otimes \Sigma_y^{-1}\right)\mathbf{F}\right)^{-1}. \tag{18b}$$

[34] If a proper prior for $\boldsymbol{\theta}$ is selected, (18a) is well defined. However, if $|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| \rightarrow \infty$, then (18a) may be computationally undefined, so (18b) should be used when Σ_θ^{-1} is easily obtained and $|\Sigma_\theta^{-1}|$ is finite. Since $\mathbf{Q}$ is ($nr \times nr$), $\mathbf{Q}^{-1}$ should be calculated as

$$\mathbf{Q}^{-1} = \left(\mathbf{I}_n \otimes \Sigma_y^{-1}\right) - \left(\mathbf{I}_n \otimes \Sigma_y^{-1}\right)\mathbf{F}\mathbf{C}\mathbf{F}^T\left(\mathbf{I}_n \otimes \Sigma_y^{-1}\right) \tag{19}$$

with $\mathbf{C}$ obtained from (12b), rather than by directly inverting (17). The advantage of (19) is that Σ_y^{-1} is only ($r \times r$) and $\mathbf{C}$ is only ($m \times m$) in contrast to the ($nr \times nr$) matrix $\mathbf{Q}^{-1}$.

[35] We now focus on the derivation of the conditional distributions of Σ_y given $\boldsymbol{\theta}$. In general, we have

$$\begin{aligned}p\left(\boldsymbol{\Sigma}_y|\mathbf{Y}, \mathbf{F}, \boldsymbol{\theta}\right) &\propto p\left(\boldsymbol{\Sigma}_y|\mathbf{F}, \boldsymbol{\theta}\right)p\left(\mathbf{Y}^{\mathrm{v}}|\mathbf{F}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y\right),\\ &\propto p\left(\boldsymbol{\Sigma}_y|\mathbf{F}, \boldsymbol{\theta}\right)\prod_{t=1}^{n} p\left(\mathbf{Y}_t|\mathbf{F}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y\right)\\ &\propto p\left(\boldsymbol{\Sigma}_y|\mathbf{F}, \boldsymbol{\theta}\right)\left|\boldsymbol{\Sigma}_y\right|^{-n/2}\exp\left(-\mathrm{tr}\left(n\hat{\boldsymbol{\Sigma}}_y\boldsymbol{\Sigma}_y^{-1}\right)/2\right),\end{aligned}$$

where

$$\hat{\boldsymbol{\Sigma}}_y = n^{-1}\sum_{t=1}^{n}\boldsymbol{\upsilon}_t\boldsymbol{\upsilon}_t, \boldsymbol{\upsilon}_t = \mathbf{Y}_t - \mathbf{F}_t\boldsymbol{\theta}. \tag{20}$$

[36] Under model (20) and the assumption of inverse Wishart prior for $\boldsymbol{\Sigma}_y$, i.e., $\Sigma_y|\mathbf{F}, \boldsymbol{\theta}\ \mathrm{W}_{\boldsymbol{\upsilon}}^{-1}(\Lambda_y)$, the conditional posterior distribution of Σ_y is

$$\Sigma_y|\mathbf{Y}, \mathbf{F}, \boldsymbol{\theta}\ \ \mathrm{W}_{\upsilon+n}^{-1}\left(\boldsymbol{\Lambda}_y + \boldsymbol{\upsilon}\hat{\boldsymbol{\Sigma}}_y\right).$$

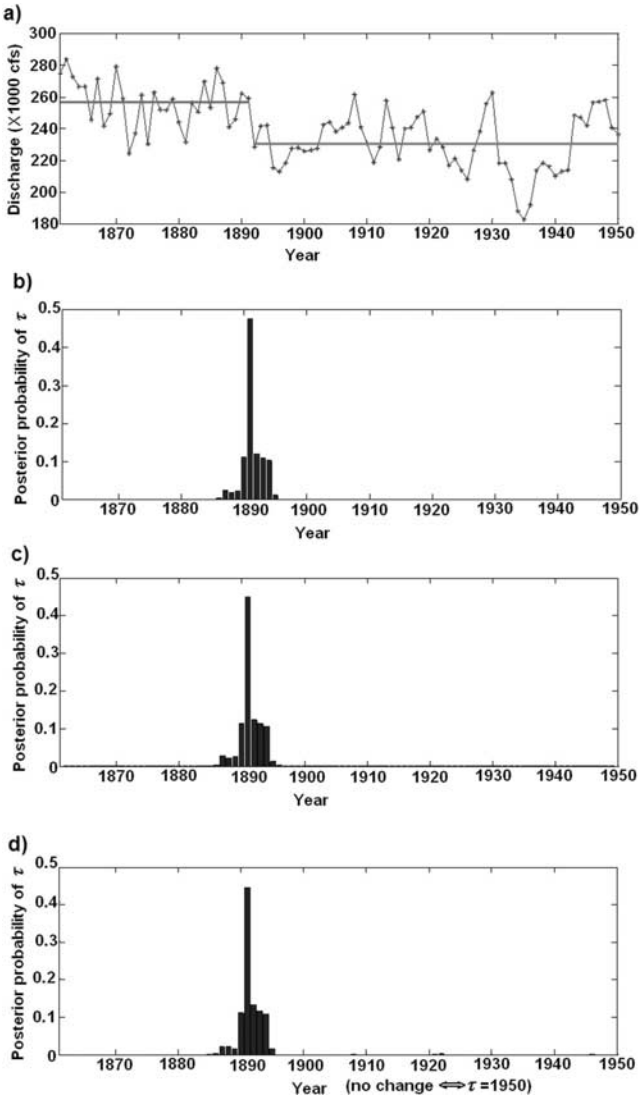

Figure 1. Comparison of the proposed methodology with those of *Rasmussen* [2001] and *Perreault et al.* [2000a] on a single shift detection in the Saint Lawrence streamflow data: (a) discharge, (b) *Perreault et al.* [2000a], (c) *Rasmussen* [2001], and (d) proposed methodology.

[37] In the case when $\boldsymbol{\Sigma}_y = \delta^2\Gamma_y$, where Γ_y is a known positive definite matrix, a conjugate inverse gamma prior (i.e., $\delta^2|\mathbf{F},\ \boldsymbol{\theta}\ \mathrm{G}^{-1}(a,\ b)$) is assumed. The corresponding conditional posterior for δ^2 is

$$\delta^2|\mathbf{Y},\mathbf{F},\boldsymbol{\theta}\ \ \mathrm{G}^{-1}\left(a+\frac{nr}{2},b+\frac{1}{2}\mathrm{tr}\left(n\hat{\boldsymbol{\Sigma}}_y\boldsymbol{\Gamma}_y^{-1}\right)\right).$$

[38] The noninformative case corresponds to imposing $a \to 0$ and $b \to 0$. In the important case of independent response variables ($\Gamma_y = \mathbf{I}_r$), we have

$$\delta^2|\mathbf{Y},\mathbf{F},\boldsymbol{\theta}\ \ \mathrm{G}^{-1}\left(a+\frac{nr}{2},b+\frac{1}{2}\mathrm{tr}\left(n\hat{\boldsymbol{\Sigma}}_y\right)\right).$$

3.4. Change Point Inference

[39] In this section, we focus on the change point inference part of the problem. As noted in the introduction, any change point in a regression formulation can be modeled by a plain regression model conditioned on the time of change point. It is then a matter of "rewriting" the design matrices $\{\mathbf{X}_t\}$ as a single matrix $\mathbf{F}$ given τ and obtaining a conditional posterior for the time of change point.

[40] For any prior $p(\tau|\{\mathbf{X}_t\}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y)$ under model [14] with $\mathbf{F}_t = \mathbf{X}_t\Delta_t^{(\tau)}$ depending on τ, we have

$$\begin{aligned} p(\tau|\mathbf{Y},\{\mathbf{X}_t\},\boldsymbol{\theta},\boldsymbol{\Sigma}_y) &\propto p(\mathbf{F},\boldsymbol{\theta},\boldsymbol{\Sigma}_y|\mathbf{Y}) \\ &\propto p(\mathbf{F},\boldsymbol{\theta},\boldsymbol{\Sigma}_y)p(\mathbf{Y}^{\mathrm{v}}|\mathbf{F},\boldsymbol{\theta},\boldsymbol{\Sigma}_y) \\ &\propto p(\mathbf{F},\boldsymbol{\theta},\boldsymbol{\Sigma}_y)|\boldsymbol{\Sigma}_y|^{-n/2}\exp\left(-\mathrm{tr}\left(n\hat{\boldsymbol{\Sigma}}_y\boldsymbol{\Sigma}_y^{-1}\right)/2\right), \end{aligned} \quad (21)$$

where $\hat{\boldsymbol{\Sigma}}_y$ is obtained from (20). This result can be used to sample τ under any prior assumption on $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}_y$ and the missing values. Equation (21) is the "regression" version of (11): it is the exact posterior density of all unknown parameters. Hence this equation would remain valid for any structure built in $\mathbf{F}$. The use of the Metropolis-Hastings algorithm with (21) provides a general method to generate from the joint posterior of $\{\mathbf{F}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y\}$, although this may be computationally difficult in practice, which explains why direct Gibbs sampling with conjugate priors is often preferred.

[41] Although Gibbs sampling of τ from (21) is always possible (provided that some regularity conditions are satisfied), it is possible to do better under further prior assumptions. In section 3.3, we have assumed a normal prior for $\boldsymbol{\theta}$. With this additional assumption, we can integrate (21) with respect to $\boldsymbol{\theta}$ and we have

$$\begin{aligned} p(\boldsymbol{\tau}|\mathbf{Y},\{\mathbf{X}_t\},\boldsymbol{\Sigma}_y) &\propto p(\mathbf{F},\boldsymbol{\Sigma}_y|\mathbf{Y}) \\ &\propto p(\mathbf{F},\boldsymbol{\Sigma}_y)p(\mathbf{Y}^{\mathrm{v}}|\mathbf{F},\boldsymbol{\Sigma}_y), \end{aligned} \quad (22)$$

where

$$\mathbf{Y}^{\mathrm{v}}|\mathbf{F},\boldsymbol{\Sigma}_y\ \ \mathrm{N}[\mathbf{f},\mathbf{Q}].$$

[42] Since the parameters $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ may be strongly dependent, the use of (22) as opposed to (21) has the desirable feature of reducing the dependencies in the series of Gibbs samplers. Therefore the use of (22) would improve mixing and would speed up convergence to the joint posterior of all parameters. Ideally, we should integrate (22) with respect to $\boldsymbol{\Sigma}_y$ as well, but our prior assumptions render this task very difficult. *Perreault et al.* [2000b] performed successfully a similar integration under a simpler model with more restraining priors.

[43] When choosing the prior for τ, since the particular event $\tau = n$ stands for the absence of a change point, it might be appropriate to place more or less prior probability mass on this event, depending on the question of interest or on the prior knowledge of the data. In their application

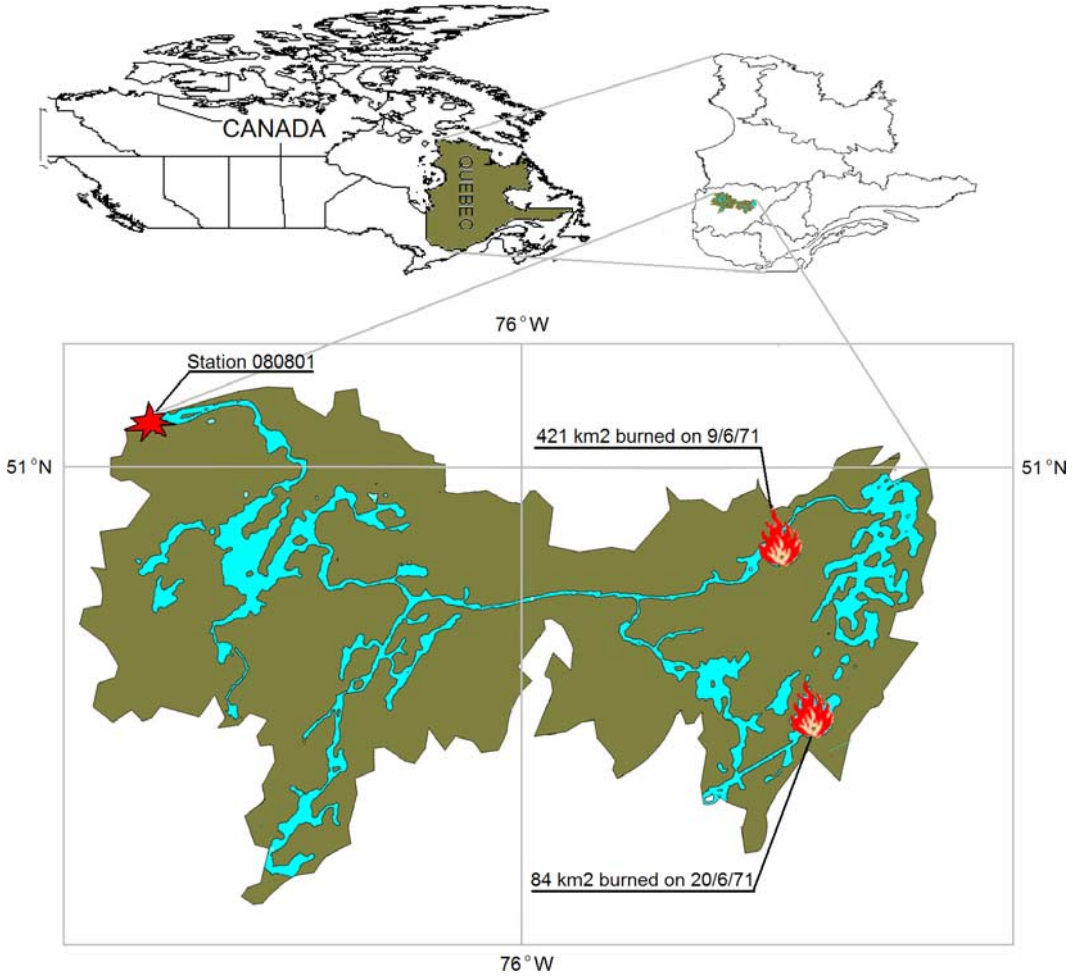

Figure 2. Location map of station 080801.

example, *Carlin et al.* [1992] used the discrete uniform on $\{1, 2, \ldots, n\}$ as a prior pmf for τ.

3.5. Extension to Missing Data

[44] In practice, the data set could contain missing values. Bayesian methods cope with this problem elegantly by replacing the missing values with unknown parameters that are updated in the Gibbs sampling routine in the same way it is done for the parameters of interest. In order to update a missing value through Gibbs sampling, we need its conditional distribution given all other parameters and data. Appendix B provides the conditional distributions that allow Gibbs sampling of missing values in $\mathbf{Y}^{v}$ or $\mathbf{F}$.

4. Case Studies

[45] In this section, three examples in hydrology are analyzed using applicable methods among those that have been described in section 2. The first example was presented by *Rasmussen* [2001]. It was considered to allow for a rational comparison of the original methodologies with the approach proposed in this paper. These examples are as follows.

[46] Example 1: This example deals with a single shift detection in the Saint Lawrence streamflows data at Ogdensbourg, New York. The analysis was performed using the methodology of *Perreault et al.* [2000a] (model (1)), *Rasmussen* [2001] (model (7)) and the proposed methodology (model (12)).

[47] Example 2: This example deals with single change point detection in the multiple linear regression between mean basin-scale precipitation at four different periods of the year and the summer–autumn flood peaks of the Broadback River located in northern Quebec, Canada. Inference was performed for models (7) and (12).

[48] Example 3: The data of five rivers located in the Côte-Nord region of the province of Quebec, Canada, are investigated for a single shift using model (12),. Model (4) [*Perreault et al.*, 2000b] could not be used in this case because of several gaps in the observations.

4.1. Single Shift Detection in the Saint Lawrence Streamflow Data

[49] We consider the 1861–1950 annual streamflows of the Saint Lawrence River at Ogdensbourg, New York. This data set was analyzed by *Rasmussen* [2001]. The data are plotted in Figure 1a and seems to indicate that mean annual flow of that river displays either a downward trend of a negative shift. As this example is very simple, all the models presented in section 2 can be used except that of *Perreault et al.* [2000b] which is intended to work on multivariate data sets only. Models (1), (7) and (12), were

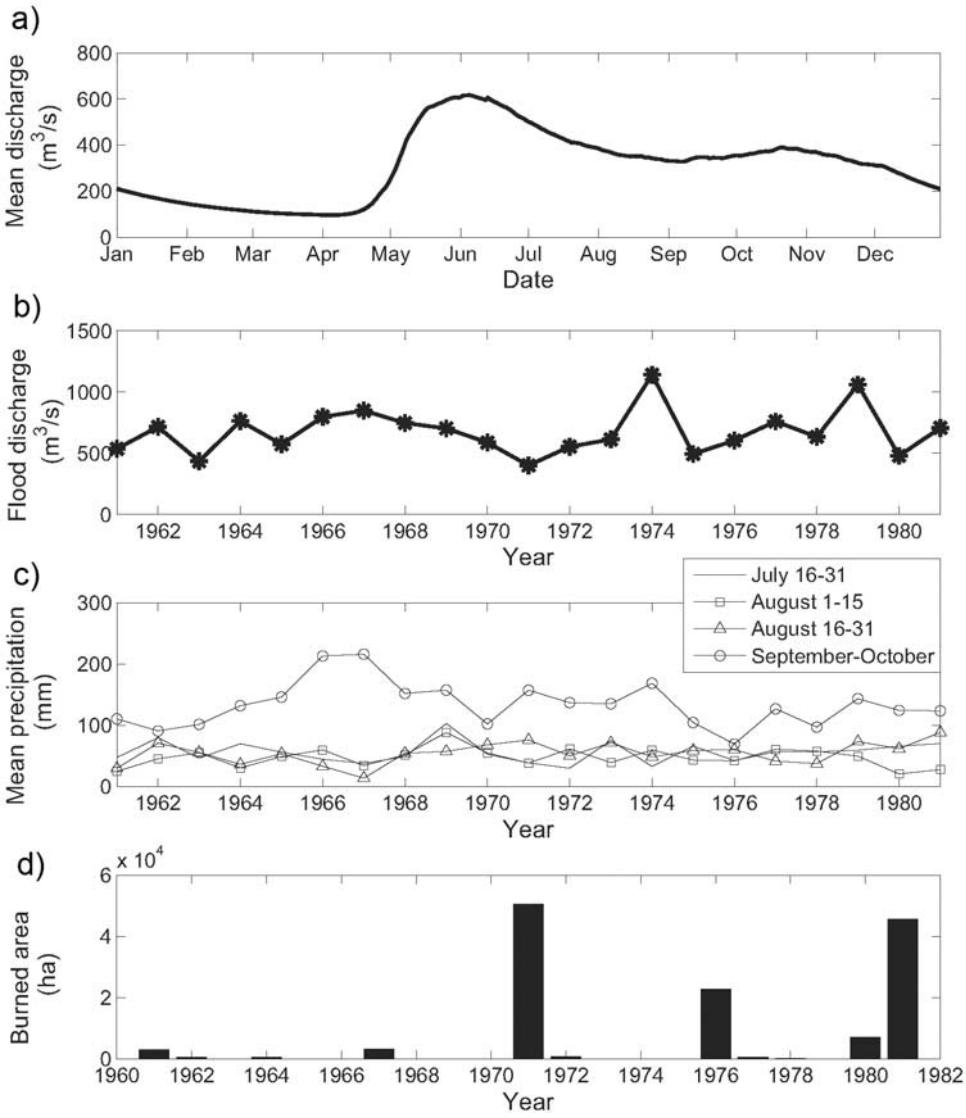

Figure 3. Data for change point detection in summer–autumn flood peaks of the Broadback River: (a) mean hydrograph, (b) summer–autumn flood peak time series, (c) precipitation time series, and (d) burned area time series.

thus applied to the data set. As there are no explicative variables in this example, the vector $\mathbf{X}$ in the regression equation is a simple column for which each element has value 1.

4.1.1. Prior Specification and Inferences on Model Parameters

[50] The posterior distributions for model (3) [*Rasmussen*, 2001] were obtained using Jeffrey's noninformative priors for the parameters (the Jeffrey's prior is a noninformative prior distribution which is invariant under reparameterization of the parameter vector. It thus expresses the same belief no matter the scale used). Consequently, no prior specifications are required for this particular approach. The prior distributions for the parameters of the two other models were thus set to be noninformative in order to allow for a rational comparison of the various approaches. τ was assumed to be uniformly distributed over $\{1,\ldots,n\}$ for all models. The parameters α and β for model (1) were set to 2 and var($\mathbf{Y}$) which corresponds to an inverse gamma distribution of mean var($\mathbf{Y}$) and infinite variance [see *Perreault et al.*, 2000a], while λ was set to 10000. For model (12), the prior mean for $\boldsymbol{\theta}$ was set to the sample mean, and the prior variance of $\boldsymbol{\theta}$ was set to 10000 times the sample variance.

[51] The posterior distributions of models (1) and (4) were obtained using their analytical expressions (equations (3) and (9)). To make inferences on the parameters of model (12), 10000 iterations of the Gibbs sampler were performed. Convergence was successfully assessed at iteration 100. Inferences on model parameters were performed using the 9900 last iterations.

4.1.2. Results

[52] The posterior distributions of the date of change are plotted in Figures 1b, 1c, and 1d for models (1), (7) and (12), respectively. It appears that the three models display the same shape for the posterior probability of the date of

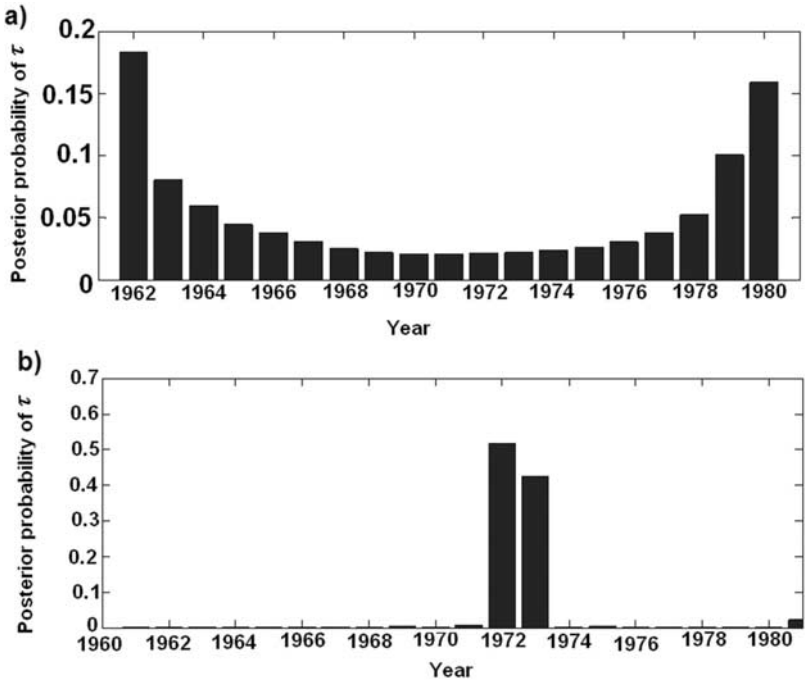

Figure 4. Change point detection in summer–autumn flood peaks of the Broadback River: (a) posterior probability of change point obtained with the methodology of *Rasmussen* [2001] and (b) posterior probability of change point obtained with the proposed methodology.

change: the mode and 95% credibility interval obtained with all the three models are 1891 and [1886 1894]. The results of model (1) [*Perreault et al.*, 2000a] and (7) [*Rasmussen*, 2001] are particularly similar, although there are very small differences in the posterior distributions because of different model parameterizations. Model (12), gives a posterior distribution that is also very close to the two others. Note that it was not necessarily expected that empirical distributions computed from MCMC chains would fit exactly the analytical solution. Variability due to numerical errors and the limited size of MCMC chains will always be present. The results presented in Figure 1 are thus very satisfying and can be considered as a successful validation of the proposed methodology for the case of univariate normal data with a single shift.

4.2. Change Detection in a Multivariate Regression Model: Influence of Forest Fires on Summer–Autumn Flood Peaks of the Broadback River

[53] The change point detection methods will now be applied to the relationship between summer–autumn maximum flood discharge and precipitation at station 80801 located on the Broadback River, Quebec, Canada. This river has a catchment of 17100 km2 and experiences from time to time forest fire bursts (Figure 2). According to the Canadian Large Fire Database [*Stocks et al.*, 2002] (see also Natural Resources Canada, Canadian large fires database, 2005, available at http://fire.cfs.nrcan.gc.ca/Downloads/LFDB/LFD_5999_e.ZIP), major forest fires occurred during the summer of 1971, burning 506 km^2 in the upper parts of the catchment (1/34 of the total basin area). It is hypothesized that the deforestation due to these fires has changed the basin response function to meteorological inputs. In order to perform the analysis, the 1961–1981 daily flood discharges at station 80801 were obtained from Quebec Ministry of Environment. The Broadback River is subject to two types of floods: spring flood, which are dominated by snowmelt, and summer–autumn floods which are caused by direct liquid precipitation. Figure 3a presents the mean daily discharge at this station for 1961–1981. It appears that the summer–autumn maximum flood peak is generally observed at the end of October (Figure 3a). Daily precipitation of July–October 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 mean precipitation on the Broadback River catchment for every half month 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 mean precipitation of 16–31 July, (2) the sum of precipitation of 1–15 August, (3) the sum of precipitation of 16–31 August and (4) the sum of precipitation of September–October. The values of 1961–1981 summer–autumn flood peaks are presented in Figure 3b and those of the chosen explanatory variables in Figure 3c. Figure 3d presents the burned areas on the catchment for each year of the period of study.

4.2.1. Prior Specification and Inferences on Model Parameters

[54] An equal weight was set for the probability of change ($\tau = 1,\ldots,n-1$) and the absence of change ($\tau = n$). Note that this prior is more realistic than the flat prior in example 1 because it acknowledges the fact that a change may not happen. The prior for $\boldsymbol{\theta}$ was set as follows: since in this application $\boldsymbol{\mu}_t = \mathbf{F}_t\boldsymbol{\theta}$ represents the expectation of the flood

Table 1. Mean Value and Credibility Intervals Before and After the Change Point for the Coefficients of the Linear Regression Describing the Relationship Between Summer–Autumn Flood Peaks and Precipitation on the Broadback River's Basin

	Before the Change Point		After the Change Point	
	Mode	95% Credibility Interval	Mode	95% Credibility Interval
Coefficient of the sum of precipitation of 16–31 July	4.69	[0.52 7.98]	−6.83	[−12.25 0.83]
Coefficient of the sum of precipitation of 1–15 August	−0.18	[−5.25 4.30]	6.90	[1.61 11.92]
Coefficient of the sum of precipitation of 16–31 August	−0.32	[−3.89 3.00]	6.43	[−0.50 10.82]
Coefficient of the sum of precipitation of September–October	2.99	[1.87 4.26]	3.38	[1.05 5.94]

peak at date t, it seems reasonable to give to its mean a prior distribution for which the 95% lower confidence interval is positive, i.e., $\mathbf{F_t}\hat{\boldsymbol{\theta}}^p - 1.96\mathbf{F}_t\boldsymbol{\Sigma}_\theta^p\mathbf{F_t^T} > 0$, $t = 1, .., n$ where $\hat{\boldsymbol{\theta}}^p$ and $\boldsymbol{\Sigma}_\theta^p$ represent the prior mean and the prior variance for θ. These considerations led to $\hat{\boldsymbol{\theta}}^p = \hat{\boldsymbol{\theta}}_{\mathbf{reg}}$ and $\hat{\boldsymbol{\Sigma}}_\theta^p = k\hat{\boldsymbol{\Sigma}}_\theta^{reg}$ where $\boldsymbol{\theta}^{reg}$ and $\hat{\boldsymbol{\Sigma}}_\theta^{reg}$ are the mean and variance of the regression parameters obtained using ordinary least squares, and $k = \max\{k|\forall t \in \{1,\ldots,n\}, \mathbf{F}_t\boldsymbol{\theta}^p - 1.96k\,(\mathbf{F_t}\,\hat{\boldsymbol{\Sigma}}_\theta^{\mathbf{p}}\,\mathbf{F_t^T})^{0.5} > 0\}$.

[55] As for the first example, 10000 iterations of the Gibbs sampler were performed. Convergence was successfully assessed at iteration 100. Inferences on model parameter were performed using the 9900 last iterations.

4.2.2. Results

[56] Figure 4a presents the posterior probability of the date of change in the linear relationship that was obtained with the approach of *Rasmussen* [2001]. The maximum posterior distribution of the date of change is maximal at the beginning and at the end of the series, and displays no peak. This kind of shape of posterior distribution of date of change is typical of model (7) when applied to homogeneous series. Thus the application of this approach leads to a 'no change' conclusion.

[57] The posterior probability of the date of change obtained with model (12), is given in Figure 4b. The mode and credibility interval for this distribution are 1972 and (1972, 1978) respectively. It shows a clear peak in 1972 leading to a strong conclusion of change between 1972 and 1973. The mode and credibility intervals of the posteriors distributions of each coefficient of the linear regression before and after the change point were also computed from the MCMC chains and listed in Table 1. The posterior probability distributions of each of these coefficients before and after the change point are provided in Figure 5. Inspection of Figure 5 shows that the weight of the sum of precipitation of 16–31 July decreased from positive to negative values while that of the sums of precipitation of 1–15 August and 16–31 August increased significantly. The negative values in the regression coefficients after the change point can be explained by the dependence between the sums of precipitation of consecutive periods. This dependence could have been removed using techniques such as principal component analysis (PCA), but such task is beyond the scope of this paper and is not supposed to change the existence and date of change in the linear relationship. The uncertainty on the regression coefficients is also higher after the change point since the 95% credibility interval is wider in all cases (Table 1), and the distributions have a larger support (Figure 5).

[58] Since the two approaches give dramatically different results, an alternative procedure was sought to check whether there was a change in 1972 or not. We used the change point detection method in multivariate regression described by *Vincent* [1998]. This method is applied as follows.

[59] 1. For each year i between 1961 and 1981: (1) Fit two linear regression models to segments [1961–i] and [i + 1, 1981] and compute the series of residuals. (2) Test the autocorrelation of the residuals using the Durbin-Watson test at 5% significance level. (3) If the test is positive, use the classical F test at 5% significance level to compare the model with a change point at year i with the model without change point (See Vincent, 1988 for details on how to compute the test statistics and the associated critical values).

[60] 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.

[61] The application of the above-mentioned method gave two potential dates of change (1972 and 1973) with a higher F statistics for 1972. While the residuals of the model without change point were found autocorrelated, no significant autocorrelation was found in the residuals of the fitted model. Figure 6 presents the normal probability plot of the residuals given a change in 1972. It can be seen that the residuals are reasonably normal, as required by linear regression theory.

[62] Hence the method of *Vincent* [1998] supports the hypothesis of change in 1972. The main reason for which the model of *Rasmussen* [2001] failed to detect the change point is the use of the Jeffrey's noninformative prior for the regression parameters. The choice of Jeffrey's noninformative prior is motivated by mathematical convenience, but such a choice gives the same prior probability for each value between $-\infty$ and $+\infty$, while the range of physically possible values is much smaller. It is well known in Bayesian statistics that when the prior is too vague, weak signals in the observations series are not detected. The prior used with the proposed method is vague but at a much less extent than the Jeffrey's noninformative prior (see section 4.2.1). It led thus to the clear detection of changes in the data series.

4.3. Single Shift Detection in a Multivariate Data Set With Missing Data

[63] Five rivers in the Côte-Nord region (province of Quebec, Canada) were selected for this application. These stations have a sufficiently long common period of observation to set up the prior distributions. As all the other stations of the same hydrological region display a significant amount of missing data, only the approach proposed in this paper can be used in this case. The selected rivers are the Godbout River (station 71401), the Moisie River

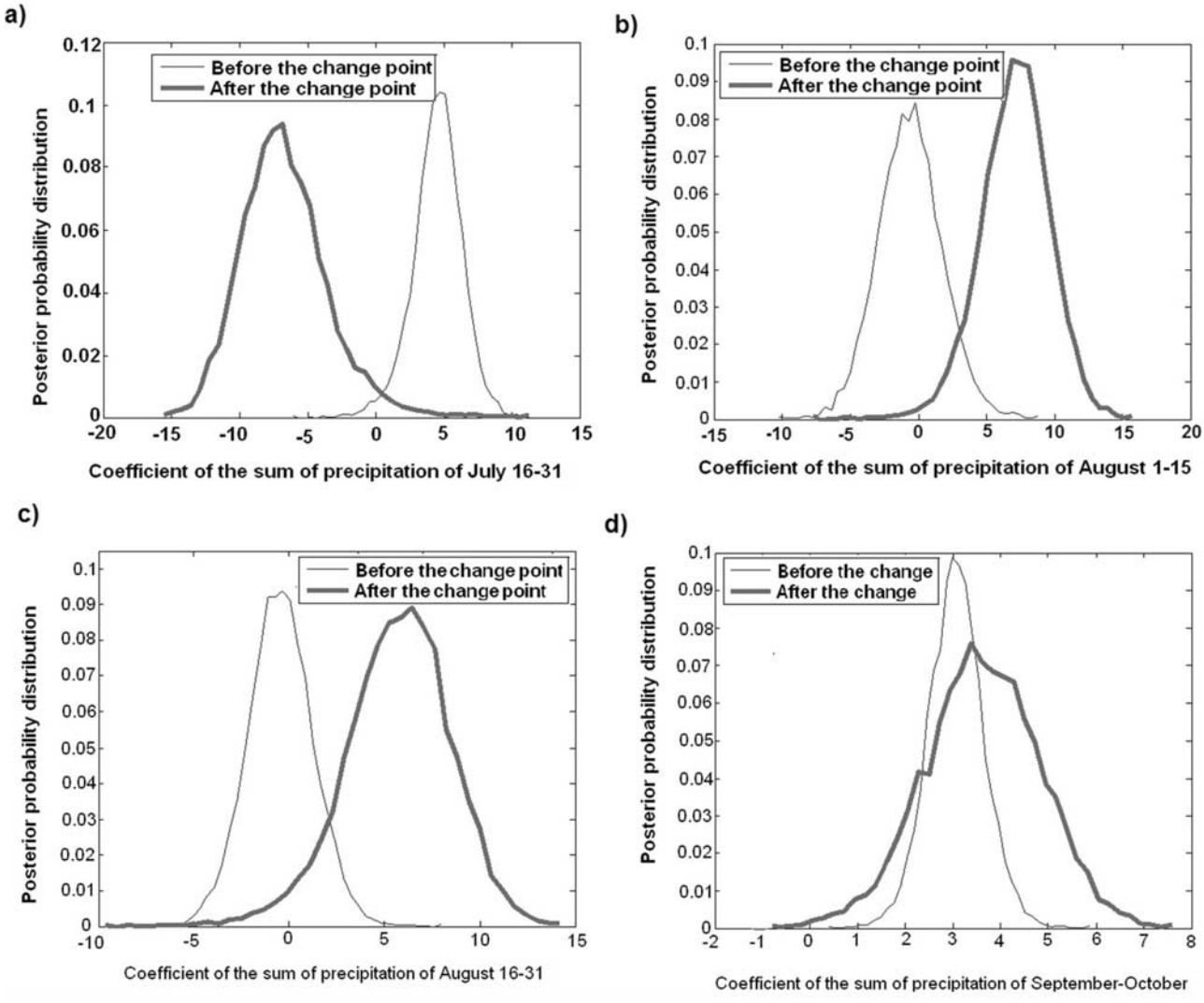

Figure 5. Posterior probability distributions of the coefficients of the linear regression describing the relationship between summer–autumn flood peaks and precipitation on the Broadback River's basin.

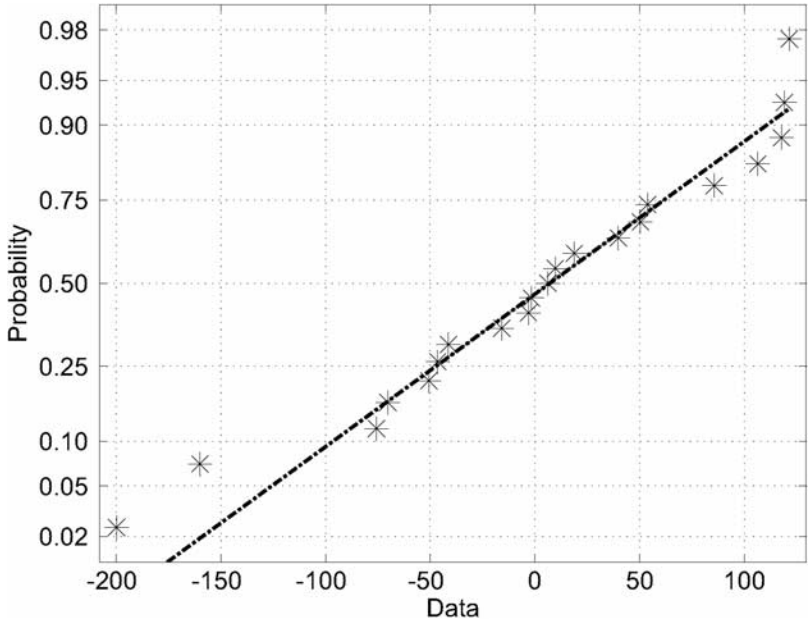

Figure 6. Normal probability plot of the residuals given a change in 1972.

Table 2. Characteristics of the Five Rivers of Northern Quebec

Station Number	Station Name	Longitude	Latitude	Catchment Area, km^2	Observation Period	Number of Missing Values Between 1957 and 2001
71401	Godbout	−67.65	49.33	1 570	1972–2000	19
72301	Moisie	−66.18	50.35	19 000	1966–2001	12
73503	Magpie	−64.58	50.68	7 230	1971–2001	22
73801	Romaine	−63.62	50.30	13 000	1957–2001	1
74601	Nabissipi	−62.21	50.25	2 060	1963–1987	20

(station 72301), the Magpie River (station 73503), the Romaine River (station 73801) and the Saint Paul River (station 74601), which all have observations during the period 1975–1987. The characteristics of these rivers are listed in Table 2, and their annual maximum flood peaks are plotted in Figure 7a.

4.3.1. Prior Specification and Inferences on Model Parameters

[64] The prior specification for $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ are the same as in section 4.2.1 except that only the common period of observation was used to compute $\hat{\boldsymbol{\theta}}^{\text{reg}}$, $\hat{\boldsymbol{\Sigma}}_{\theta}^{reg}$ and k. There are no covariates in this model, except the intercept, given as a

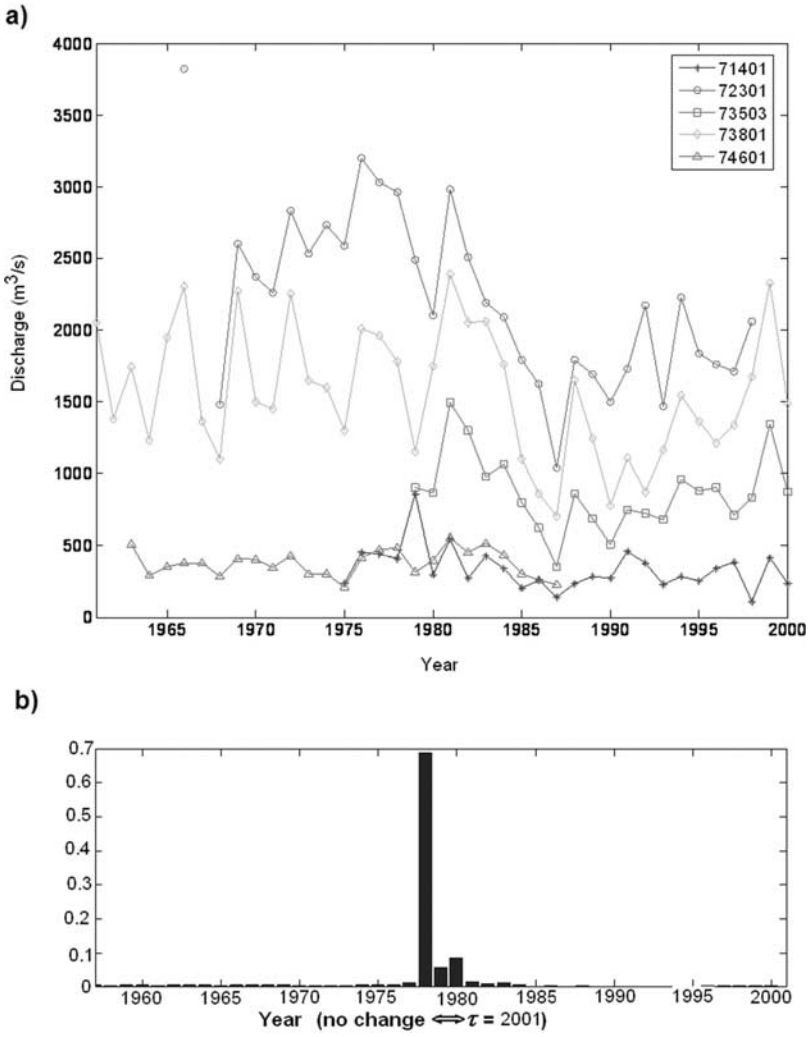

Figure 7. Change point detection on the five rivers of northern Quebec: (a) flood peak time series and (b) posterior probability of change point.

column of ones. Jeffrey's noninformative prior was first used for $\mathbf{\Sigma}_y$ ($v \to -1$ and $|\Lambda_y| \to 0$). The flood discharges times series were also standardized to verify the hypothesis of common variance assumed by the proposed method. 100000 iterations of the Gibbs sampler were performed and convergence was successfully assessed after iteration 1000. The number of iterations is higher because there is much more variables to update than in the first four examples.

4.3.2. Results

[65] Figure 7b presents the posterior probability distribution of the date of change for model (12). Inspection of Figure 7 shows that the posterior probability of the date is concentrated between 1977 and 1984 with clear peak in 1977. The posterior probability of the date of change is maximal in 1978. It could be concluded that there is an evidence of regional change in river flows of the Côte-Nord region in the province of Quebec around 1978.

[66] The most interesting aspect of this application is the straightforward estimation of missing data in a context of nonstationarity. As mentioned earlier, there was a significant number of gaps in the streamflow data of the Côte-Nord region. Estimation of the missing values is not an easy task even with a stationarity hypothesis. The proposed methodology addresses this issue in a straightforward manner, and the obtained posterior distributions allow a full assessment of the uncertainty associated with the results. The reconstituted streamflows in which missing values are estimated by the mean of their posterior distributions are given in Figure 8. The credibility intervals for missing data are also provided in Figure 8.

5. Discussion

[67] The three case studies presented in this paper show that the proposed approach is very flexible and can be applied to a wide range of problems in hydrology. In Example 1, it is compared to published change point detection approaches with the same priors and data and it gave exactly the same results. In example 2, it is shown that it gives better results than *Rasmussen* [2001] on the problem of change point detection in summer–autumn flood peaks of the Broadback River probably because it allows for a more realistic but still vague prior specification on regression parameters as well as on the variance parameter.

[68] This flexibility leads to nonexplicit solutions for the posterior probability distributions, thus to MCMC simulations, while the approaches of *Rasmussen* [2001] and *Perreault et al.* [2000a, 2000b] provide posterior distributions in closed forms. However, model flexibility is a requirement for a realistic analysis of hydrological data sets and the proposed methodology can be applied to a much broader range of problems: for instance, example 3 is of particular importance for hydrologists since it also allows the estimation of missing data in a nonstationary context, along with a full uncertainty assessment of the results. The posterior probability distribution of the missing data takes into account the uncertainty on the date of change, on regression parameters as well as on the variance-covariance structure. The results are thus much more informative than any classical estimation with confidence intervals often based on unverified regularity hypotheses.

[69] A number of other hydrological problems can be analyzed with the change point detection methodology such as homogenization of historical data or estimation of missing data in the explanatory variables. An interesting but quite straightforward topic of further work would be the generalization of the approach to multiple change point problems.

[70] Another interesting future development would be the extension of the approach to the analysis of series of unequal variances between series and/or before and after the change point. For instance, multivariate data sets have been standardized before applying the methodology provided in this paper. If the shifts in the mean were very severe (e.g., more than five times the standard deviation), the variance of the standardized series would have be dramatically different before and after the change and the proposed method does not apply. Generalization to thz.

[71] Finally, as one may not know how much change points exist in a given series, the extension to multiple change points is desirable. A possible solution in this case is to use the method recursively (i.e., segments between consecutive change points of between a change point and the limits of the series are tested until all segments are found homogeneous). See for example *Beaulieu et al.* [2007] for examples of application of method designed for one inhomogeneity to multiple change point problems. Another solution for multiple changes is to exploit the same idea for more complicated models, including a multiple change point model or the more general segmented multivariate regression. The latter is formally described as

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\beta}_i + \boldsymbol{\upsilon}_t, \quad \text{if } x_t \in]\boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_i],$$

for $t = 1,\ldots,n$ and $i \in \{1,\ldots,l\}$. The $(r \times 1)$ observations $\{\mathbf{Y}_t\}$ are modeled as piecewise regressions depending on the covariates $\{\mathrm{x}_t\}$. $\{\mathbf{X}_t\}$ are $(r \times m)$ design matrices, $\{\boldsymbol{\beta}_i\}$ are $(m \times 1)$ regression parameters, and $\{\boldsymbol{\upsilon}_t\}$ are $(r \times 1)$ residual vectors. A natural approach for the analysis of this model is to obtain estimation equations for the β_i s separately. However, this problem is greatly simplified if the model is written as its equivalent multivariate regression form. Define the $(1 \times l)$ row vector

$$\delta^{(t)} = \left(\overbrace{0, 0, \ldots, 0, 1}^{i}, 0, \ldots, 0\right), \quad \text{if } x_t \in (\alpha_{i-1}, \alpha_i].$$

[72] With $\boldsymbol{\theta}^T = (\boldsymbol{\beta}_1^T,\ldots,\boldsymbol{\beta}_l^T)$, the segmented multivariate regression is equivalent to

$$\mathbf{Y}_t = \mathbf{X}_t\left(\delta^{(t)} \otimes \mathbf{I}_m\right)\boldsymbol{\theta} + \boldsymbol{\upsilon}_t.$$

[73] Under the appropriate assumptions on the residuals $\{\boldsymbol{\upsilon}_t\}$, results of sections 4 and 5 are immediately applicable. With $\mathbf{F}_t = \mathbf{X}_t\ (\delta^{(t)} \otimes \mathbf{I}_m)$, the conditional posterior (18), (or (19) if $\boldsymbol{\theta}$ has a normal prior) can be used to obtain the conditional posterior of the parameters $\{\boldsymbol{\alpha}_i\}$ and perform their Gibbs sampling. There is no doubt that the same idea

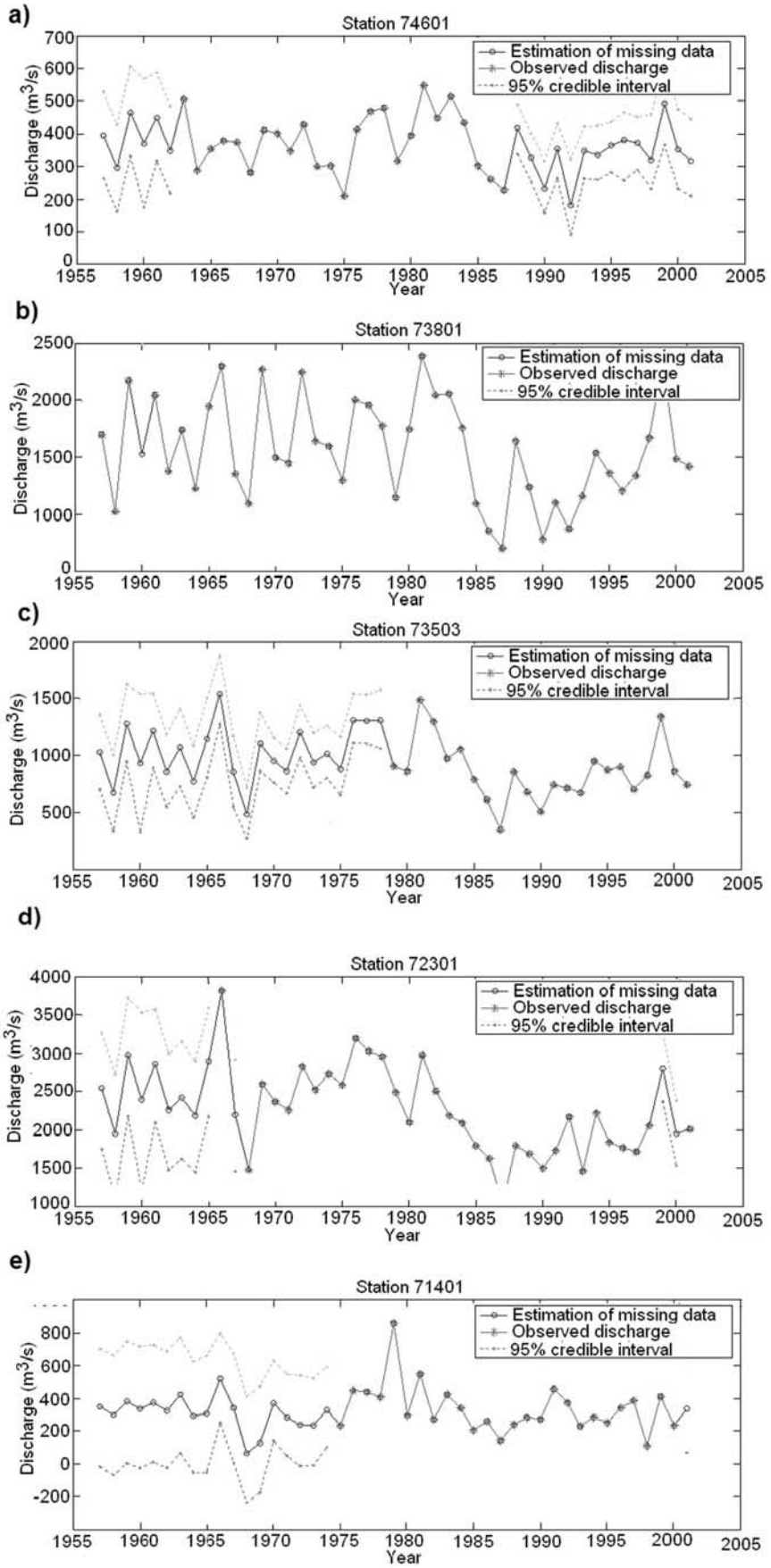

Figure 8. Estimations and credible intervals for missing data: (a) station 74601, (b) station 73801, (c) station 73503, (d) station 72301, and (e) station 71401.

can be used to obtain a practical solution for a wide variety of switching models.

6. Conclusions

[74] This paper provided an implementation of Bayesian analysis for multivariate regression via Gibbs sampling. The method was extended to the inclusion of missing values and to the inclusion of a change point structure in the model. An attractive feature of the approach presented in this paper is that it can be applied to cases that cannot be analyzed with recently published change point detection methodologies such as *Rasmussen* [2001] and *Perreault et al.* [2000a, 2000b]: It can readily be applied to cases where the change point simultaneously occurs in several response variables, to cases where the change does not occur with certainty and to cases where informative priors are appropriate. Three applications that highlight these features are presented.

[75] An interesting future development would be to relax the assumption of constant residual variance over time and the one of normality. A potential approach for this would be to introduce dependencies in the variance evolution over time, hence allowing for variable variance estimation. The scope of possibilities for the developed approach goes beyond the analysis of the single change point problem. Potential applications of this model include not only change point models, but also other switching models such as segmented multivariate regression or shifting-level models. Such generality is made possible by the fact that the design matrices $\{\mathbf{F}_t\}$ can be structured in accordance to such models. This opens the door for a practical approach to analyze these models and apply them in the field of water resources.

Appendix A: Design Matrix F When There Is a Continuity Constraint at the Change Point

[76] When there is a continuity constraint at the change point, the expression of the design matrix $\mathbf{F}$ is slightly different of that presented at section 3. We give here its expression for two practical cases.

A1. Continuity Constraint at the Change Point

$$\mathrm{Y}_t = \begin{cases} \mathrm{X}_t\beta_1' + \nu_t \ if \ t \le \tau \\ (\mathrm{X}_t - \mathrm{X}_\tau)\beta_2' + \mathrm{X}_\tau\beta_1' + \nu_t \ if \ t > \tau \end{cases}$$

$$\mathrm{X}_t^* = (\mathrm{X}_\mathrm{t}, \mathrm{X}_\tau)$$

$$\theta = (\beta_1\beta_2)$$

$$\Delta_t^\tau = \begin{pmatrix} (\delta_t)^\tau \mathrm{I}_m & (1-(\delta_t)^\tau)\mathrm{I}_m \\ (1-(\delta_t)^\tau)\mathrm{I}_m & -(1-(\delta_t)^\tau)\mathrm{I}_m \end{pmatrix}$$

$$\mathrm{F_t} = \mathrm{X_t}^* \Delta_\mathrm{t}^\tau.$$

[77] Note that in this special case, if $\mathbf{X}$ has a column with constant values, the coefficient of the first element of $\boldsymbol{\beta}_2$ is always null, thus this parameter should not be updated in the MCMC computations.

A2. Linear Relationship Before the Change, Constant Mean After the Change, and Continuity of the Mean Model at the Change Point

$$\mathrm{Y}_t = \begin{cases} \mathrm{X_t}\beta' + \nu_\mathrm{t} \ if \ t \le \tau \\ \mathrm{X}_\tau\beta' + \nu_\mathrm{t} \ if \ t > \tau \end{cases}$$

$$\mathrm{X_t^*} = (\mathrm{X_t}, \mathrm{X}_\tau)$$

$$\Delta_t^\tau = \begin{pmatrix} (\delta_t)^\tau \mathrm{I}_m^* \\ (1-(\delta_t)^\tau)\mathrm{I}_m^* \end{pmatrix}$$

$$\theta = \beta'$$

$$\mathrm{F_t} = \mathrm{X_t^*}\Delta_\mathrm{t}^\tau$$
$$\mathrm{Y_t} = \mathrm{F_t}\theta + \nu_\mathrm{t}.$$

Appendix B: Extension to Missing Data

[78] The case where missing values are present in $\mathbf{Y}^\mathrm{v}$ is examined. For any given matrix (or vector) $\mathbf{a}$, let $\mathbf{a}_{(u)}$ be the matrix (or vector) composed of the values in $\mathbf{a}$ corresponding to the set of indices U. Hence define $\mathbf{Y}^\mathrm{v}_{(M)}$ to be the vector of missing values in $\mathbf{Y}^\mathrm{v}$, where M is the set of indices corresponding to the missing values in $\mathbf{Y}^\mathrm{v}$, and define $\mathbf{Y}^\mathrm{v}_{(O)}$ to represent the vector of observed values in $\mathbf{Y}^\mathrm{v}$, where O is the set of indices corresponding to the observed values.

[79] From (15), the posterior distribution of the missing values in $\mathbf{Y}^\mathrm{v}$ is

$$\mathbf{Y}^\mathrm{v}_{(M)} | \mathbf{Y}^\mathrm{v}_{(O)}, \mathbf{F}, \Sigma_y \quad \mathrm{N}\left(\tilde{\mathbf{f}}_{(M)}, \tilde{\mathbf{Q}}_{(M\times M)}\right),$$

where

$$\tilde{\mathbf{f}}_{(M)} = \mathbf{f}_{(M)} + \mathbf{Q}_{(M\times O)}\left(\mathbf{Q}_{(O\times O)}\right)^{-1}\left(\mathbf{Y}^\mathrm{v}_{(O)} - \mathbf{f}_{(O)}\right)$$

and

$$\tilde{\mathbf{Q}}_{(M\times M)} = \mathbf{Q}_{(M\times M)} - \mathbf{Q}_{(M\times O)}\left(\mathbf{Q}_{(O\times O)}\right)^{-1}\mathbf{Q}_{(O\times M)}.$$

[80] If $\mathbf{F}$ has missing data, it can also be generated by Gibbs sampling. With this approach, the model for $\mathbf{F}$ cannot be ignored and prior distributional assumptions on $\mathbf{F}$ must be considered. For instance, in the case of model [13], the prior must account for the change point structure $\mathbf{F}_t = \mathbf{X}_t \boldsymbol{\Delta}_t^{(\tau)}$. We present a solution for this special case.

[81] Define $\mathbf{X}_t^\mathrm{v}$ by stacking the columns of $\mathbf{X}_t$ into a single column vector. Inference on missing data will be obtained from the conditional joint distribution of $\mathbf{Y}_t$ and $\mathbf{X}_t^\mathrm{v}$. These two components are related by model [13], which can also be written as

$$\mathbf{Y}_t = \left(\boldsymbol{\theta}_t^{(\tau)'} \otimes \mathbf{I}_r\right)\mathbf{X}_t^\mathrm{v} + \boldsymbol{\upsilon}_t. \tag{B1}$$

[82] Let $\{\mathbf{X}_1^{\mathrm{v}},\ldots,\mathbf{X}_n^{\mathrm{v}}\}$ be independent with

$$\mathbf{X}_t^{\mathrm{v}}|\boldsymbol{\mu}_{\mathbf{x}},\boldsymbol{\Omega}_{\mathbf{x}} \sim \mathrm{N}[\boldsymbol{\mu}_{\mathbf{x}},\boldsymbol{\Omega}_{\mathbf{x}}],\quad t=1,\ldots,n$$

and define for a given time t when some values are missing

$$\mathbf{Z}_t=\begin{pmatrix}\mathbf{Y}_t\\ \mathbf{X}_t^{\mathrm{v}}\end{pmatrix}$$

From (B1), it follows that

$$\mathbf{Z}_t|\tau,\boldsymbol{\theta},\boldsymbol{\Sigma}_{\mathbf{y}},\boldsymbol{\mu}_{\mathbf{x}},\boldsymbol{\Omega}_{\mathbf{x}} \sim \mathrm{N}[\mathbf{g}_t,\mathbf{R}_t],$$

where

$$\mathbf{g}_t=\begin{pmatrix}\left(\boldsymbol{\theta}_t^{(\tau)}\otimes\mathbf{I}_r\right)\boldsymbol{\mu}_{\mathbf{x}}\\ \mathbf{m}_{\mathbf{x}}\end{pmatrix}$$

$$\mathbf{R}_t=\begin{pmatrix}\boldsymbol{\Sigma}_y+\left(\boldsymbol{\theta}_t^{(\tau)}\otimes\mathbf{I}_r\right)\boldsymbol{\Omega}_x\left(\boldsymbol{\theta}_t^{(\tau)}\otimes\mathbf{I}_r\right)\left(\boldsymbol{\theta}_t^{(\tau)'}\otimes\mathbf{I}_r\right)\boldsymbol{\Omega}_{\mathbf{x}} & \\ \boldsymbol{\Omega}_{\mathbf{x}}\left(\boldsymbol{\theta}_t^{(\tau)}\otimes\mathbf{I}_r\right) & \boldsymbol{\Omega}_{\mathbf{x}}\end{pmatrix}.$$

[83] From model assumptions, $\{\mathbf{Z}_1,\ldots,\mathbf{Z}_n\}$ are independent. Hence the posterior distribution of the missing values in $\mathbf{Z}_t$ is directly obtained by normal theory, that is,

$$\mathbf{Z}_{t(M)}|\mathbf{Z}_{t(O)},\tau,\boldsymbol{\theta},\boldsymbol{\Sigma}_{\mathbf{y}},\boldsymbol{\mu}_{\mathbf{x}},\boldsymbol{\Omega}_{\mathbf{x}} \sim \mathrm{N}\left[\tilde{\mathbf{g}}_{t(M)},\tilde{\mathbf{R}}_{t(M\times M)}\right],$$

where

$$\tilde{\mathbf{g}}_{t(M)}=\mathbf{g}_{t(M)}+\mathbf{R}_{t(M\times O)}\left(\mathbf{R}_{t(O\times O)}\right)^{-1}\left(\mathbf{Z}_{t(O)}-\mathbf{g}_{t(O)}\right) \quad \text{(B2a)}$$

$$\tilde{\mathbf{R}}_{t(M\times M)}=\mathbf{R}_{t(M\times M)}-\mathbf{R}_{t(M\times O)}\left(\mathbf{R}_{t(O\times O)}\right)^{-1}\mathbf{R}_{t(O\times M)}. \quad \text{(B2b)}$$

[84] However, $(\mathbf{R}_t(O\times O))$ may not be strictly positive definite, but only nonnegative definite. This will happen if, for example, an intercept parameter is part of model (7). It can be shown that (B2) remains valid if the g inverse (generalized inverse) $(\mathbf{R}_{t(O\times O)})^{-}$ is used instead of $(\mathbf{R}_{t(O\times O)})^{-1}$. The g inverse of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{-}$ and can be calculated by $\mathbf{A}^{-}=\Gamma\Lambda^{-1}\Gamma'$, where Γ is a column orthonormal matrix of eigenvectors corresponding to the s nonzero eigenvalues $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_s)$ of $\mathbf{A}$. A more general definition of the g inverse is reviewed by *Mardia et al.* [1979].

[85] To provide an estimation tool for the parameters $\boldsymbol{\mu}_{\mathbf{x}}$ and $\boldsymbol{\Omega}_{\mathbf{x}}$, we consider the conjugate normal inverse Wishart prior

$$\boldsymbol{\Omega}_{\mathbf{x}} \sim \mathrm{W}_{\upsilon_0}^{-1}(\mathbf{V}_0) \quad \text{(B3a)}$$

$$\boldsymbol{\mu}_{\mathbf{x}}|\boldsymbol{\Omega}_{\mathbf{x}} \sim \mathrm{N}[\boldsymbol{\mu}_{\mathbf{0}},\boldsymbol{\Omega}_{\mathbf{x}}/k_0]. \quad \text{(B3b)}$$

[86] From *Gelman et al.* [1995], the conditional posterior is

$$\Omega_x|\{\mathbf{X}_t\} \sim \mathrm{W}_{\upsilon_n}^{-1}(\mathbf{V}_n)$$
$$\mu_x|\{\mathbf{X}_t\},\Omega_x \sim \mathrm{N}[\mu_n,\Omega_x/k_n],$$

where

$$\mu_n=\frac{1}{k_n}\left(k_0\mu_0+n\overline{\mathbf{X}}\right)$$
$$k_n=k_0+n$$
$$\upsilon_n=\upsilon_0+n$$
$$\mathbf{V}_n=\mathbf{V}_0+\mathbf{S}+\frac{k_0 n}{k_n}\left(\overline{\mathbf{X}}-\mu_0\right)\left(\overline{\mathbf{X}}-\mu_0\right)'$$

and

$$\overline{\mathbf{X}}=\frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t^{\mathrm{v}}$$
$$\mathbf{S}=\sum_{t=1}^{n}\mathbf{X}_t^{\mathrm{v}}\mathbf{X}_t^{\mathrm{v}}-n\overline{\mathbf{X}}\overline{\mathbf{X}}'.$$

[87] We can easily obtain samples from this joint posterior by first sampling $\boldsymbol{\Omega}_{\mathbf{x}}|\{\mathbf{x}_t\}$ and then sampling $\boldsymbol{\mu}_{\mathbf{x}}|\{\mathbf{x}_t\}$, $\boldsymbol{\Omega}_{\mathbf{x}}$. The noninformative multivariate Jeffrey's prior density for $\{\boldsymbol{\mu}_{\mathbf{x}},\boldsymbol{\Omega}_{\mathbf{x}}\}$ is

$$[\boldsymbol{\mu}_{\mathbf{x}},\boldsymbol{\Omega}_{\mathbf{x}}] \propto |\boldsymbol{\Omega}_{\mathbf{x}}|^{-\left(rm^{*}+1\right)/2}.$$

[88] This is the limiting case of the normal inverse Wishart prior in (B3) $k_0 \to 0$, $\upsilon_0 \to -1$ and $|\mathbf{V}_0| \to 0$. The posterior distribution $\{\boldsymbol{\mu}_{\mathbf{x}},\boldsymbol{\Omega}_{\mathbf{x}} \mid \{\mathbf{X}_t\}\}$ for this case can be written as

$$\boldsymbol{\Omega}_{\mathbf{x}}|\{\mathbf{X}_{\mathbf{t}}\} \sim \mathrm{W}_{n-1}^{-1}(\mathbf{S})$$
$$\boldsymbol{\mu}_{\mathbf{x}}|\{\mathbf{X}_{\mathbf{t}}\},\boldsymbol{\Omega}_{\mathbf{x}} \sim \mathrm{N}\left[\overline{\mathbf{X}},\Omega_x/n\right],$$

in which we find the sample estimators of the parameters $\boldsymbol{\mu}_{\mathbf{x}}$ and $\boldsymbol{\Omega}_{\mathbf{x}}$.

[89] **Acknowledgments.** The financial support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Northern Study Center (CEN) of Laval University, and the Ouranos consortium is gratefully acknowledged. The authors wish to thank the four reviewers as well as the associate editor for their insightful comments that greatly helped to improve the quality of the paper.

References

Asselin, J. J., T. B. M. J. Ouarda, and O. Seidou (2005), Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables, model development research report, Eau, Terre and Environ., Inst. Natl. de la Rech. Sci., Quebec, Que., Canada.

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 (2007), Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations (in French), *J. Sci. Hydrol.*, *51*, 18–37.

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

Bowman, A. W., A. Pope, and B. Ismail (2006), Detecting discontinuities in nonparametric regression curves and surfaces, *Stat. Comput.*, *16*, 377–390.

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

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.

Chernoff, H., and S. Zacks (1963), Estimating the current mean of a normal distribution which is subjected to changes in time, *Ann. Math. Stat.*, *35*, 999–1028.

Deng, J. L. (1989), Introduction to grey system theory, *J. Grey Syst.*, *1*(1), 1–24.

Easterling, D. R., and T. C. Peterson (1995), A new method for detecting undocumented discontinuities in climatological time series, *Int. J. Climatol.*, *15*, 369–377.

Fearnhead, P. (2005), 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.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, London.

Geman, S., and D. Geman (1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell.*, *6*, 721–741.

Green, P. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, *82*, 711–732.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*, 97–109.

Lee, A. S. F., and S. M. Heghinian (1977), A shift of the mean level in a sequence of independent normal random variables—A Bayesian approach, *Technometrics*, *19*, 503–506.

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

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979), *Multivariate Analysis*, Academic, New York.

Metropolis, N., and S. Ulam (1949), The Monte Carlo method, *J. Am. Stat. Assoc.*, *44*, 335–341.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092.

Moore, R. E. (1979), *Method and Application of Interval Analysis*, Soc. for Ind. and Appl. Math., Philadelphia, Pa.

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

Perreault, L., É. Parent, J. Bernier, and B. Bobée (2000b), 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.

Peterson, T. C., et al. (1998), Homogeneity adjustments of in situ atmospheric climate data: A review, *Int. J. Climatol.*, *18*, 1493–1517.

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., and T. B. M. J. Ouarda (2006), Recursion-based multiple changepoint detection in multivariate linear regression and application to river streamflows, *Water Resour. Res.*, *43*, W07404, doi:10.1029/2006WR005021.

Smith, A. F. M. (1975), A Bayesian approach to inference about change-point in sequence of random variables, *Biometrika*, *62*, 407–416.

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

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

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].

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.

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

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.

Zhao, X., and P.-S. Chu (2006), Bayesian multiple change point analysis of hurricane activity in the eastern North Pacific: A Markov chain Monte Carlo approach, *J. Clim.*, *19*, 564–578.

J. J. Asselin and T. B. M. J. Ouarda, INRS-ETE, 490 rue de la Couronne, Quebec, QC, Canada G1K 9A9.

O. Seidou, Department of Civil Engineering, University of Ottawa, 161 Louis Pasteur Office A113, Ottawa, ON, Canada K1N 6N5. (oseidou@uottawa.ca)