Journal of Hydrology 235 (2000) 221–241

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited

L. Perreault[a,b,*], J. Bernier[a,b], B. Bobée[b], E. Parent[a]

[a]*ENGREF, laboratoire de Gestion du Risque en Sciences de l'eau (GRESE), 19 Avenue du Maine, 75732 Paris Cedex 15, France*
[b]*Chaire en Hydrologie Statistique Hydro-Québec/CRSNG, INRS-Eau, 2800 rue Einstein, Sainte-Foy, Québec, Canada, G1V 4C7*

Received 22 June 1999; revised 2 March 2000; accepted 25 May 2000

Abstract

A Bayesian method is presented for the analysis of two types of sudden change at an unknown time-point in a sequence of energy inflows modeled by independent normal random variables. First, the case of a single shift in the mean level is revisited to show how such a problem can be straightforwardly addressed through the Bayesian framework. Second, a change in variability is investigated. In hydrology, to our knowledge, this problem has not been studied from a Bayesian perspective. Even if this model is quite simple, no analytic solutions for parameter inference are available, and recourse to approximations is needed. It is shown that the Gibbs sampler is particularly suitable for change-point analysis, and this Markovian updating scheme is used. Finally, a case study involving annual energy inflows of two large hydropower systems managed by Hydro-Québec is presented in which informative prior distributions are specified from regional information. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Change-point analysis; Energy inflows; Gibbs sampling; Change in the variance; Change in the mean level

1. Introduction

This paper provides a Bayesian approach to characterize when and by how much a single change has occurred in a sequence of hydrometeorological random variables. Inferences herein are based on the analysis of posterior distributions and are conditional upon the fact that a change happened with certainty. The problem of testing its existence, and of identifying its type, is investigated in Perreault et al. (2000a).

Although such a hypothesis is rarely stated explicitly, the assumption that stochastic time series remain stationary plays a crucial role in water resources management. Under the assumption that tomorrow will statistically behave like yesterday, stochastic models are fitted to hydrometeorological variables such as river flow, precipitation and temperature. The estimated models are then used for many engineering purposes, in particular for simulating the operation of hydropower systems (energy planning, design of power plants, operation of reservoirs). Consequently such models and the decisions stemming from them are based on the hypothetical stationary behavior of hydrometeorological inputs. However, some hydrometeorological time series can exhibit abrupt changes maybe caused by site-specific factors (e.g. land-use effect on water yield) or induced by a climatic change. This leads to questioning the stationarity hypothesis in hydrometeorological time series analysis.

Hydro-Québec is a public company that produces,

* Corresponding author. Fax: +1-418-654-2600.
E-mail address: perrealu@inrs-eau.uquebec.ca (L. Perreault).

PII: S0022-1694(00)00270-5

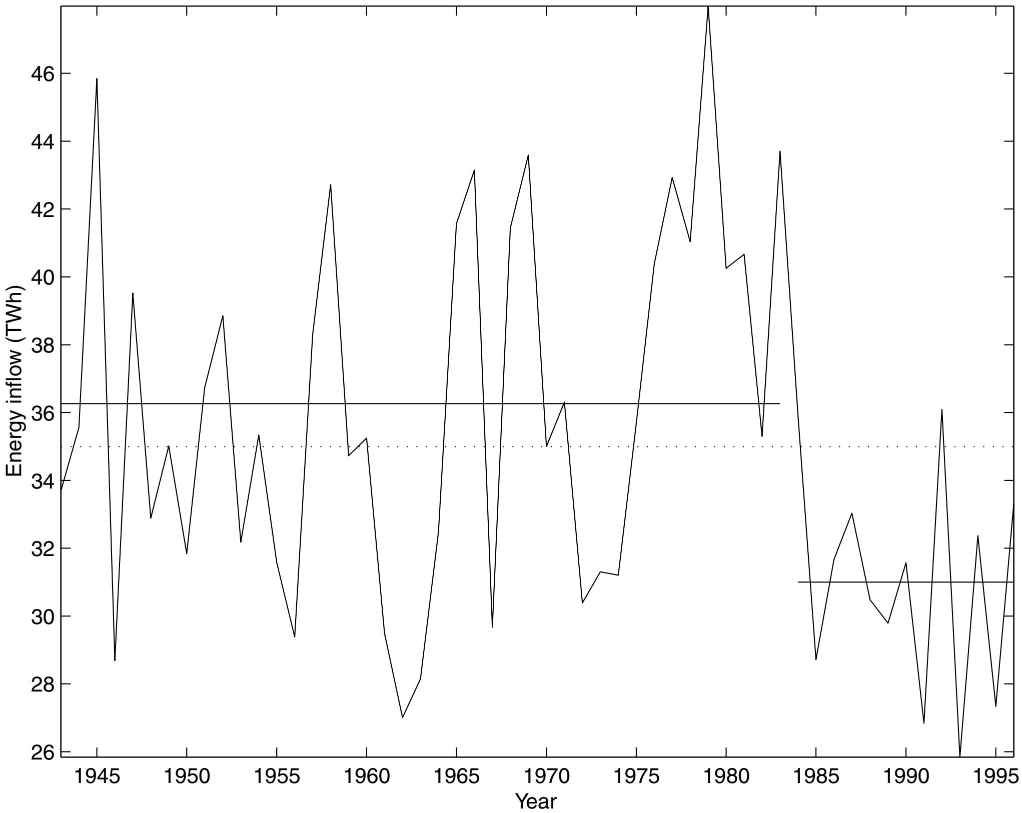

Fig. 1. Annual energy inflow for Churchill Falls power system.

transmits and distributes electricity throughout the province of Québec. It currently operates 54 power plants supplied by 26 large reservoirs. The sites are assembled to form 8 major hydropower systems: St-Laurent, Outaouais, La Grande, St-Maurice, Bersimis, Manicouagan, Outardes and Churchill Falls. For each of these systems, annual energy inflows are evaluated by multiplying the net basin supply of each reservoir into the system by a factor based on the production capacity of the corresponding power plant. Such annual energy inflows are therefore subject to hydro-meteorological changes, if any. These data are of high importance for energy planning because some of their statistical characteristics (namely mean and variance) are used as inputs to construct scenarios or to forecast future energy availability.

Consider the time plot in Fig. 1 which shows the annual energy inflow in terawatt-hour (TWh) for the hydropower system Churchill Falls. This 92 432 km^2 watershed is situated in northeastern Québec, in the Labrador (province of Newfoundland).

Examining this time series, one may suspect that an abrupt change in the mean level has occurred around 1983, and two distinct partial means could be evaluated: $\bar{x}_1$ from 1943 to 1983 and $\bar{x}_2$ from 1984 to 1996 (continuous lines). On the other hand, it can be argued that this sudden fluctuation may only be due to the natural variability of the hydrologic regime, and one would rather still consider the overall mean $\bar{x}$ as the representative available annual amount of energy (dotted line). Let's now suppose that based on these historical observations, the construction of a power plant has to be planned for hydropower generation. To meet the energy demand and eventually export generated hydroelectricity, the hydropower company would like to rely on one and only one of the possible situations $(\bar{x}_1, \bar{x}_2)$ or $\bar{x}$. Since for this case $(\bar{x}_2 - \bar{x}_1)/\bar{x}$ may be as high as 15%, decisions for the future

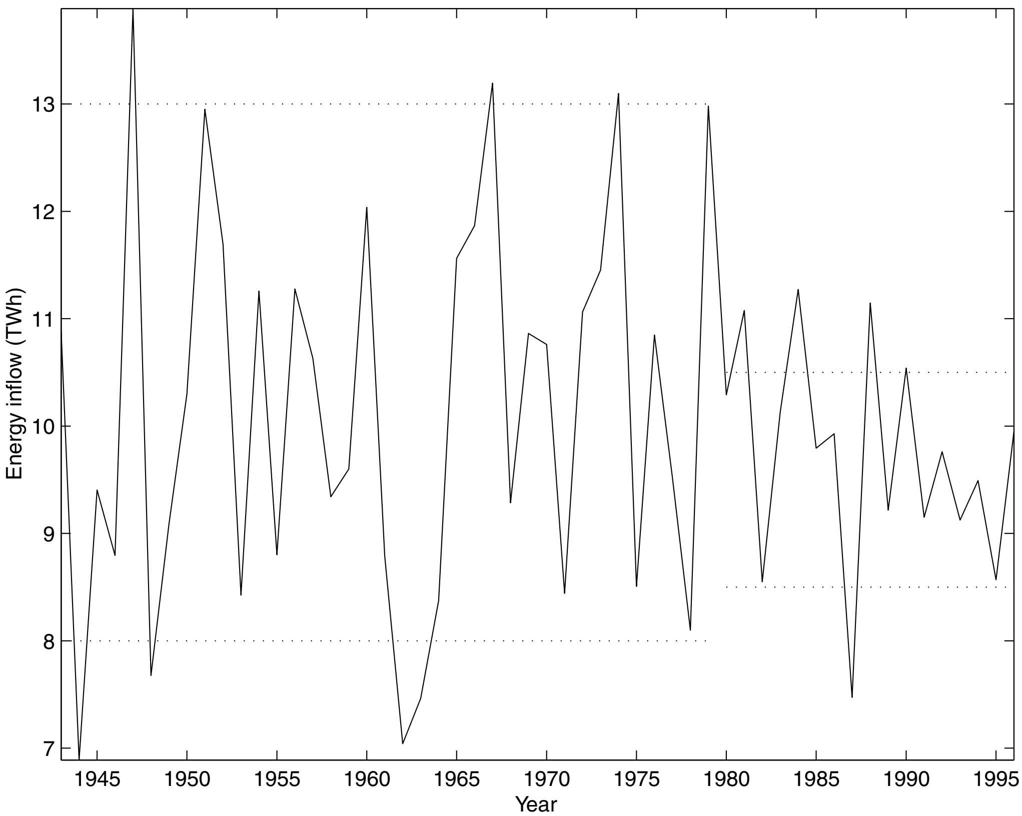

Fig. 2. Annual energy inflow for Outaouais power system.

relying on $\bar{x}$ can be dramatically different from anticipations based on $(\bar{x}_1, \bar{x}_2)$. Analyzing a change in the mean level energy inflow data is then clearly an essential step before planning hydropower systems. Other hydrological contexts such as rainfall estimation, flood analysis or climatological studies of global warming are concerned with abrupt shifts in the mean level. References can be found in Perreault et al. (1999).

As an example for a possible change of a different kind, consider the annual series of energy inflow for the Outaouais power system located in the southwestern Québec (Fig. 2).

The overall variability of these observations seems to have suddenly decreased after 1980. The energy planning system of forecasts depends strongly upon energy inflow variability. For instance, a sharp knowledge of this characteristic is the key point to determine which range of scenarios is needed to make a good decision for reservoir releases and interannual storage strategy. After examining this series, an engineer will typically base his decision on one of the two possible estimates: evaluating the variance over the entire period or considering only the last observations, i.e. using more but maybe not relevant information about future realizations or taking into account less but maybe more representative information. As in the case of a shift in the mean, a change in variance may induce important decisions and have large economic consequences (it is well known for instance that the size of reservoirs will be directly related to inflow variance); inference about existence and characteristics of these changes can thus also be considered as a valuable precaution before developing management rules in water resources systems.

In this paper, interest is mainly focussed on the estimation of the unknown time-point and intensity of the change, assuming an abrupt change has in fact occurred. The aim is to outline that, in the case of a single sudden change in the mean level or variance, these problems can be straightforwardly addressed through the Bayesian framework, which in turn can be easily solved by implementation of explicit or Monte Carlo based inference techniques. We emphasize that the purpose of this paper is not to determine why such sudden changes occurred (e.g. climatic change or site-specific factors). Of course, attribution is an important and interesting research topic, but we are only concerned with developing statistical tools to infer about the point and intensity of a change to help decision making.

In Section 2, we formulate the general Bayesian change-point setting and consider two specific models: a single change in the mean and a single change in the variance of a sequence of independent normal random variables. Change in the mean level leads to solutions in closed form and results are therefore directly available. However, such is not the case for a change in the variance. The Gibbs sampler, as presented in Section 3, provides an elegant and convenient answer to this problem. In Section 4, the practitioner's point of view is taken when applying the approach to the series of Churchill Falls and Outaouais energy inflows presented above. A simple regional analysis was performed to specify the parameters of the prior distributions which are assumed to represent the prior state of belief. Finally, Section 5 offers a general discussion and conclusions.

2. Bayesian analysis of normal sequences with an unknown change-point

The simplest formulation of the change-point problem is the following. Let us assume that the densities $p_1(x)$ and $p_2(x)$ belong to a known parametric class of probability densities $\mathbf{P} = \{p(x|\theta); \theta \in \Theta\}$ indexed by an unknown parameter θ such that, for a sequence of n independent random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, we have

$$\begin{aligned} X_i &\sim p_1(x) = p(x_i|\theta_1), \qquad i = 1, \ldots, \tau \\ X_i &\sim p_2(x) = p(x_i|\theta_2), \qquad i = \tau + 1, \ldots, n \end{aligned} \tag{1}$$

where $\theta_1 \neq \theta_2$ and $\tau = 1, 2, \ldots, n-1$ is an unknown parameter, called the change-point. That is, the first and second parts of the sequence of random variables are distributed as statistical distributions which belong to the same class, but with different unknown parameter θ. Since the unknown change-point τ can take values between 1 and $n - 1$, this model assumes that exactly one abrupt change occurred with certainty. Indeed other types of changes such as trends can exist, and model (1) would not be appropriate. However, recent understanding of global climate interactions such as the El Nino/La Nina and the North Atlantic Oscillation phenomena give credence to the idea that climate may operate in two or more quasi-stationary states, and that it can rapidly switch from one state to another (Rodriguez-Iturbe et al., 1991; Kerr, 1992, 1999). Therefore, an abrupt change-point model such as expression (1) may be representative of several hydrological and climatic time series. For instance, there is evidence of such behavior in climatic and hydrological data series in the Africa (Paturel et al., 1997; Servat et al., 1997) and in the Pacific region (Kerr, 1992). As regards Hydro-Québec energy inflow series, many annual runoff series for rivers situated in Northern Québec seem to exhibit this type of sudden drastic change (Perreault et al., 2000b). Thus, model (1) must be considered, at least as a first methodological step. Anyhow, we think that there is a need to consider first this simple model in detail since application of Bayesian change-point analysis in hydrological literature is very limited.

The likelihood function resulting from n observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ generated by model (1) becomes

$$p(\mathbf{x}|\theta_1, \theta_2, \tau) = \prod_{i=1}^{\tau} p(x_i|\theta_1) \prod_{i=\tau+1}^{n} p(x_i|\theta_2) \tag{2}$$

from which, for instance, maximum likelihood estimates for τ, θ_1 and θ_2 can be obtained (Hinkley, 1970). In the Bayesian perspective, a joint prior distribution $p(\theta_1, \theta_2, \tau)$ is assumed for the parameters. Bayes theorem then provides the joint posterior distribution $p(\theta_1, \theta_2, \tau|\mathbf{x})$ of θ_1, θ_2, τ given the data by normalization over $\mathbf{x}$ of the joint density written as

$$p(\mathbf{x}|\theta_1, \theta_2, \tau)p(\theta_1, \theta_2, \tau) \tag{3}$$

Just as the prior distribution $p(\theta_1, \theta_2, \tau)$ reflects beliefs about the parameters prior to experimentation, the posterior distribution $p(\theta_1, \theta_2, \tau|\mathbf{x})$ reflects the updated beliefs after observing the sample data. In the Bayesian framework, all statistical inference about the unknown parameters is based on the posterior distribution (Berger, 1985). The analysis of posterior distributions is often referred to as "scientific reporting".

Interest is now on making inference about the change-point τ, the structure parameters θ_1 and θ_2, and any suitable function $W(\theta_1,\ \theta_2)$ which can describe the amount of shift. To achieve this, all marginal posterior distributions must be found by integration. For example, to evaluate $p(\tau|\mathbf{x})$, θ_1 and θ_2 must be integrated out of $p(\theta_1,\ \theta_2,\ \tau|\mathbf{x})$. Using conjugate prior distributions for fixed τ, and assuming prior independence between τ and the other parameters $(\theta_1,\ \theta_2)$, i.e. $p(\theta_1, \theta_2, \tau) = p(\theta_1, \theta_2)p(\tau)$, solutions in closed form can be obtained for some simple models (see Section 2.1). Likelihoods for which conjugate prior density functions exist are those corresponding to exponential family models (Bernardo and Smith, 1994). In more complex models (see Section 2.2), even with simple prior distributions and independence, integration may turn to a very difficult numerical task. However, use of Gibbs sampler, a tool particularly suitable for change-point analysis, enables a straightforward solution to such problems as developed in Section 3.

2.1. A single change in the mean (model M_1)

As in most published approaches on change-point studies, we assume univariate sequences of independent normal random variables. This simple model is appropriate for Hydro-Québec large hydropower system annual energy inflows. One can invoke the central limit theorem to justify the normal assumption, since annual energy inflows for a given hydropower system are calculated as a summation over time and space of the monthly energy inflows. In addition, changes in the parameters of the normal model have a direct and simple hydrological interpretation. If needed, one may transform the data to achieve normality. However, in this case of reparametrization, interpretation of a change in the new parameters should be done very carefully. From the hydrological point of view, the independence assumption between successive values may be more questionable. It is made here for convenience as a first methodological step, and can be relaxed eventually by considering, for example, a shift in the mean-vector of an autoregressive model. Hydro-Québec annual energy inflows are proportional to the net basin supplies, which in turn are evaluated following the water balance equation. Therefore, the data considered herein are implicitly adjusted for surface and subsurface storage effect that could induce interannual correlations, and the assumption of independence seems reasonable to us. Still, it can be verified on similar annual series for which it is believed that no change has occurred.

Consider a set of random variables observed at consecutive equally spaced time points. Suppose that, due to some exogenous factors, the first and second parts of the sequence of random variables operate at two different mean levels, respectively, μ_1 and μ_2, but with the same variance σ^2. This situation can be represented by the following model denoted by M_1:

$$\begin{aligned} X_i &\sim \mathcal{N}(x_i|\mu_1, \sigma^2), \qquad i = 1, \ldots, \tau \\ X_i &\sim \mathcal{N}(x_i|\mu_2, \sigma^2), \qquad i = \tau + 1, \ldots, n \end{aligned} \tag{4}$$

where $\mathcal{N}(x_i|\mu, \sigma^2)$ stands for the usual normal probability density function (p.d.f.) with parameters $\mu \in \Re$ and $\sigma \in \Re^+$.

The Bayesian framework for inference regarding model M_1 dates back to Chernoff and Zacks (1963). Main theoretical contributions can be found in Smith (1975), Lee and Heghinian (1977), Booth and Smith (1982) and Broemeling (1985). Their approaches differ mainly by the prior distributions specified to represent the unknown parameters. The method proposed by Lee and Heghinian (1977) has been used in several practical cases in hydrology, for instance by Bruneau and Rassam (1983), and more recently by Paturel et al. (1997) and Servat et al. (1997). Also, Bernier (1994) and Perreault et al. (1999) considered a Bayesian approach for model M_1 to analyze changes in hydrometeorological time series.

To illustrate the principle of Bayesian change-point analysis, the posterior distributions of the parameters of interest for model M_1 are derived in Appendix A.

Conjugate prior distributions are considered and independence between the date of change τ and the model parameters is assumed. With these assumptions, the joint posterior distribution is a finite mixture of conjugate distributions, and analytical expressions are obtained for all marginals and conditionals. The assumption of independence signifies that what is known about date of change in the mean level of the annual energy inflows does not depend upon the means before and after the change. Note that this hypothesis could be questionable if for instance one believes that a change can only occur if the mean level reaches a given threshold. These results are well known in the statistical literature, but change-point inference analysis does not belong to the standard statistical toolbox of most practitioners in hydrology.

2.2. A single change in the variance (model M_2)

In this section we consider the change of variance case represented by model M_2:

$$\begin{aligned} X_i &\sim \mathcal{N}(x_i|\mu, \sigma_1^2), \qquad i = 1,\ldots,\tau \\ X_i &\sim \mathcal{N}(x_i|\mu, \sigma_2^2), \qquad i = \tau+1,\ldots,n \end{aligned} \tag{5}$$

The problem of a change in the variance occurring at an unknown time-point has been less widely covered than a single shift in the mean level. In hydrology, to our knowledge, it has not been studied from a Bayesian perspective although the variance of hydrological series is an important parameter in water resources management. In the statistical literature, a change in the variance of a sequence of normal random variables was first treated within the Bayesian framework by Smith (1975) who assumed noninformative prior distributions. In this particular case, closed form for the posterior densities can be obtained. Menzefricke (1981) generalized Smith's approach by considering informative priors.

The likelihood function resulting from n observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of model M_2, a series of annual energy inflows such as shown in Fig. 2, can be expressed as

$$\begin{aligned} p(\mathbf{x}|\mu, \boldsymbol{\sigma}^2, \tau, M_2) &= \prod_{i=1}^{\tau} \mathcal{N}(x_i|\mu, \sigma_1^2) \prod_{i=\tau+1}^{n} \mathcal{N}(x_i|\mu, \sigma_2^2) \\ &= \left(\frac{1}{2\pi}\right)^{(n/2)} \left(\frac{1}{\sigma_1^2}\right)^{(\tau/2)} \left(\frac{1}{\sigma_2^2}\right)^{((n-\tau)/2)} \\ &\quad \times \exp\left\{-\frac{\tau}{2\sigma_1^2}[s_\tau^2 + (\bar{x}_\tau - \mu)^2]\right\} \\ &\quad \times \exp\left\{-\frac{(n-\tau)}{2\sigma_2^2}[s_{n-\tau}^2 + (\bar{x}_{n-\tau} - \mu)^2]\right\} \end{aligned} \tag{6}$$

We assumed the same joint prior distribution as Menzefricke (1981) to represent prior knowledge about the parameters of model M_2:

$$\begin{aligned} p(\mu, \boldsymbol{\sigma}^2, \tau|M_2) &= \mathcal{N}(\mu|\phi, \lambda\sigma_1^2)\mathcal{IG}(\sigma_1^2|\alpha_1, \beta_1)\mathcal{IG}(\sigma_2^2|\alpha_2, \beta_2)p(\tau|M_2) \\ &= \mathcal{NIGIG}(\mu, \boldsymbol{\sigma}^2|\phi, \lambda, \boldsymbol{\alpha}, \boldsymbol{\beta})p(\tau|M_2) \end{aligned} \tag{7}$$

As usual, the prior distribution of μ is normal with precision depending on σ_1, and priors for the variance are inverted gamma densities. However, strictly speaking, the prior distribution (7) although closely related is not a complete conjugate even for fixed τ. In fact, given τ, expression (7) assumes prior independence between σ_1^2 and σ_2^2, while no algebraic manipulations on the likelihood (6) allows us to separate these parameters in two independent expressions. Therefore, the exact conjugate should consider a fixed dependence structure between σ_1^2 and σ_2^2. This would lead to a nonstandard joint prior density for which it is difficult to specify the hyperparameters: eliciting knowledge about dependence between variances may be at least a quite complicated task if not unrealistic. We therefore adopt Menzefricke's prior assumptions letting the data account for the strength of dependence, if any.

Selection of particular values in expression (7) for the hyperparameters ϕ, λ, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ belongs to modeler's tasks (Section 4.1). These can be chosen to give various general shapes for the joint prior distribution and take into account a variety of prior beliefs about the studied phenomenon. The hyperparameter values may come from historical or regional information, even from subjective knowledge. When prior information about the phenomenon is limited, it may be desirable to let the prior knowledge for the unknown parameters be vague (Box and Tiao, 1973). The prior distribution can be turned to a particular form of noninformative density by letting $p(\mu, \boldsymbol{\sigma}^2, \tau | M_2) \propto p(\tau|M_2)\sigma_1^{-2}\sigma_2^{-2}$, i.e. $\lambda \to \infty$, $\alpha_1 = \alpha_1 = \beta_1 = \beta_2 \to 0$ in expression (7). This corresponds to Smith's noninformative joint prior distribution for M_2.

Using Bayes theorem to combine expressions (6) and (7), the joint posterior distribution of $(\mu, \boldsymbol{\sigma}^2, \tau)$, given the observed data $\mathbf{x}$, is proportional to

$$p(\mu, \boldsymbol{\sigma}^2, \tau | \mathbf{x}, M_2) \propto \times \exp\left\{ -\frac{\lambda'(n-\tau)\sigma_1^2 + \sigma_2^2}{2\lambda'\sigma_1^2\sigma_2^2}(\mu - \phi'')^2 \right\} \times \left(\frac{1}{\sigma_1^2}\right)^{\alpha_1'+1}\left(\frac{1}{\sigma_2^2}\right)^{\alpha_2'+1} \times \exp\left\{ -\left[\frac{\beta_1'}{\sigma_1^2} + \frac{\beta_2'}{\sigma_2^2} + \frac{(n-\tau)(\phi' - \bar{x}_{n-\tau})^2}{2[\lambda'(n-\tau)\sigma_1^2 + \sigma_2^2} \right]\right\} p(\tau|M_2) \quad (8)$$

where

$$\lambda' = \lambda/(1+\tau\lambda), \ \phi' = (1-\lambda'\tau)\phi + \lambda'\tau\bar{x}_\tau,$$

$$\phi'' = \frac{1}{\lambda'(n-\tau)\sigma_1^2 + \sigma_2^2}[\sigma_2^2\phi' + \lambda'(n-\tau)\sigma_1^2\bar{x}_{n-\tau}]$$

$$\beta_1' = \frac{\tau}{2}[s_\tau^2 + (1-\lambda'\tau)(\phi - \bar{x}_\tau)^2] + \beta_1,$$

$$\beta_2' = \frac{n-\tau}{2}s_{n-\tau}^2 + \beta_2,$$

$$\alpha_1' = \alpha_1 + \frac{\tau+1}{2}, \ \alpha_2' = \alpha_2 + \frac{n-\tau}{2}$$

It is possible to obtain an expression for the marginal posterior density of the intensity of change, defined as $\eta = \sigma_2^2/\sigma_1^2$, which is a parameter of great interest for describing a change in variance. First, transforming to $\theta = \sigma_1^2$ and $\eta = \sigma_2^2/\sigma_1^2$, and integrating μ and θ out from this joint posterior distribution (8), the joint posterior distribution of η and τ is obtained. Then, summing over τ, the marginal posterior distribution $p(\eta|\mathbf{x}, M_2)$ is seen to be proportional to

$$p(\eta|\mathbf{x}, M_2) \propto \sum_{\tau=1}^{n-1}\left[\frac{\lambda'}{\lambda'(n-\tau)+\eta}\right]^{1/2} \times \left(\frac{1}{\eta}\right)^{\alpha_2'+1/2} \frac{\Gamma(\alpha_1' + \alpha_2' + 1/2)}{[g(\tau,\eta)]^{\alpha_1'+\alpha_2'+1/2}} p(\tau|M_2), \quad (9)$$

where

$$g(\tau,\eta) = \left\{ \beta_1' + \beta_2'\eta^{-1} + \frac{(n-\tau)(\phi' - \bar{x}_{n-\tau})^2}{2[\lambda'(n-\tau)+\eta]} \right\}$$

Writing expression (8) differently, it is also possible to find a closed form for the marginal posterior distribution of μ by integrating σ_1^2 and σ_2^2 out of this expression and summing over τ. This expression is not given here since μ is not the parameter of major interest for model M_2. However, unlike the noninformative case or the problem of a single change in the mean level, recourse to numerical integration is necessary to make inferences about the other parameters (σ_1^2, σ_2^2 and τ). How can we avoid performing integration to get the normalizing constants in expressions (8) and (9)? Fortunately, the Gibbs sampler is an alternative handy tool, easy to implement for the hydrologist without numerical analysis skills, as will be showed in the next section.

3. Implementation of the Gibbs sampler for model M_2

The Gibbs sampler is one of a class of Markov Chain Monte Carlo (MCMC) algorithms that facilitate practical statistical problem solving (Tanner, 1992). In a Bayesian perspective, the objective is to produce posterior densities for parameters of interest. For simple models analytic calculation is possible, and results are directly available. For more complex models, recourse to numerical integration is necessary. However, in some models conventional numerical techniques are often insufficiently accurate or at least difficult to implement without being gifted in computer programming and numerical analysis. In contrast, the Gibbs sampler is generally straightforward to implement and is easily accessible to the average statistical practitioner. The Gibbs sampler was developed formally by Geman and Geman (1984) in the context of image restoration. Gelfand and Smith (1990) showed its applicability to general Bayesian computations while Carlin et al. (1992) and Stephens (1994) applied it to change-point problems. In hydrology, Lu and Berliner (1999) used the Gibbs algorithm, while other MCMC approaches were adopted by Kuczera and Parent (1998) and Campbell et al. (1999).

The idea of MCMC methods is to construct a Markov chain whose stationary and ergodic distribution is precisely the posterior distribution of interest that is intractable analytically. In the context of the general change-point model (1), the Gibbs sampler is an updating scheme that involves drawing t random values from each of the set of "full posterior distributions" $p(\tau|\theta_1, \theta_2, \mathbf{x})$, $p(\theta_1|\theta_2, \tau, \mathbf{x})$ and $p(\theta_2|\theta_1, \tau, \mathbf{x})$. A brief description in a general context of how these values should be sampled is offered in Lu and Berliner (1999). More details on theoretical properties of the algorithm can be found in Gelfand and Smith (1990). Subject to certain mild regularity conditions on the joint and conditional densities, it can be proven that the obtained sampled vector $\{(\tau^{(i)}, \theta_1^{(i)}, \theta_2^{(i)});\ i = 1, \ldots, t\}$ tends in distribution to a sample from the target density $p(\tau, \theta_1, \theta_2|\mathbf{x})$ as t tends to infinity. Hence, replicating the entire process in parallel m times and keeping the values obtained in the last iteration t provides independent and identically distributed triplets $(\tau_j^{(t)}, \theta_{1j}^{(t)}, \theta_{2j}^{(t)})$, $j = 1, \ldots, m$. These values can then easily be used for estimating different features of the posterior joint distribution, namely the marginal densities. For example, as an estimate for $p(\theta_1|\mathbf{x})$, one can draw a histogram with $(\theta_{11}^{(t)}, \ldots, \theta_{1m}^{(t)})$ or use a more general kernel-type of estimate. But since for any marginal the corresponding full conditional has been assumed available, more efficient estimation is obtained by using this set of distributions. In this paper, we use the so-called "Rao–Blackwellized" estimate (Gelfand and Smith, 1990); for density $p(\theta_1|\mathbf{x})$, the estimate is given by

$$\hat{p}(\theta_1|\mathbf{x}) = \frac{1}{m} \sum_{j=1}^{m} p(\theta_1|\tau_j^{(t)}, \theta_{2j}^{(t)}, \mathbf{x}) \tag{10}$$

The formal argument for the higher efficiency of expression (10) is based on the Rao–Blackwell theorem (Lehmann, 1983, p. 50). There may also be interest in a function of the random variables $W(\tau, \theta_1, \theta_2)$. Because each triplet $(\tau^{(t)}, \theta_1^{(t)}, \theta_2^{(t)})$ provides an observed $W^{(t)} = W(\tau^{(t)}, \theta_1^{(t)}, \theta_2^{(t)})$ whose marginal distribution is approximately $p(W)$, an estimate analogous to expression (10) can be obtained. In fact, if θ_1 actually appears as an argument of W, the full conditional density $p(W|\theta_2, \tau, \mathbf{x})$ can be deduced by univariate transformation of variable from $p(\theta_1|\theta_2, \tau, \mathbf{x})$, and used in expression (10). Note that complete implementation of the Gibbs sampler requires the determination of t and, across iterations, the choice of m. These values may vary between individual problems and are thus subject of much current research (a Web site which includes a lot of information and references about convergence diagnostics is http://www.ensae.fr//crest/statistique/robert/McDiag). Generally, experimentation with different settings of t and m are necessary. We do not view this as a drawback since random generation is now computationally inexpensive.

For the general change-point model (1), each of the full conditional distributions $p(\tau|\theta_1, \theta_2, \mathbf{x})$, $p(\theta_1|\theta_2, \tau, \mathbf{x})$ and $p(\theta_2|\theta_1, \tau, \mathbf{x})$ is proportional to expression (3). Moreover, whatever the prior dependence structure between the parameters θ_1 and θ_2 may be, $p(\tau|\theta_1, \theta_2, \mathbf{x})$ is exactly of the form

$$p(\tau|\theta_1, \theta_2, \mathbf{x}) = \frac{p(\mathbf{x}|\theta_1, \theta_2, \tau)p(\tau)}{\sum_{\tau=1}^{n-1} p(\mathbf{x}|\theta_1, \theta_2, \tau)p(\tau)} \tag{11}$$

and can be easily sampled since it involves a discrete density. Suppose that θ_1 and θ_2 are assumed independent so that $p(\theta_1, \theta_2) = p(\theta_1)p(\theta_2)$. If $p(\theta_1)$ is conjugate with $p(x_i|\theta_1)$ and $p(\theta_2)$ is conjugate with $p(x_i|\theta_2)$ then $p(\theta_1|\theta_2, \tau, \mathbf{x})$ does not depend on θ_2 and is merely the prior $p(\theta_1)$ updated by the data $x_1, \ldots, x_\tau$ while $p(\theta_2|\theta_1, \tau, \mathbf{x})$ is $p(\theta_2)$ updated by $x_{\tau+1}, \ldots, x_n$. Since $p(\theta_1)$ and $p(\theta_2)$ belong to standard parametric families, it is straightforward to sample from them. More generally, this property still remains when only conditional independence upon a "nuisance" parameter, which describes the unknown but constant structure of the distribution on both sides of τ (σ^2 in M_1 and μ in M_2). Therefore, with these hypotheses, the Gibbs sampler is particularly suitable for change-point analysis. If the assumption of conjugate priors and of independence is dropped, the full conditionals may not be expressed in a simple form and we must sample from nonstandardized densities of the form (3). In this case, more general Markov Chain Monte Carlo approaches are needed (Robert, 1996, 1998). The most popular tool that one can implement is undoubtedly the Metropolis–Hasting algorithm (Metropolis et al., 1953).

Because prior distribution (7) assumes independence between σ_1^2 and σ_2^2, the collection of full conditional densities $p(\mu|\boldsymbol{\sigma}^2, \tau, \mathbf{x}, M_2)$, $p(\sigma_1^2|\mu, \sigma_2^2, \tau, \mathbf{x}, M_2)$ and $p(\sigma_2^2|\mu, \sigma_1^2, \tau, \mathbf{x}, M_2)$ can readily be determined for model M_2. Each of these distributions is proportional to expression (8) and belong to the same parametric family as their corresponding prior. They are therefore available for sampling. Finally, as it was outlined above, $p(\tau|\mu, \boldsymbol{\sigma}^2, \mathbf{x}, M_2)$ is exactly of the form (11).

More precisely, from expressions (6)–(8) it is straightforward to verify that the full conditionals for M_2 are as

$$p(\mu|\boldsymbol{\sigma}^2, \tau, \mathbf{x}, M_2) = \mathscr{N}\left(\mu\left|\phi'', \frac{\lambda'\sigma_1^2\sigma_2^2}{\lambda'(n-\tau)\sigma_1^2 + \sigma_2^2}\right.\right)$$

$$p(\sigma_1^2|\mu, \sigma_2^2, \tau, \mathbf{x}, M_2) = \mathscr{IG}\left(\sigma_1^2\left|\alpha_1', \frac{\tau}{2}\left[s_\tau^2 + (\bar{x}_\tau - \mu)^2 + \frac{(\mu-\phi)^2}{\tau\lambda}\right] + \beta_1\right.\right)$$

$$p(\sigma_2^2|\mu, \sigma_1^2, \tau, \mathbf{x}, M_2) = \mathscr{IG}\left(\sigma_2^2\left|\alpha_2', \frac{n-\tau}{2}[s_{n-\tau}^2 + (\bar{x}_{n-\tau} - \mu)^2] + \beta_2\right.\right)$$

$$p(\tau|\mu, \boldsymbol{\sigma}^2, \mathbf{x}, M_2) = \frac{p(\mathbf{x}|\mu, \boldsymbol{\sigma}^2, \tau, M_2)p(\tau|M_2)}{\sum_{\tau=1}^{n-1} p(\mathbf{x}|\mu, \boldsymbol{\sigma}^2, \tau, M_2)p(\tau|M_2)}$$

Highly efficient estimates for the marginal posterior distributions of each parameters $\mu, \sigma_1^2, \sigma_2^2$ and τ from Gibbs sampler simulated values can then be obtained using a Rao–Blackwellized estimator. Following the discussion above, we can also transform σ_2^2 to $\eta\sigma_1^2$ to obtain the full conditional distribution of η as

$$p(\eta|\mu, \sigma_1^2, \tau, \mathbf{x}, M_2) = \mathscr{IG}\left(\eta\left|\alpha_2', \frac{n-\tau}{2\sigma_1^2}[s_{n-\tau}^2 + (\bar{x}_{n-\tau} - \mu)^2] + \frac{\beta_2}{\sigma_1^2}\right.\right)$$

and use it to evaluate the Rao–Blackwellized estimate for $p(\eta|\mathbf{x}, M_2)$, i.e.

$$\hat{p}(\eta|\mathbf{x}, M_2) = \frac{1}{m}\sum_{j=1}^{m} p(\eta|\mu_j^{(t)}, \sigma_{1j}^{2(t)}, \tau_j^{(t)}, \mathbf{x}, M_2) \qquad (12)$$

4. Applications

Bayesian change-point analysis procedure is now applied to annual energy inflow series presented in the introduction. Model M_1 is assumed for the Churchill Falls hydropower system, while model M_2 is considered for the Outaouais hydropower system.

4.1. Specifying prior distributions

The first step in Bayesian analysis is to set up a *full probability model*. That is, in addition to modeling the observable quantities (model M_1 or M_2 for the annual energy inflow), we must represent the prior degree of belief about the unknowns, i.e. the parameters of the models. In Bayesian analysis, specifying a prior distribution for the parameters is an integral part of the modeling task, with all hypotheses that modeling involves. Prior elicitation is therefore a crucial component of the Bayesian approach. Generally, a compromise has to be reached between mathematical simplicity and realism. The conjugate prior distribution assumption adopted here (Section 2) facilitates the derivation of posterior distributions and may be criticized for lack of realism. Still, the family of normal-inverted gamma is flexible and allows for the representation of a wide spectrum of prior knowledge. To be fully operational, it is desirable to specify the prior distributions by eliciting knowledge of an expert and/or by using other information than the data themselves. An engineer should have valuable prior information about annual energy inflow behavior (subjective knowledge, regional information, etc.). To our knowledge, the first convincing approach to specify realistic priors in a hydrological context has

Table 1
Data set with $y_{\bar{e}}$, y_{var} and $\bar{P}$ calculated before the prior expected change-point 1970 ($y_{\bar{e}}$: average annual energy inflow; y_{var}: variance of annual energy inflow; A: basin area; $\bar{P}$: average annual precipitation; C(TW): installed generating capacity)

System	$y_{\bar{e}}$ (TWh)	y_{var} (TWh^2)	A (km^2)	$\bar{P}$ (mm)	C
Bersimis	7.65	0.56	15 695	975	0.00171
Churchill Falls	35.36	28.14	92 432	891	0.00543
La Grande	79.10	102.14	176 472	764	0.01524
Churchill Falls	22.24	4.92	45 480	1030	0.00502
Outaouais	10.05	3.62	225 280	934	0.00188
Outardes	9.92	1.28	18 798	1008	0.00184
St-Maurice	9.01	2.14	51 813	1016	0.00164
St-Laurent	11.62	1.41	794 000	982	0.00226

been proposed by Bernier (1967) for frequency analysis with the lognormal distribution. Hyperparameter values were specified using estimated quantiles from a preliminary study. Posterior quantiles and credible intervals were then obtained using Bayes theorem. More recently, Coles and Tawn (1996) used a similar approach for the analysis of extreme rainfall data with the GEV probability distribution. To specify the prior density for the parameters of the GEV, they elicited subjective knowledge of an expert hydrologist within the quantile space, a scale with which he has familiarity. For more discussion and references on prior specification see Berger (1985) and Bernardo and Smith (1994).

Experts in Hydro-Québec do not agree about the existence of a shift either in the mean level or in the variability of annual energy inflows. Moreover, specialists who suspect the existence of a sudden change are not willing to specify their degree of belief about the year at which it occurred without referring to the data of interest. This information is therefore biased. In fact, the use of any of these data in the prior formulation is not formally acceptable in a Bayesian analysis (Berger, 1985). Finally, if we examine the annual energy inflows for the other Hydro-Québec hydropower systems, no clear change-point can be identified. Therefore, one cannot reasonably favor any year of change, and we assume τ is distributed as a uniform discrete distribution for both model M_1 and M_2. Note that in this case, the prior expected change-point is 1970, i.e. the mean of a discrete uniform probability distribution on the interval [1943, 1995].

For the other parameters, the prior degrees of belief were assumed to be represented by normal-inverted gamma type of distributions (see A2 for M_1 and expression (7) for M_2). Complete specification of prior knowledge only requires the choice of the hyperparameters in these expressions. Because it is felt that the annual energy inflow of all hydropower systems are related in some way, we based the determination of these quantities on a regional analysis. The information available to establish a transfer function between the site of interest and the other systems consists, for each hydropower system, of four variables: average annual energy inflow ($y_{\bar{e}}$), variance of annual energy inflow (y_{var}), basin area (A), average annual precipitation ($\bar{P}$) and installed generating capacity (C). Two data sets were considered, corresponding, respectively, to the values of $y_{\bar{e}}$, y_{var} and $\bar{P}$ evaluated before and after the prior expected year of change $E\{\tau\} = 1970$. Table 1 gives the first data set.

The idea is to construct a simple model to predict the average energy inflow $y_{\bar{e}}$ and the variance y_{var} of the energy inflows for the site of interest, before and after 1970. These predictions, along with their standard error, are then used to elicit the mean and the variance of the prior distributions. Finally, solving a simple system of equations for the first two moments leads to estimated values for the hyperparameters based on regional information. Since a regional model is used to specify hyperparameters by transferring knowledge from the nearby sites, this approach, in some way, has much in common with an empirical hierarchical Bayesian analysis (Berger, 1985).

To illustrate the procedure, let us suppose that we want to specify the hyperparameters of a normal-inverted gamma prior distribution $\mathcal{N}(\mu|\phi, \lambda\sigma^2)\mathcal{IG}(\sigma^2|\alpha, \beta)$. If we denote the predicted values of $y_{\bar{e}}$ and y_{var}, respectively, by $\hat{y}_{\bar{e}}$ and $\hat{y}_{\mathrm{var}}$, and

Table 2
Results of regression analyses for Churchill Falls and Ouataouais

Systems		1943–1970			1971–1996		
		$\hat{\gamma}$	$\hat{y}$	$s^2(\hat{y})$	$\hat{\gamma}$	$\hat{y}$	$s^2(\hat{y})$
Churchill Falls	$y_{\bar{e}}$	5120	27.80	2.80	5121	27.81	4.50
	y_{var}	436 980	12.88	7.45	617 000	18.19	6.00
Outaouais	$y_{\bar{e}}$	5253	9.88	11.26	5242	9.85	11.20
	y_{var}	444 950	1.57	44.83	624 000	2.21	34.26

their corresponding variances by $s^2(\hat{y}_{\bar{e}})$ and $s^2(\hat{y}_{\mathrm{var}})$, it is natural to take as prior mean and variance for μ and σ^2

$$E\{\mu\} = \hat{y}_{\bar{e}}, \qquad \mathrm{Var}\{\mu\} = s^2(\hat{y}_{\bar{e}}), \qquad E\{\sigma^2\} = \hat{y}_{\mathrm{var}},$$
$$\mathrm{Var}\{\sigma^2\} = s^2(\hat{y}_{\mathrm{var}})$$

Using expressions for prior mean and variance of normal and inverted gamma distributions (Bernardo and Smith, 1994, p. 431), and then solving this system of equations leads us to

$$\alpha = 2 + \hat{y}^2_{\mathrm{var}}/s^2(\hat{y}_{\mathrm{var}}), \qquad \beta = \hat{y}_{\mathrm{var}}(\alpha - 1),$$
$$\phi = \hat{y}_{\bar{e}} \text{ and } \lambda = s^2(\hat{y}_{\bar{e}})/\hat{y}_{\mathrm{var}}. \tag{13}$$

Linear regression analysis was performed for each data set (the first one is given in Table 1), discarding the site of interest. A simple regression considering only the generating capacity C appeared to be the best regional models. More precisely, we have

$$y_{\bar{e}} = \gamma_{\bar{e}}C + \epsilon \text{ and } y_{\mathrm{var}} = \gamma_{\mathrm{var}}C^2 + \epsilon.$$

In the particular case of hydropower systems in Québec, precipitation and drainage area are very poorly related to the average energy inflow or the variance compared to generating capacity C. Indeed, variable C is a good summary of what an engineer would think of the potential annual energy output the company may derive when installing a hydropower plant. It can be used as a prior knowledge of what can be "intuitively" expected from the site, before the production started, that is in the state of knowledge before collecting a long series of observations in working conditions. The estimated coefficients and the corresponding prediction along with its variance are given in Table 2.

Using the results in Table 2 in expression (13) leads to the values of the hyperparameters for the two series. To summarize, the prior mean and standard deviation of each parameter are listed in Table 3 for the two cases.

This table will be used in the next sections to show how prior state of belief is updated by the data.

4.2. Churchill Falls power system

Expressions given in Appendix A are used to analyze the annual energy inflow of Churchill Falls hydropower complex (Fig. 1). Fig. 3 displays the annual inflows and all marginal posterior densities of the parameters of interest together with their corresponding prior density (in dotted line). We first observe how the data modified or updated prior

Table 3
Prior specifications from regional analysis

		τ	μ_1	μ_2	δ	σ^2
Churchill Falls	Exp. value	1970	27.80	27.81	0.01	12.88
(model M_1)	Stand. deviation	16	1.67	2.12	2.70	2.73
		τ	σ_1^2	σ_2^2	η	μ
Outaouais	Exp. value	1970	1.57	2.21	1.41	9.88
(model M_2)	Stand. deviation	16	6.70	5.85	4.03	3.36

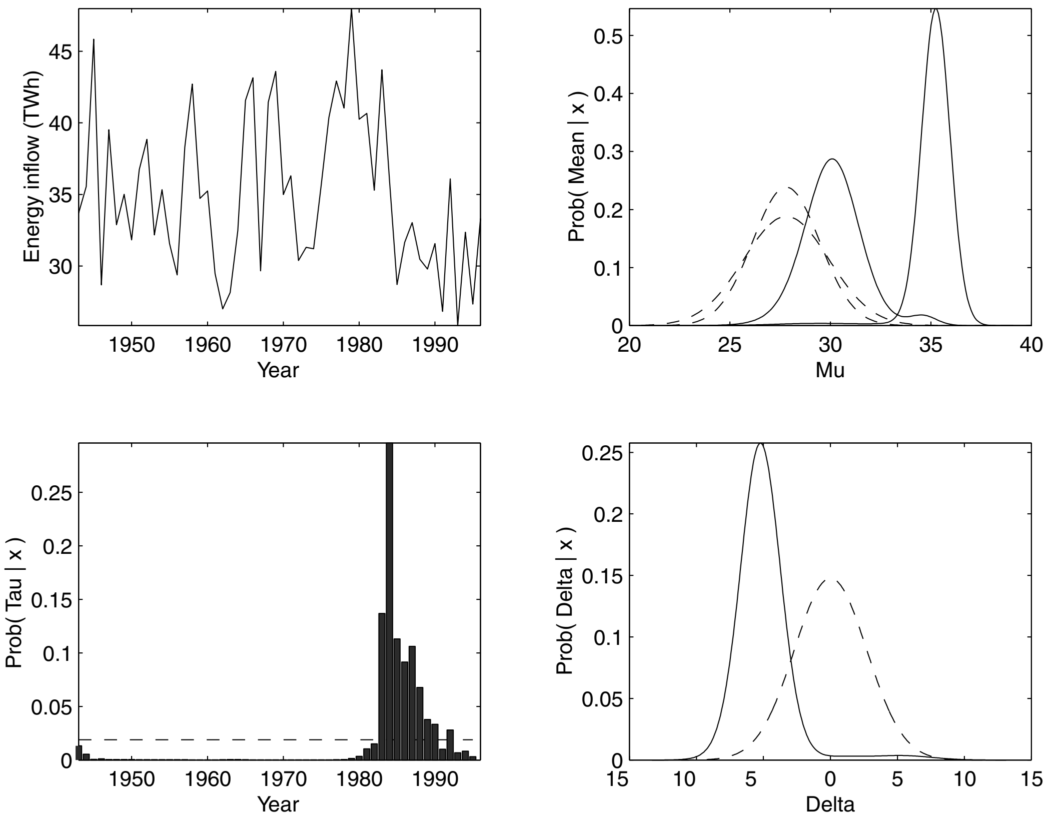

Fig. 3. Marginal posterior distributions for model M_1: Churchill Falls power system.

information by comparing the prior and the posterior distributions. Such graphs are also useful tools to interpret the parameter's uncertainty. Table 4 summarizes estimated posterior quantities (mode, mean and standard deviation). These estimates are to be compared with their corresponding prior values specified in Section 4.1 (Table 3). Table 4 also gives, for each parameter, a symmetric 90% Bayesian credible interval.

This analysis indicates first, under the hypothesis of a change of type M_1 in the annual energy inflow, that the change-point occurred around 1984 with approximately a 7-year standard deviation. Also, the mean level of energy inflow after the shift seems to have decreased by an amount of 5 TWh with 1.5 TWh of standard deviation. The credible interval for the intensity of the shift shows that a decrease of almost 8 TWh is still plausible at a 90% credible level. Moreover, this interval do not contain zero which suggests a negative change of at least 2 TWh. However, this observation provides no justification for actions such as "rejecting the no change hypothesis" (the model assumes a change did occur). Under such hypothesis, the most probable intensity of the shift being -5 TWh, the hydropower company would better take action to compensate this potential loss of energy inflow and ensure balance between supply and demand for the next few years. The output of this Bayesian analysis

Table 4
Posterior moments and credible interval for Churchill Falls

Parameters	Mode	Mean	Stand. dev.	90% Credible int.
τ	1984	1984	6.99	[1981; 1990]
μ_1	35.28	35.11	0.74	[33.75; 36.28]
μ_2	30.05	30.16	1.33	[27.61; 32.76]
δ	-5.21	-4.95	1.52	[-7.86; -2.13]

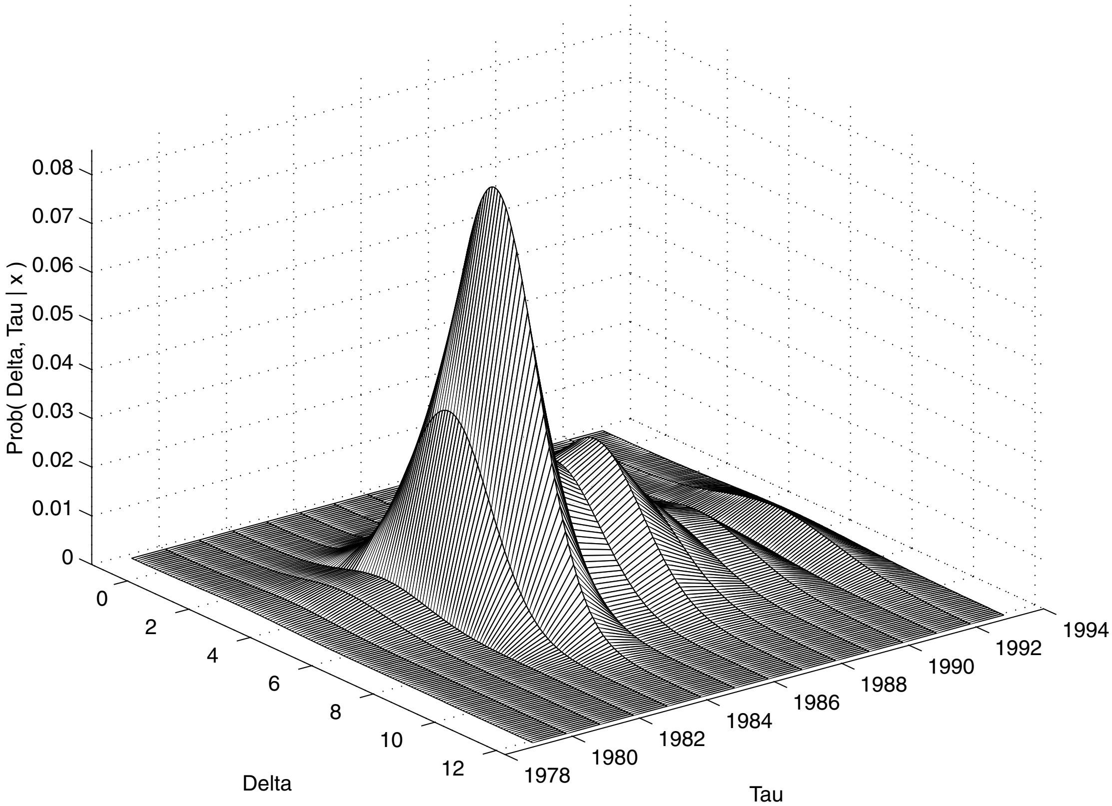

Fig. 4. Joint posterior distribution of (δ, τ) for model M_1: Churchill Falls power system.

(estimates and uncertainties) can be used to choose among different alternatives (anticipating construction of new power plants, buying electricity, using thermal power station, etc.).

One of the advantages of a Bayesian analysis is that it offers not only marginal but also all joint posterior distributions, allowing for an assessment of dependencies among parameters. As an illustration, Fig. 4 presents the joint posterior distribution of δ and τ. It shows in particular how, after considering the data, δ and τ are very closely related, even though we assumed prior independence between the change-point and the other parameters.

Our approach assumes no interannual correlations on both sides of τ. Autocorrelations (lag 1–5) are given in Table 5 separately for the period 1943–1984 and 1985–1996.

The first subseries does not exhibit any significant interannual correlations. However, after the estimated change-point the autocorrelations are significant. One could revise the analysis by considering for example an AR model with a changing coefficient of

Table 5
Autocorrelations for Churchill Falls (Note: p-values for H_0: “no correlation” appear in parentheses)

Lag	1	2	3	4	5
1943–1984	0.24	0.17	0.02	−0.10	−0.17
	(0.1066)	(0.1419)	(0.2707)	(0.3521)	(0.3124)
1985–1996	−0.74	0.49	−0.22	0.26	−0.67
	(0.0040)	(0.0021)	(0.0011)	(0.0016)	(0.0036)

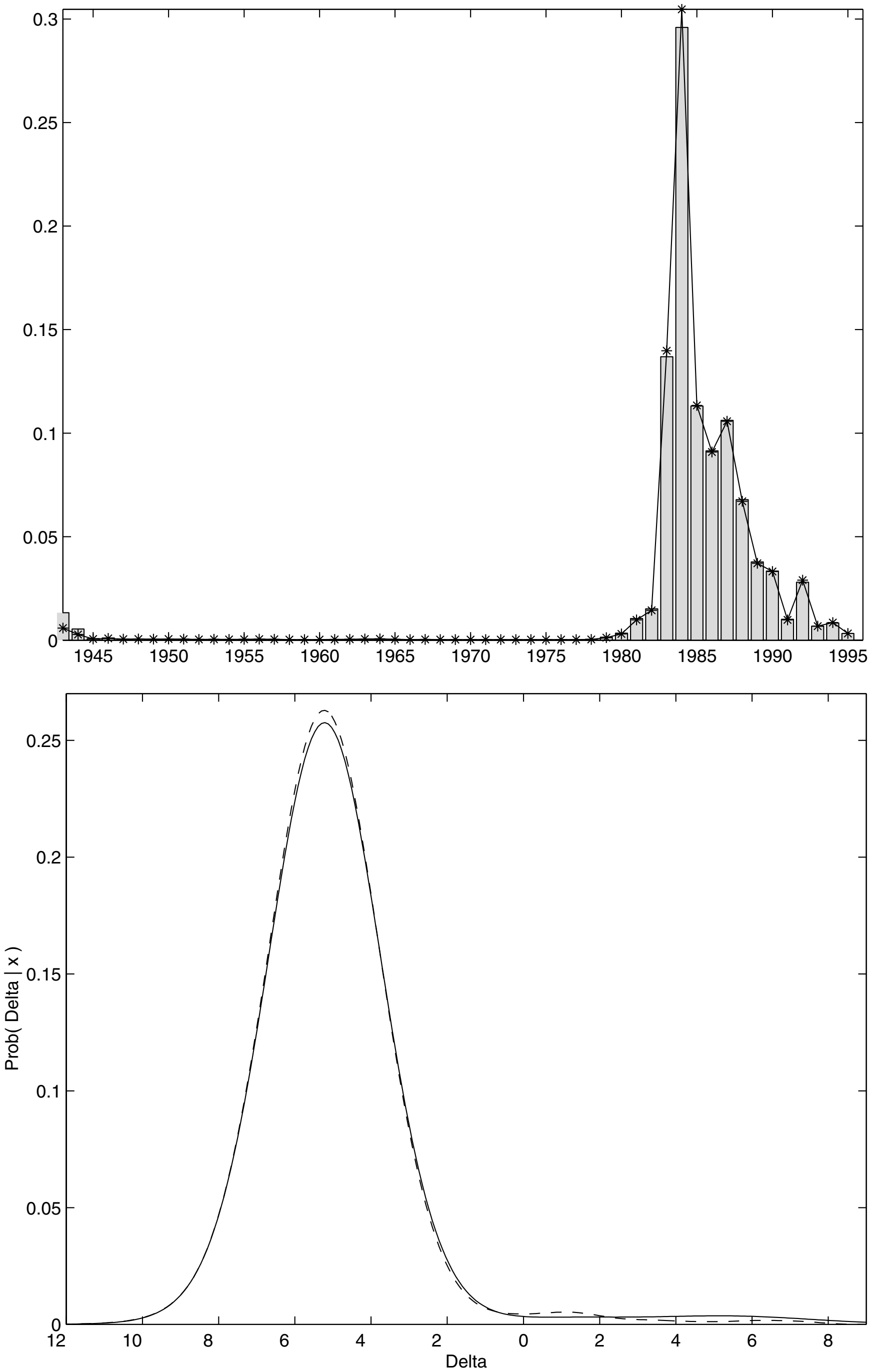

Fig. 5. Comparison of exact and approximate posterior distributions ($t = 1000$, $m = 1000$).

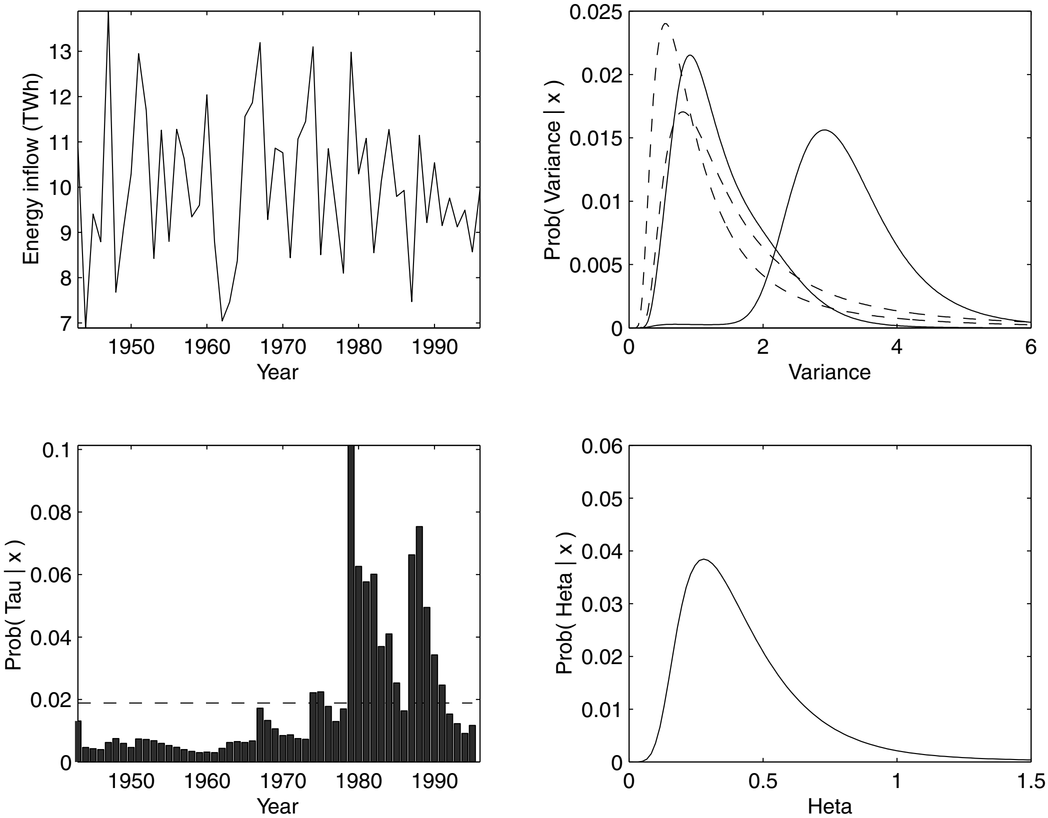

Fig. 6. Marginal posterior distributions for model M_2: Outaouais power system.

autocorrelation. Note that the estimated correlations after the change-point are only based on ten observations. Therefore, it may not be relevant to increase the number of parameters based on such uncertain estimates.

Before analyzing the change in the variance for the annual energy inflow of the Outaouais power system using Gibbs sampler, we first outline the way that the algorithm would be typically implemented by reconsidering model M_1. We attempt to reproduce the analytical results for the Churchill Falls example by estimating the marginal posterior densities of interest using Gibbs algorithm. Under the assumption of model M_1 and the joint prior distribution in Appendix A, the full conditional posterior distributions for each parameter are given in Appendix B. They are readily available for sampling using standard routines. Note that since $p(\tau|\boldsymbol{\mu}, \sigma^2, \mathbf{x}, M_1)$ is univariate and discrete, sampling from this distribution during the Gibbs sampler cycle is straightforward by simple function inversion.

Fig. 5 compares the results obtained above with those evaluated using the Gibbs sampler for the marginal posterior distributions of the change point τ (continuous line with stars at each time point) and the intensity of shift δ (dotted line). Discarding the first $t = 1000$ values of the chain, the conditional densities were averaged over $m = 1000$ replications using the Rao–Blackwellized density estimate (see

Table 6
Posterior moments and credible interval for Outaouais

Parameters	Mode	Mean	Stand. dev.	90% Credible int.
τ	1979	1978	12.81	[1948; 1991]
σ_1^2	2.77	3.03	0.77	[1.82; 4.48]
σ_2^2	0.90	1.43	0.52	[0.53; 2.80]
η	0.30	0.58	0.20	[0.15; 1.30]

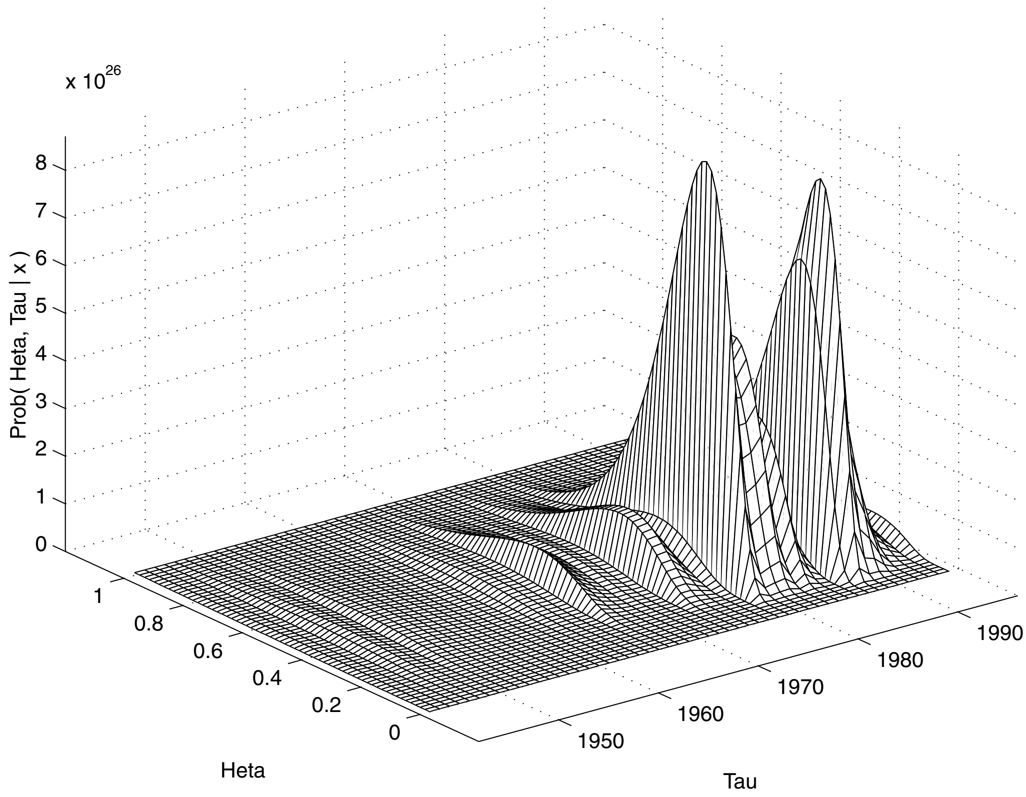

Fig. 7. Joint posterior distribution of (η, τ) for model M_2: Outaouais power system.

(B.3) for $\hat{p}(\delta|\mathbf{x}, M_1)$). There is a clear agreement between the exact and approximate densities in this case. Further experimentation showed that the Gibbs sampler could reproduce the exact posterior distributions in considerably fewer than $t = 1000$ iterations. This is confirmed by inspection of 1000 replicates after, say, 100 iterations.

4.3. Outaouais power system

Using the full conditionals given in Section 3 and Gibbs sampler, we analyze the annual energy inflows calculated at the Outaouais hydropower system. Inferences presented below are also based on $t = 1000$ and $m = 1000$. Fig. 6 displays the annual energy inflows and all marginal posterior densities of the parameters of interest together with their corresponding prior density (in dotted line). In Table 6 we list the estimated posterior quantities and 90% credible intervals. Finally, Fig. 7 presents the joint posterior distribution of η and τ.

The analysis indicates, under the hypothesis of a change of type M_2 in the annual energy inflow, that the variance after the change is at least twice as small than it was before the change (the posterior mode and mean for η are, respectively, 0.30 and 0.55). The change-point seems to be 1979 with approximately 12 years of standard deviation. Comparing these values to the prior expectation (Table 3), clearly the information stemming from the data contributed to modifying considerably the prior state of belief. The credible interval for η shows that an almost sevenfold decrease is still plausible at a 90% credible level. But this interval also suggests a much smaller change, since it contains one, i.e. $\sigma_1^2 \approx \sigma_2^2$.

These results may seem statistically convincing in favor of a change, even though the procedure is not a statistical test. But, is such a decrease in the variability ($\eta \approx 0.5$) enough to change the strategy of the

Table 7
Autocorrelations for Outaouais (Note: p-values for H_0: “no correlation” appear in parentheses)

Lag	1	2	3	4	5
1943–1979	−0.04	−0.05	−0.23	−0.03	0.04
	(0.7824)	(0.9121)	(0.4832)	(0.6432)	(0.7643)
1980–1996	−0.35	0.13	−0.27	0.26	0.05
	(0.1157)	(0.2408)	(0.2110)	(0.1831)	(0.2790)

company’s development for the future? Clearly, decisional consequences have to be considered together with the statistical analysis. If Hydro-Québec changes its policy only in light of these statistical results, the company may face costly consequences. In the particular case of a change in the variability of energy inflows, one should take into account the fact that the losses associated with the decisions, such as contracting or not new agreements to sell electricity, are probably asymmetrical. Suppose the variance is in fact one half as large as before. The uncertainty around the mean becomes smaller and the future is somewhat less variable. Therefore, the mean value, as a prediction for future observations, is more credible that it was in the past, so that new agreements to sell electricity are relatively less risky. The company can then contract for delivering more power. If, in reality, the variance has not changed, as it was previously believed, future values of energy inflows may deviate from the mean much more frequently than anticipated. In this case, the company would not respect its contractual engagements, resulting in dramatic consequences. On the other hand, if the company would decide not to change its policy of development, when indeed the variance has decreased, good opportunities may be lost but catastrophic consequences would not to be encountered as in the former case.

Finally, the two subseries do not exhibit any significant interannual correlations (Table 7).

5. Discussion and conclusions

The Bayesian method presented in this paper can be viewed as an extension of the normal models and practitioners can perform such change-point analysis routinely using standard statistical toolboxes. This approach can be generalized to other type of normal models (a simultaneous change in the mean and variance for instance, see Perreault et al., 2000a) and even to other types of p.d.f.s, for instance the gamma distribution. More precisely, use of probability distributions which belong to the exponential class of p.d.f.s allows for exactly the same line of reasoning based on conjugacy.

The following conclusions have been reached in this study:

- Hydrologists can take full advantage of the existence of conjugate distributions when studying a single change-point in a statistical model belonging to the exponential family. If prior independence between the epoch of change and model parameters is assumed, the joint posterior distribution is a finite mixture of conjugate distributions. This allows for easy computation of posterior odds, as illustrated by the univariate normal model under the configuration of a single change in the mean level.
- It was showed that more complex change-point problems can be readily addressed by Gibbs sampling, which is easily accessible to the average statistical practitioner. This was illustrated in the case of a single change in the variance for which no explicit expressions for the posterior distributions can be found, unless one considers noninformative priors. In hydrology, the case of a single change in the variance occurring at an unknown time point was neglected although this parameter is important for water resources management. The Gibbs sampling scheme could also be used if one is interested to relax the independence assumption. For instance, change-point in AR models may be treated by generalizing the approach proposed by Chib (1993).
- Two case studies involving annual energy inflow series were presented to illustrate how Bayesian change-point can be performed, from the specification of the prior distributions up to the

interpretation of posterior odds. It was shown that several questions can be answered simultaneously from the posterior distributions. For instance, one can not only estimate the parameters but can also directly obtain accuracy measures for the estimate and interpret dependence between them. This is in contrast to classical statistics, for which obtaining estimates and determining their reliability are two different problems. It was also seen that interpretation of the results is easy and straightforward. The frequentist interpretation is difficult to understand for a nonstatistician, and water resources managers often find it much too abstract for operational concerns. For example, interpretation of classical confidence intervals involves averaging over all possible data, while it is known which data did in fact occur. Practitioners usually interpret the results of a classical confidence interval as a Bayesian credible interval (Lecoutre, 1997). Moreover, recourse to the usual asymptotic assumption required for classical inference (see Hinkley, 1970) is not needed here, and therefore Bayesian change-point analysis can be applied to modest sample sizes.

- Unlike many published Bayesian analyses, the problem of eliciting prior knowledge was not skirted by using noninformative priors or even worse, by specifying hyperparameters based on a subset of the actual sample. To ensure maximum efficiency, all observations must in fact be used to update the prior state of belief, and not to specify hyperparameters. In this paper, prior elicitation relied on expert knowledge and on external information about the studied phenomena. A simple regional analysis was performed to specify the hyperparameters and represent the prior state of belief. Effort in this direction should be developed to take full advantage of the Bayesian framework.

The problem of model uncertainty has not been addressed here. All results herein are conditional upon a given model. Both M_1 and M_2 assume with certainty that a change has occurred. This statistical reporting framework is essential but incomplete. It forces the hydrologist to choose among the different models to perform a change-point analysis, and to exclude some other kinds of changes. To take into account the “no change hypothesis” and diverse types of changes, one should assign a prior probability to the different alternatives, and consider the change-point study as a Bayesian model selection problem among the various situations that may occur. Using the results presented in this paper, this perspective for hydrological change-point analysis is explored in Perreault et al. (2000a). The important operational issue of forecasting in changing sequences is also studied.

Acknowledgements

We wish to express our deep gratitude to Professor Lucien Duckstein for his critical review of this paper. We are also grateful to René Roy and Raymond Gauthier of Hydro-Québec for providing data. Finally, we gratefully acknowledge the constructive comments of two anonymous referees on an earlier version that considerably improved this paper. This research was supported by la Direction Générale de l’Enseignement et de la Recherche de France of the French ministery of Agriculture (DGER), les fonds pour la Formation de Chercheurs et l’Aide à la Recherche du Québec (FCAR), and la Chaire en Hydrologie Statistique Hydro-Québec/CRSNG (INRS-Eau).

Appendix A

A.1. Bayesian inference for model $\boldsymbol{M_1}$

The likelihood function resulting from n observations $\mathbf{x} = (x_1, x_2, \dots, x_n)$ generated by model M_1 can be written as

$$p(\mathbf{x}|\boldsymbol{\mu}, \sigma^2, \tau, M_1) = \prod_{i=1}^{\tau} \mathcal{N}(x_i|\mu_1, \sigma^2) \prod_{i=\tau+1}^{n} \mathcal{N}(x_i|\mu_2, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{(n/2)} \exp\left\{-\frac{\tau}{2\sigma^2}[s_\tau^2 + (\bar{x}_\tau - \mu_1)^2]\right\} \times \exp\left\{-\frac{(n-\tau)}{2\sigma^2}[s_{n-\tau}^2 + (\bar{x}_{n-\tau} - \mu_2)^2]\right\} \quad \text{(A1)}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and

$$\bar{x}_\tau = \sum_{i=1}^{\tau} \frac{x_i}{\tau}, \qquad \bar{x}_{n-\tau} = \sum_{i=\tau+1}^{n} \frac{x_i}{n-\tau},$$

$$s_\tau^2 = \sum_{i=1}^{\tau} \frac{(x_i - \bar{x}_\tau)^2}{\tau}, \qquad s_{n-\tau}^2 = \sum_{i=\tau+1}^{n} \frac{(x_i - \bar{x}_{n-\tau})^2}{n-\tau}.$$

For fixed τ, the likelihood (A1) has the structure of a product of two normal distributions with one inverted gamma distribution, which suggests an ad hoc normal-inverted gamma type of distribution to represent prior knowledge about $\boldsymbol{\mu}$ and σ^2. Assuming for model M_1 prior independence between τ and the other parameters $(\boldsymbol{\mu}, \sigma^2)$, and that $p(\tau|M_1)$ is any discrete distribution on the set $\{1, 2, \ldots, n-1\}$, leads to the joint prior parameter p.d.f.:

$$\begin{aligned} &p(\boldsymbol{\mu}, \sigma^2, \tau|M_1) \\ &\quad = \mathscr{N}(\mu_1|\phi_1, \lambda_1\sigma^2)\mathscr{N}(\mu_2|\phi_2, \lambda_2\sigma^2)\mathscr{IG}(\sigma^2|\alpha, \beta) \\ &\qquad \times p(\tau|M_1) = \mathscr{NNIG}(\boldsymbol{\mu}\sigma^2|\boldsymbol{\phi}, \boldsymbol{\lambda}, \alpha, \beta)p(\tau|M_1) \end{aligned} \tag{A2}$$

where $\phi = (\phi_1, \phi_2)$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, and $\mathscr{IG}(x|\alpha, \beta)$ stands for the inverted gamma p.d.f. with parameters $\alpha \in \Re^+$ and $\beta \in \Re^+$ (see Bernardo and Smith, 1994, for the density).

Because of conjugate properties (Berger, 1985), under M_1, the conditional joint posterior distribution $p(\boldsymbol{\mu}, \sigma^2|\tau, \mathbf{x}, M_1)$ given τ and the observed data $\mathbf{x}$ also belongs to the class of normal-inverted gamma distributions, but with updated parameters $(\boldsymbol{\phi}', \boldsymbol{\lambda}', \alpha', \beta')$. More precisely,

$$p(\boldsymbol{\mu}, \sigma^2|\tau, \mathbf{x}, M_1) = \mathscr{NNIG}(\boldsymbol{\mu}, \sigma^2|\boldsymbol{\phi}', \boldsymbol{\lambda}', \alpha', \beta'), \tag{A3}$$

where

$$\lambda_1' = \lambda_1/(1 + \tau\lambda_1), \quad \phi_1' = (1 - \lambda_1'\tau)\phi_1 + \lambda_1'\tau\bar{x}_\tau,$$

$$\begin{aligned} &\lambda_2' = \lambda_2/[1 + (n-\tau)\lambda_2], \quad \phi_2' \\ &\quad = [1 - \lambda_2'(n-\tau)]\phi_2 + \lambda_2'(n-\tau)\bar{x}_{n-\tau}, \end{aligned}$$

$$\begin{aligned} \beta' &= \frac{\tau}{2}[s_\tau^2 + (1 - \lambda_1'\tau)(\phi_1 - \bar{x}_\tau)^2] \\ &+ \frac{n-\tau}{2}[s_{n-\tau}^2 + (1 - \lambda_2'(n-\tau))(\phi_2 - \bar{x}_{n-\tau})^2] + \beta, \end{aligned}$$

$$\alpha' = \alpha + n/2.$$

To simplify notation, dependence upon τ and $\mathbf{x}$ was omitted by writing for example λ_1' instead of $\lambda_1'(\mathbf{x}, \tau)$. The prior predictive density, i.e. the p.d.f. of the data $\mathbf{x}$ only conditioned upon the change-point $\tau, p(\mathbf{x}|\tau, M_1)$, which makes the Bayes theorem denominator, can be determined by dividing the joint p.d.f. $p(\boldsymbol{\mu}, \sigma^2, \mathbf{x}|\tau, M_1)$ by expression (A3), and canceling factors involving $\boldsymbol{\mu}$ and σ^2. The result is

$$p(\mathbf{x}|\tau, M_1) = \left(\frac{1}{2\pi}\right)^{n/2} \sqrt{\frac{\lambda_1'\lambda_2'}{\lambda_1\lambda_2}} \frac{\beta^\alpha}{(\beta')^{\alpha'}} \frac{\Gamma(\alpha')}{\Gamma(\alpha)}. \tag{A4}$$

Integrating the appropriate parameters out of expression (A3), leads to the conditional posterior distributions $p(\mu_1|\tau, \mathbf{x}, M_1)$ and $p(\mu_2|\tau, \mathbf{x}, M_1)$ of the means before and after the change-point. The conditional posterior density of the intensity of the shift $\delta = \mu_2 - \mu_1$, $p(\delta|\tau, \mathbf{x}, M_1)$, is then deduced by a simple univariate transformation of variable. These distributions are Student t-distributions, $\mathscr{ST}(a, b, c)$ (expression for the p.d.f. can be found in Bernardo and Smith, 1994, p. 432):

$$p(\mu_i|\tau, \mathbf{x}, M_1) = \mathscr{ST}(\phi_i', \alpha'(\lambda_i'\beta')^{-1}, 2\alpha'), \quad i = 1, 2$$

$$\begin{aligned} &p(\delta|\tau, \mathbf{x}, M_1) \\ &\quad = \mathscr{ST}(\phi_2' - \phi_1', \alpha'(\lambda_1' + \lambda_2')^{-1}, (\beta')^{-1}, 2\alpha') \end{aligned}$$

In our problem, τ is unknown and its marginal posterior distribution $p(\tau|\mathbf{x}, M_1)$ has to be derived. Using Bayes theorem and the prior predictive density (A4), the marginal posterior density of the change-point $\tau = 1, 2, \ldots, n-1$ under model M_1 is seen to be

$$\begin{aligned} p(\tau|\mathbf{x}, M_1) &= \frac{p(\mathbf{x}|\tau, M_1)p(\tau|M_1)}{\sum_{\tau=1}^{n-1} p(\mathbf{x}|\tau, M_1)p(\tau|M_1)} \\ &\propto p(\tau|M_1)\sqrt{\lambda_1'\lambda_2'}(\beta')^{-\alpha'} \end{aligned} \tag{A5}$$

This distribution is discrete and gives, for each time point, the posterior probability of shift occurrence in the mean level assuming a change occurred with certainty. Finally, to draw conclusions regarding the means before and after the change-point and the intensity of shift, the marginal posterior distributions $p(\mu_1|\mathbf{x}, M_1)$, $p(\mu_2|\mathbf{x}, M_1)$ and $p(\delta|\mathbf{x}, M_1)$ must be evaluated. This is done by averaging the corresponding conditional posterior distribution with respect to the posterior mass function of τ:

$$p(\mu_i|\mathbf{x}, M_1) = \sum_{\tau=1}^{n-1} p(\mu_i|\tau, \mathbf{x}, M_1) p(\tau|\mathbf{x}, M_1)$$

$$p(\delta|\mathbf{x}, M_1) = \sum_{\tau=1}^{n-1} p(\delta|\tau, \mathbf{x}, M_1) p(\tau|\mathbf{x}, M_1)$$

where $i = 1, 2$. The marginal distributions of μ_1, μ_2 and δ appear as finite mixtures of Student *t*-distributions weighted by the $n - 1$ values of expression (A5).

A.2. Full conditionals for model M_1

The collection of full conditionals for model M_1 is proportional to the product

$$p(\mathbf{x}|\boldsymbol{\mu}, \sigma^2, \tau, M_1) p(\boldsymbol{\mu}, \sigma^2, \tau|M_1), \tag{B1}$$

and can be seen to be

$$p(\mu_1|\mu_2, \sigma^2, \tau, \mathbf{x}, M_1) = \mathcal{N}(\mu_1|\phi'_1, \lambda'_1\sigma^2)$$

$$p(\mu_2|\mu_1, \sigma^2, \tau, \mathbf{x}, M_1) = \mathcal{N}(\mu_2|\phi'_2, \lambda'_2\sigma^2)$$

$$p(\sigma^2|\boldsymbol{\mu}, \tau, \mathbf{x}, M_1)$$
$$= \mathcal{IG}\left(\sigma^2|\alpha' + 1, \frac{1}{2}\left[\frac{(\mu_1 - \phi'_1)^2}{\lambda'_1} + \frac{(\mu_2 - \phi'_2)^2}{\lambda'_2}\right] + \beta'\right)$$

$$p(\tau|\boldsymbol{\mu}, \sigma^2, \mathbf{x}, M_1) \propto \frac{p(\mathbf{x}|\boldsymbol{\mu}, \sigma^2, \tau, M_1) p(\tau|M_1)}{\sum_{\tau=1}^{n-1} p(\mathbf{x}|\boldsymbol{\mu}, \sigma^2, \tau, M_1) p(\tau|M_1)}$$

The other parameter of interest is $\delta = \mu_2 - \mu_1$. As in Section 3, we transformed μ_2 to δ to obtain a full conditional distribution:

$$p(\delta|\mu_1, \sigma^2, \tau, \mathbf{x}, M_1) = \mathcal{N}(\delta|\phi'_2 - \mu_1, \lambda'_2\sigma^2) \tag{B2}$$

and used it to evaluate a Rao–Blackwellized estimate for $p(\delta|\mathbf{x},M_1.)$:

$$\hat{p}(\delta|\mathbf{x}, M_1) = \frac{1}{m}\sum_{j=1}^{m} p(\delta|\mu_{1j}^{(t)}, \sigma_j^{2(t)}, \tau_j^{(t)}, \mathbf{x}, M_1). \tag{B3}$$

References

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, Springer, New York.

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory, Wiley, New York.

Bernier, J., 1994. Statistical detection of changes in geophysical series. In: Duckstein, L., Parent, E. (Eds.). Engineering Risk in Natural Resources Management, NATO Advanced Studies Institute SeriesKluwer, Dordrecht, pp. 159–176.

Bernier, J., 1967. Les méthodes bayésiennes en hydrologie statistique, Proceedings of the first Hydrology Symposium, Fort Collins, pp. 461–470.

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

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, MA.

Broemeling, L.D., 1985. Bayesian Analysis of Linear Models, Collection Statistics: Textbooks and Monographs, vol. 60, Marcel Dekker, New York.

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

Campbell, E.P., Fox, D.R., Bates, C.B., 1999. A Bayesian approach to parameter estimation and pooling in nonlinear flood event models. Water Resour. Res. 35, 211–220.

Carlin, B.P., Gelfand, A.E., Smith, A.F.M., 1992. Hierarchical Bayesian analysis of changepoint problems. Appl. Statist. 41, 389–405.

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

Chib, S., 1993. Bayes regression with autoregressive errors. A Gibbs sampling approach. J. Econometrics 58, 275–294.

Coles, S.G., Tawn, J.A., 1996. A Bayesian analysis of extreme rainfall data. Appl. Statist. 45, 463–478.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398–409.

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

Hinkley, D.V., 1970. Inference about the change-point in a sequence of random variables. Biometrika 57, 1–16.

Kerr, R.A., 1992. Unmasking a shifty climate system. Science 255, 1508–1510.

Kerr, R.A., 1999. A new force in high-latitude climate. Science 284, 241–242.

Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. J. Hydrol. 211, 69–85.

Lecoutre, B., 1997. C'est bon à savoir ! Et si vous étiez un bayésien qui s'ignore. Modulad 18, 81–87.

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

Lehmann, E.L., 1983. Theory of Point Estimation, Wiley, New York.

Lu, Z.Q., Berliner, L.M., 1999. Markov switching time series models with application to a daily runoff series. Water Resour. Res. 35, 523–534.

Menzefricke, U., 1981. A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. Appl. Statist. 30, 141–146.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1091.

Paturel, J.E., Servat, E., Kouamé, B., Lubès, H., Ouedraogo, M., Masson, J.M., 1997. Climatic variability in humid Africa along the Gulf of Guinea Part II: an integrated regional approach. J. Hydrol. 191, 16–36.

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

Perreault, L., Bernier, J., Bobée, B., Parent, E., 2000a. Journal of Hydrology. 235, 242–263.

Perreault, L., Parent, E., Bernier, J., Bobée, B., Slivitzky, M., 2000b. Environmental Research and Risk Assessment. (In press).

Robert, C., 1996. Méthodes de Monte Carlo par chaînes de Markov, Economica, Paris.

Robert, C., 1998. Discretization and MCMC Convergence Assessment, Lecture Notes in Statistics, vol. 135, Springer, New York.

Rodriguez-Iturbe, I., Entekhabi, D., Bras, R.L., 1991. Nonlinear dynamics of soil moisture at climate scales 1. Stochastic analysis. Water Resour. Res. 27, 1899–1906.

Servat, E., Paturel, J.E., Lubès, H., Kouamé, B., Ouedraogo, M., Masson, J.M., 1997. Climatic variability in humid Africa along the Gulf of Guinea Part I: detailed analysis of the phenomenon in Côte d'Ivoire. J. Hydrol. 191, 1–15.

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

Stephens, D.A., 1994. Bayesian retrospective multiple-changepoint identification. Appl. Statist. 43, 159–178.

Tanner, M.H., 1992. Tools for Statistical Inference: Observed Data and Data Augmentation Methods, Lecture Notes in Statistics, vol. 67, Springer, New York.