Journal of Hydrology 235 (2000) 242–263

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting

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

This paper provides a methodology to test existence, type, and strength of changes in the distribution of a sequence of hydrometeorological random variables. Unlike most published work on change-point analysis, which consider a single structure of change occurring with certainty, it allows for the consideration in the inference process of the no change hypothesis and various possible situations that may occur. The approach is based on Bayesian model selection and is illustrated using univariate normal models. Four univariate normal models are considered: the no change hypothesis, a single change in the mean level only, a single change in the variance only, and a simultaneous change in both the mean and the variance. First, inference analysis of posterior distributions via Gibbs sampling for a given change-point model is recalled. This scientific reporting framework is then generalized to the problem of selecting among different configurations of a single change and the no change hypothesis. The important operational issue of forecasting a future observation, often neglected in the literature on change-point analysis, is also treated in the previous model selection perspective. To illustrate the approach, a case study involving annual energy inflows for eight large hydropower systems situated in Québec is detailed. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Change-point analysis; Energy inflows; Bayesian model selection; Forecasting; Model uncertainty; Gibbs sampling

1. Introduction

It is argued in this paper that, when a change in meteorological time series at an unknown epoch is suspected, hydrologists must entertain a range of possible belief models, at least one of which assumes a given type of change and another that represents the no change hypothesis. When facing important decisions based on these data, decision makers should not put all their eggs in the same basket. In fact, any individual who considers only a single model for a decision problem, for instance forecasting future observations, ignores model uncertainty, which can be a major part of overall uncertainty about quantities of interest. As hydrological studies are mainly based on samples of limited sizes, this issue becomes very important for change-point analysis since uncertainty about the existence and the configuration of a change will always remain.

Following the emergence of published studies on climate changes, a number of hydrologists have used models which describe certain types of changes to represent hydrometeorological time series. Bayesian methods to study a single unknown change-point in a

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

PII: S0022-1694(00)00271-7

sequence of random variables have been increasingly used in hydrological literature. In most of these papers, a given type of change (often in the mean level) occurring with certainty is assumed, and focus has been put on the characterization of the change. Even if such inference is vital to help with decision making, it is not sufficient. First, conclusions stemming from this type of inference are entirely conditional upon the given configuration of change that the model assumes (for instance, in the mean only). This scientific reporting framework in turn forces the hydrologist to choose one situation among different possible models to perform the change-point analysis, and consequently to exclude all other kinds of changes. Second, these models proposed in the hydrological literature always assume a change with certainty. Therefore, the analysis of posterior distributions of parameters cannot be used, in addition to inference about the characteristics of a possible change, to "test" its existence. However, although the no change hypothesis does not have any status in this type of formulation, such questionable analysis can be found in published papers for change-point models, namely in Bruneau and Rassam (1983), Rao and Tirtotjondro (1996), and Lubès-Niel et al. (1998). In fact, these authors use posterior estimates from Bayesian analysis under such models to reject or accept the no change hypothesis. Entertaining several models in change-point analysis allows for a clear formulation of each possible situations, and therefore to proceed adequately when diagnosis about the different alternatives is needed.

In a companion paper, Perreault et al. (2000) presented a Bayesian approach for the inference about the parameters of a single change-point model. A sudden change in the mean level and in the variance in a sequence of normal random variables were investigated. This article provides a more general formulation of change-point analysis, which allows us to take into account model uncertainty when making decisions. Single change-point analysis is viewed as a model choice problem among the various possible situations that may occur. Bayesian model selection is adopted here under the perspective where the range of models under consideration is assumed to include the "true" belief model. Two different decision problems are examined within this framework: (1) the verification of the existence of a change and the identification of its type, i.e. only model choice; and (2) model choice followed by the prediction of a future observation. Although these decision problems cannot be viewed in water resources management as terminal actions, they are very important steps in the decision-making process. As in Perreault et al. (2000), the purpose of this paper is not to determine why changes occurred (e.g. climatic change or site-specific factors), but is only concerned with developing statistical tools to help decision making.

In Section 2, we recall the general formulation of a single change-point model and show how marginal posterior distributions of the parameters of interest can be evaluated analytically or, if necessary, via a straightforward iterative Markov Chain Monte Carlo method, namely the Gibbs sampler. In Section 3, we show how predictive distributions can be used to test whether a change has in fact occurred, to identify its type among several alternatives, and finally to derive a prediction that takes into account the no change hypothesis and all configurations of change. Section 4 is dedicated to the evaluation of the predictive densities when no expression in closed form can be obtained. The methods are based on posterior simulation outputs by Gibbs sampling. Finally, this new perspective for change-point analysis is illustrated for a sequence of independent normal random variables in Section 5. Four models are considered: the no change hypothesis, a single change in the mean level only, a single change in the variance only, and a simultaneous change in both the mean and the variance. An application to eight series of Hydro-Québec annual energy inflows is offered.

2. Inference for univariate single change-point models

The general formulation for a single change-point model assumes a sequence of n independent random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, for instance Hydro-Québec annual energy inflows, such that

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

where the density $p(x|\theta)$ belongs to a known class of probability densities indexed by a parameter θ taking values in the set Θ so that $\theta_1 \neq \theta_2$. The parameter $\tau = 1, 2, \ldots, n-1$ is the unknown change-point. Model (1) assumes that the time series exhibits an abrupt change. Indeed, in hydrometeorological time series, other types of changes such as trends can exist. 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). Hence, a sudden change may be representative for several hydrological and climatic time series. The independence assumption between successive values can also be questioned. 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, at least as a first methodological step. It will be relaxed in further studies.

The corresponding likelihood function for the realizations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is

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

In the Bayesian perspective, a joint prior distribution $p(\theta_1, \theta_2, \tau)$ is assumed for the parameters. Bayes theorem provides the joint posterior distribution $p(\theta_1, \theta_2, \tau|\mathbf{x})$ of θ_1, θ_2, τ given the data $\mathbf{x}$. This probability density function (p.d.f.) is proportional to

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

Evaluation of this joint density and all desired marginal posterior densities can be a very difficult task. But the Gibbs sampler enables a straightforward solution to such a problem, if conjugate prior distributions for fixed τ are assumed (see Carlin et al., 1992; Stephens, 1994; Perreault et al., 2000). Implementation requires sampling from the full conditional distributions $p(\tau|\theta_1, \theta_2, \mathbf{x})$, $p(\theta_1|\theta_2, \tau, \mathbf{x})$ and $p(\theta_2|\theta_1, \tau, \mathbf{x})$, which are all proportional to Eq. (3). Note that $p(\tau|\theta_1, \theta_2, \mathbf{x})$ is exactly of the form

$$p(\tau|\theta_1, \theta_2, \mathbf{x}) = \frac{p(\mathbf{x}|\theta_1, \theta_2, \tau)p(\tau)}{\sum_{\tau=1}^{n-1} p(\mathbf{x}|\theta_1, \theta_2, \tau)p(\tau)} \tag{4}$$

and can be easily sampled since it involves a discrete density.

As an illustration for inference about the parameters of change-point models, let us consider, as in most published approaches on change-point studies, a sequence of independent normal random variables $X_1, X_2, \ldots, X_n$. The normality assumption is appropriate as a first step to analyze a single change in 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. Even if the normal model is rather simple, it allows for various types of changes, which are interesting in hydrology. Considering only a single change as in Eq. (1) along with the no change hypothesis, some basic situations with the usual parametrization of the normal model are as follows:

- Model M_0:

$$X_i \sim \mathcal{N}(x_i|\mu, \sigma^2), \quad i = 1, \ldots, n$$

- Model M_1:

$$X_i \sim \mathcal{N}(x_i|\mu_1, \sigma^2), \quad i = 1, \ldots, \tau$$

$$X_i \sim \mathcal{N}(x_i|\mu_2, \sigma^2), \quad i = \tau+1, \ldots, n$$

- Model M_2:

$$X_i \sim \mathcal{N}(x_i|\mu, \sigma_1^2), \quad i = 1, \ldots, \tau$$

$$X_i \sim \mathcal{N}(x_i|\mu, \sigma_2^2), \quad i = \tau+1, \ldots, n$$

- Model M_3:

$$X_i \sim \mathcal{N}(x_i|\mu_1, \sigma_1^2), \quad i = 1, \ldots, \tau$$

$$X_i \sim \mathcal{N}(x_i|\mu_2, \sigma_2^2), \quad i = \tau + 1, \ldots, n$$

where $\mathcal{N}(x_i|\mu, \sigma^2)$ stands for the normal p.d.f. with parameters $\mu \in \Re$ and $\sigma \in \Re^+$. Assuming conjugate prior distributions and independence between τ and the other parameters, all posterior marginals can be obtained in closed form for models M_0, M_1 and M_3. Inference about the parameters can therefore be made directly to characterize each of these situations. Expressions for the posterior distributions for model M_0 are well-known results and can be found, for instance, in Berger (1985). Those corresponding to the change-point model M_1 are given in Perreault et al. (2000). The general idea is to first evaluate the joint and marginal posterior distributions of μ_1, μ_2 and σ^2 assuming the change-point τ is known. Then, the unconditional posterior densities are deduced by averaging these distributions over τ. The posterior marginals appear to be finite mixtures of Student t-distributions conditional upon τ, each weighted by the corresponding posterior mass $p(\tau|\mathbf{x})$. This last discrete density can be deduced from the Bayes theorem denominator given τ, which is easily evaluated following the conjugacy properties. In a similar way, all desired posterior marginals can be derived analytically for the more general situation of a single change in both the mean and the variance, i.e. model M_3. In our opinion, this model is important for hydrometeorological series, since a shift in the mean level often seems to come with a change in variability. In Appendix A, the posterior distributions of the parameters of interest, assuming this type of change, are derived using conjugate priors. Results for M_0 and M_1 assuming conjugate priors are also recalled in Appendix A. Finally, a single change only in the variance (model M_2) does not lead to solutions in closed form as for M_0, M_1 and M_3. However, the Gibbs sampler provides an elegant and convenient answer to this problem (details can be found in Perreault et al., 2000).

Up to this point, exactly one change of a given type has been assumed, i.e. it is known that a change of this configuration has occurred. Therefore, the no change hypothesis has no status in that formulation nor any other type of change, and inference about the existence of a change or about its configuration cannot satisfactorily be made based only on a single model. For example, if only M_1 is considered, examination of the posterior density of $\delta = \mu_2 - \mu_1$ to assess the no change hypothesis is not valid. Bayesian perspective adopted in this paper, has little role for the non-Bayesian concept of hypothesis test, especially where these relate to point null hypotheses such as $H_0 : \delta = \mu_2 - \mu_1 = 0$ (no change). Inferences about a specific value δ under a given hypothesis (change) in the absence of a well-specified alternative (as in classical statistics) cannot be made. In order for a Bayesian analysis to yield a non-zero probability for a point null hypothesis, it must begin with a non-zero prior probability for that hypothesis. However, δ a continuous parameter and the probability for an event $\{\delta = 0\}$ is zero. Therefore, actions such as “accept or reject no change”, i.e. M_0, based on the examination of a credible interval about δ under M_1 is not justified. That is why we cannot test the existence of a change only on the basis of a change-point model, and must consider separately the alternative M_0 in order that the no change situation becomes a possible outcome. A well-known simple example that illustrates the problem is the so-called experiment on used electronic and quantum-mechanical random event generators with visual feedback. This design can be modeled by a Binomial experiment. Suppose a Bernoulli trial where $\theta = \Pr(\text{success})$, $n = 104\ 490\ 000$ trials and X is defined as the number of successes. Therefore, X follows a $\text{Bin}(n,\theta)$. Let the realization of X be $x = 52\ 263\ 446$. We want to test $H_0 : \theta = 1/2$ against $H_1 : \theta \neq 1/2$. With uniform prior $p(\theta)$, we obtain as a 95% credible interval $\theta \in (0.50008,\ 0.50027)$ under H_1. On the other hand, if we consider model H_0 and assume $\Pr(H_0) = \Pr(H_1) = 1/2$, we have $\Pr(H_0|x) = 0.94$, which highly favor H_0. Clearly, one cannot do Bayesian tests using posterior distribution under a given hypothesis without considering an alternative. Therefore, the approach should be extended by including in the analysis the possibility of no change, and all other types of change likely to occur over the period of observation. Bayesian model selection is a solution to such a problem, since it allows for the consideration of a range of possible models in the inference process.

3. Entertaining a range of possible models for change-point analysis

In this section, Bayesian model selection is presented in the context of change-point analysis. Two different decision problems are examined within this perspective: the verification of the existence of a change and the identification of its type, if any, and the prediction of a future observation. But first, predictive analysis, on which model selection and forecasting are based, is briefly recalled. In the rest of the paper, since a range of possible models is now considered, the model indicator M_k is introduced in the list of unknown parameters.

3.1. Definitions

The general problem of statistical prediction may be described as that of inferring the values of unknown observable variables, say $\mathbf{z}$, from the current state of belief denoted H. In the Bayesian perspective, this is done through the predictive distribution, which is obtained by integrating the parameters ψ out of the joint density $p(\mathbf{z}, \psi|H)$ conditional upon the actual state of belief H:

$$p(\mathbf{z}|H) = \int p(\mathbf{z}|\psi, H)p(\psi|H)\,\mathrm{d}\psi \tag{5}$$

If the change-point model (1), denoted M_k, is assumed before any data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are considered, then $\mathbf{z} = \mathbf{x}$, $\psi = (\theta_1, \theta_2, \tau)$ and $H = M_k$. Therefore, assuming independence between τ and (θ_1, θ_2), the predictive distribution of the unknown but observable $\mathbf{x}$ is

$$\begin{aligned} p(\mathbf{x}|M_k) &= \sum_{\tau=1}^{n-1} \int\int p(\mathbf{x}|\theta_1, \theta_2, \tau, M_k) \\ &\quad\times p(\theta_1, \theta_2, \tau|M_k)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2 \\ &= \sum_{\tau=1}^{n-1} p(\tau|M_k)p(\mathbf{x}|\tau, M_k) \end{aligned} \tag{6}$$

This density is often called the marginal distribution of $\mathbf{x}$, but a more informative name is the prior predictive distribution. That is, this density is the probability of seeing the data, which were actually observed assuming a given change-point model, calculated before any data became available. This predictive density, implied by the likelihood and the prior distribution, provides a basis for assessing the compatibility of the data with our prior state of belief, i.e. model M_k.

After the data $\mathbf{x}$, assumed to be generated by model M_k, have been observed, we can predict an unknown future value y from the same process. In this case, $\mathbf{z} = y$, $\psi = (\theta_1, \theta_2, \tau)$ and $H = (\mathbf{x}, M_k)$. Using elementary probability rules (Berger, 1985), this new predictive distribution $p(y|\mathbf{x}, M_k)$ appears as finite mixtures of the predictive densities for fixed τ, weighted by the appropriate posterior probability of a shift occurring at that point:

$$\begin{aligned} p(y|\mathbf{x}, M_k) &= \sum_{\tau=1}^{n-1} \int\int p(y|\theta_1, \theta_2, \tau, M_k) \\ &\quad\times p(\theta_1, \theta_2, \tau|\mathbf{x}, M_k)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2 \\ &= \sum_{\tau=1}^{n-1} p(\tau|\mathbf{x}, M_k) \int\int p(y|\theta_2, M_k)p \\ &\quad\times(\theta_1, \theta_2|\tau, \mathbf{x}, M_k)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2 \\ &= \sum_{\tau=1}^{n-1} p(\tau|\mathbf{x}, M_k)p(y|\tau, \mathbf{x}, M_k) \end{aligned} \tag{7}$$

This distribution is called the posterior predictive distribution, posterior because it is conditional on the observed $\mathbf{x}$ and predictive because it is a prediction for an observable.

Note, in some cases the integrals in Eqs. (6) and (7) can be evaluated analytically. More often, these integrals are intractable and thus must be computed by numerical methods. In this paper, we favor an approach based on Gibbs sampling that is straightforward to implement, and easily accessible to the average statistical practitioner without numerical analysis skills (see Section 4).

3.2. Comparison of change-point models

Bayesian model selection consists of evaluating posterior probabilities for each model being true among a discrete set denoted $\mathscr{M}$. Let us assume that the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ have arisen from one of $q+1$ possible models in $\mathscr{M} = \{M_0, M_1, \ldots, M_k, \ldots, M_q\}$, where, as in Section 2, M_0 denotes the model which assumes that no change has occurred, while $M_1, \ldots, M_k, \ldots, M_q$ are change-point models. First, the Bayesian model selection proceeds by selecting prior probabilities $p(M_k)$ of each model being true. Often, equal prior probabilities are used, i.e. $p(M_k) = 1/(q+1)$. In the context of change-point analyses for water resources management, this choice based on a full symmetry of all situations can be criticized. In fact, the no change hypothesis will often represent the status quo situation, which many decision makers favor. Moreover, the remaining hypotheses will generally induce changes in policies that could be drastic, and that may involve a lot of money. Therefore, if no prior information is available about the existence of a change in the series of interest, the decision maker will be tempted to affect at least half of the weight to the no change hypothesis. Here, a more sensible choice would give equal weight to the "change" and "no change" alternatives:

$$p(M_0) = 1/2 \text{ and } p(M_1) = \cdots = p(M_q) = 1/(2q) \tag{8}$$

After the prior probabilities have been chosen, the analysis then proceeds by computing the posterior probabilities of each model being true. Using Bayes theorem, we have

$$p(M_k|\mathbf{x}) = \frac{p(\mathbf{x}|M_k)p(M_k)}{\sum_{j=0}^{q} p(\mathbf{x}|M_j)p(M_j)} \tag{9}$$

where $p(\mathbf{x}|M_k)$ is the prior predictive density of $\mathbf{x}$ under model M_k given in Eq. (6).

Let us consider only the problem involving the verification of the existence of a change and, if any, the identification of its type. This problem is essentially that of choosing a model in $\mathscr{M}$, without any subsequent decision, when the "state of the world" of interest is defined to be the "true" model. To proceed we first assume that the range of models under consideration includes the "true" belief model. This is the so-called $\mathscr{M}$-closed perspective described in Bernardo and Smith (1994). This assumption may be difficult to accept in a literal sense. However, for a change-point analysis, this hypothesis might be considered reasonable. In fact, often in hydrology, a given model, for instance model M_0, has been extensively adopted and found to be a successful predictive device in a wide spectrum of applications. Now, suppose a hydrologist suspects that a certain type of change, for example M_1, may have occurred, and in this new context wants to incorporate uncertainty about M_0. Provided the hydrologist is comfortable with assigning prior weights to the alternative formulations, the $\mathscr{M}$-closed perspective can be exploited since M_0 is generally viewed as a proxy to the true model. If the natural zero-one utility function is used in the $\mathscr{M}$-closed setting, maximizing expected utility implies that the optimal model choice is the one with largest posterior probability (9) to be true. Therefore, these posterior probabilities can be used in a formal way to verify if a change in a hydrometeorological series is a plausible hypothesis, and if necessary what type of shift may have occurred.

For the particular situation where only two models, M_j and M_k, are to be compared, the above discussion suggests that one should use the posterior odds ratio

$$\frac{p(M_j|\mathbf{x})}{p(M_k|\mathbf{x})} = \frac{p(\mathbf{x}|M_j)p(M_j)}{p(\mathbf{x}|M_k)p(M_k)} \tag{10}$$

Expression (10) reveals the key role of the prior predictive density in providing the way in which the data update the relative prior beliefs into the relative posterior belief about the true model. This leads naturally to the Bayes factor B_{jk}, defined as the ratio of posterior to prior odds on M_j against M_k,

$$B_{jk} = \frac{p(M_j|\mathbf{x})}{p(M_k|\mathbf{x})} \bigg/ \frac{p(M_j)}{p(M_k)} = \frac{p(\mathbf{x}|M_j)}{p(\mathbf{x}|M_k)}, \tag{11}$$

which can be interpreted as a measure of whether the data $\mathbf{x}$ have increased or decreased the odds on M_j relative to M_k. The Bayes factor given in Eq. (11) is similar enough to the likelihood ratio statistic, but the parameters are eliminated by integration rather than by maximization. Naturally, $B_{jk} > 1$ signifies that M_j is now more relatively plausible than M_k in light of $\mathbf{x}$; $B_{jk} < 1$ signifies that the relative plausibility of M_j has decreased. However, in many problems, such criteria for interpreting the Bayes factors can be very loose. Based on the work of Jeffreys (1961), Raftery (1996) suggested more conservative guidelines for

Table 1
Raftery's scale for interpreting the Bayes factor

$2 \ln B_{jk}$	B_{jk}	Evidence for M_j
0–2	1–3	Not worth more than a bare mention
2–6	3–20	Positive
6–10	20–150	Strong
>10	>150	Very strong

interpreting the Bayes factor, which are on the same scale as the familiar deviance and likelihood ratio test statistics. The guidelines are reproduced in Table 1.

Many questions can be answered using Bayes factors (or equivalently the posterior probability of each model) along with Table 1. For instance, if one is interested in how much the data support a given model M_k compared to its competitor alternatives in $\mathscr{M}$, the appropriate Bayes factor, denoted $B_{k(-k)}$, is given by

$$B_{k(-k)} = \left(\frac{p(M_k|\mathbf{x})}{1-p(M_k|\mathbf{x})}\right)\Big/\left(\frac{p(M_k)}{1-p(M_k)}\right) = \left[\sum_{j=0(\neq k)}^{q} B_{jk}\left(\frac{p(M_j)}{1-p(M_k)}\right)\right]^{-1} \quad (12)$$

An important particular case is the overall assessment of change versus no change. Since $B_{jk} = B_{kj}^{-1}$, we note that testing the existence of a change in a hydrometeorological series can be based on

$$B_{c0} = B_{(-0)0} = \left(\frac{1-p(M_0|\mathbf{x})}{p(M_0|\mathbf{x})}\right)\Big/\left(\frac{1-p(M_0)}{p(M_0)}\right) = \sum_{j=1}^{q} B_{j0}\left(\frac{p(M_j)}{1-p(M_0)}\right) \quad (13)$$

The right-hand side of Eq. (13) has the form of a weighted average of individual Bayes factors for specific configurations of change against no change. If, as above, equal prior probabilities are specified for the q change-point models, and equal prior weights are assigned to the change and no change alternatives, the Bayes factor B_{c0} for change against no change is just an average of the Bayes factors B_{j0}, taken over all considered change-points model. Finally, if we assume that a change has occurred, identification of its type may be carried out by evaluating the Bayes factor (12) for each alternative, excluding the no change model M_0. According to the scheme in Table 1, its type is identified after examining if one of these Bayes factor provides enough evidence in favor of a given change-point model compared to its competitors.

It is important to notice that the use of Bayes factors, and therefore of Eq. (9), needs proper informative prior distributions for all parameters to be specified. When prior distributions are improper, the prior predictive distributions $p(\mathbf{x}|M_j)$ and $p(\mathbf{x}|M_k)$ are not defined. However, this rather embarrassing problem can be overcome by using default Bayes factors, such as the "Intrinsic Bayes factor" (Berger and Pericchi, 1996) and the "Fractional Bayes factor" (O'Hagan, 1995). In this paper, we do not address model selection with non-informative priors since we assume that some prior hydrologic expertise is available, however vague it may be.

3.3. Accounting for model uncertainty when forecasting

Suppose we are interested in a decision problem that involves not only model choice but also a subsequent action, such as the prediction of a future value y. Following the discussion in the previous section, if one wishes to consider a single model, the one with largest posterior probability to be true, denoted by M_k^*, will be selected for prediction. If we consider a quadratic utility function, maximizing expected utility leads to the optimal prediction $\tilde{y}_k^*$ under M_k^* given by

$$\tilde{y}_k^* = \int y p(y|\mathbf{x}, M_k^*)\, \mathrm{d}y = E\{y|\mathbf{x}, M_k^*\} \quad (14)$$

where $p(y|\mathbf{x}, M_k^*)$ is the posterior predictive density (7). This prediction depends upon a given model M_k^*. Using it forces the hydrologist to select a single model and to ignore model uncertainty, which can be a major part of overall uncertainty about future values. The more general Bayesian model selection framework, presented in the previous section, allows for the evaluation of an "overall" posterior predictive distribution, which takes into account all models in $\mathscr{M}$. In fact, since under the $\mathscr{M}$-closed perspective each model has a posterior probability $p(M_k|\mathbf{x})$, one can maintain consideration of several models by weighting each conditional posterior predictive density $p(y|\mathbf{x}, M_k)$ by $p(M_k|\mathbf{x})$:

$$p(y|\mathbf{x}) = \sum_{k=0}^{q} p(y|\mathbf{x}, M_k)p(M_k|\mathbf{x}) \quad (15)$$

Assuming again a quadratic utility function, it is seen that the optimal overall prediction $\tilde{y}$ with respect to expected utility is

$$\tilde{y} = \int yp(y|\mathbf{x})\,\mathrm{d}y = \sum_{k=0}^{q} E\{y|\mathbf{x}, M_k\}p(M_k|\mathbf{x})$$

$$= \sum_{k=0}^{q} \tilde{y}_k p(M_k|\mathbf{x}) \quad (16)$$

Eq. (15) shows, in particular, that selecting a single model and proceeding conditionally on it may be reasonable if one of the posterior probabilities $p(M_k|\mathbf{x})$ is close to unity. If not, then prediction conditional on a single model fails to take into account all of the uncertainty about the configuration of the change, and the precision of forecasts may be overestimated.

4. Evaluations of predictive densities

As mentioned in Section 3.1, the prior and posterior predictive distributions, $p(\mathbf{x}|M_k)$ and $p(y|\mathbf{x}, M_k)$, may not be expressible in closed form, and recourse to numerical approximations is needed. Since estimation of the parameters for model (1) is done via Gibbs sampling (Section 2), it would be natural to use the simulated outputs to estimate the predictive densities. The problem is easily solved by Monte Carlo generation for posterior predictive distribution. On the other hand, the evaluation of the prior predictive distribution from Gibbs outputs is more challenging. In the following, we present how the Gibbs algorithm can be used to evaluate such integrals. The parameters of model (1) are denoted by $\boldsymbol{\theta} = (\theta_1, \theta_2, \tau)$, and we suppose $\{\boldsymbol{\theta}^{(j)};\ j = 1,\dots,m\}$ are the m draws from the joint posterior density $p(\boldsymbol{\theta}|\mathbf{x}, M_k)$ obtained using the Gibbs sampler, after the first t values of the chain have been discarded.

4.1. Posterior predictive density

The posterior predictive density can be readily approximated via the Gibbs sampling process used to estimate the posterior distributions. For each draw of the parameters $\{\boldsymbol{\theta}^{(j)};\ j = 1,\dots,m\}$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x}, M_k)$, one simply has to sample one value of y from the likelihood function $p(y|\boldsymbol{\theta}^{(j)}, M_k)$, which for model (1) is $p(y|\theta_2^{(j)}, M_k)$. The set of simulated y from all the $\boldsymbol{\theta}^{(j)}$ characterize the posterior predictive density, $p(y|\mathbf{x}, M_k)$, and can be used to estimate it. For instance, as a rough estimate for $p(y|\mathbf{x}, M_k)$, one can use a histogram. A better estimate is given by

$$\hat{p}(y|\mathbf{x}, M_k) = \frac{1}{m}\sum_{j=1}^{m} p(y|\theta_2^{(j)}, M_k) \quad (17)$$

Gelfand and Smith (1990) referred to this averaging technique as Rao–Blackwellization and argued that it improves on the usual histogram estimate. This approach was adopted in Perreault et al. (2000) for inference about the parameters of model M_2.

If one is concerned with model uncertainty, an overall posterior predictive distribution can be estimated using the posterior probabilities associated with each model. If, for all models, prior and posterior predictive distributions are intractable, we have

$$\hat{p}(y|\mathbf{x}) = \sum_{k=0}^{q} \hat{p}(M_k|\mathbf{x})\hat{p}(y|\mathbf{x}, M_k), \quad (18)$$

where $\hat{p}(y|\mathbf{x}, M_k)$ is given by Eq. (17) and $\hat{p}(M_k|\mathbf{x})$ is the estimated posterior probability of M_k being true obtained by substituting in Eq. (9) the prior predictive by an estimated value (see Section 4.2). Finally, conditional and overall point predictions for a future value, assuming a quadratic utility function, can be evaluated in a way similar to that in Section 3.3.

4.2. Prior predictive density

As it was mentioned, the evaluation of the prior predictive distribution $p(\mathbf{x}|M_k)$ from Gibbs outputs is a more complicated task. In fact, integration is done with respect to the prior density, whereas the Gibbs sampler produces the draws from the posterior. Therefore, the simulated outputs from the Gibbs method, or other Markovian iterative schemes, cannot be used directly to evaluate prior predictive distributions as is the case for posterior predictive densities.

Several approaches, based on such posterior

Step1 **Initial Gibbs cycle**

$$\theta_1^{(j)} \sim p\left(\theta_1 \left| \theta_2^{(j-1)}, \tau^{(j-1)}, \mathbf{x}, M_k\right.\right)$$
$$\theta_2^{(j)} \sim p\left(\theta_2 \left| \theta_1^{(j)}, \tau^{(j-1)}, \mathbf{x}, M_k\right.\right)$$
$$\tau^{(j)} \sim p\left(\tau \left| \theta_1^{(j)}, \theta_2^{(j)}, \mathbf{x}, M_k\right.\right)$$

$\Longrightarrow$ $\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*, \tau^*)$; $p(\tau^* | \theta_1^*, \theta_2^*, \mathbf{x}, M_k)$; $\hat{p}(\theta_1^* | \mathbf{x}, M_k)$ eq. (B.3)

$\swarrow$

Step 2 **Partial Gibbs cycle**

$$\theta_2^{(j)} \sim p\left(\theta_2 \left| \theta_1^*, \tau^{(j-1)}, \mathbf{x}, M_k\right.\right)$$
$$\tau^{(j)} \sim p\left(\tau \left| \theta_1^*, \theta_2^{(j)}, \mathbf{x}, M_k\right.\right)$$

$\Longrightarrow$ $\hat{p}(\theta_2^* | \theta_1^*, \mathbf{x}, M_k)$ eq. (B.4)

$\swarrow$

Step 3 **Chib's estimate**

$\Downarrow$

$\hat{p}_{CH}(\mathbf{x} | M_k)$

eq. (B5)

Fig. 1. Basic steps for Chib's estimate.

simulation, have been proposed to estimate prior predictive densities like $p(\mathbf{x}|M_k)$. However, many of these estimates are known to be unstable, or need large samples (see Kass and Raftery, 1995). Other more accurate estimates requires a tuning function, which can be quite difficult to determine (see Gelfand and Dey, 1994). On the other hand, Chib (1995) developed a very simple approach that is free of the problems just mentioned. This method, which is adopted here, was developed in the setting where the Gibbs sampling algorithm has been used to provide sample draws from the posterior distribution. To use it, it is necessary that all normalizing constants of the full conditional densities involved in the Gibbs sampling be known. This is usually the case when the Gibbs algorithm is implemented.

Chib (1995) exploits the fact that the prior predictive density is the normalizing constant in the Bayes theorem, and therefore can be expressed as the prior distribution times the likelihood function over the posterior density, i.e.

$$p(\mathbf{x}|M_k) = \frac{p(\mathbf{x}|\boldsymbol{\theta}, M_k)p(\boldsymbol{\theta}|M_k)}{p(\boldsymbol{\theta}|\mathbf{x}, M_k)} \tag{19}$$

This simple identity holds for any parameter value. The numerator in the last expression can be directly evaluated at a given point, say $\boldsymbol{\theta}^*$. However, no expression in closed form is available for the denominator $p(\boldsymbol{\theta}|\mathbf{x}, M_k)$. Therefore, if a density estimate for $p(\boldsymbol{\theta}|\mathbf{x}, M_k)$, denoted $\hat{p}(\boldsymbol{\theta}|\mathbf{x}, M_k)$, is available, an estimate for the prior predictive distribution using expression (19) can be derived as

$$\hat{p}_{\mathrm{CH}}(\mathbf{x}|M_k) = \frac{p(\mathbf{x}|\boldsymbol{\theta}^*, M_k)p(\boldsymbol{\theta}^*|M_k)}{\hat{p}(\boldsymbol{\theta}|\mathbf{x}, M_k)}, \tag{20}$$

where $\boldsymbol{\theta}^*$ is any parameter value. The simple solution proposed by Chib (1995) to evaluate $\hat{p}(\boldsymbol{\theta}|\mathbf{x}, M_k)$ is to perform additional Gibbs cycles. A brief description of the implementation for our particular problem is offered in Appendix B. A scheme summarizing the approach is presented in Fig. 1. See Chib (1995) for more general discussions and properties.

Although Chib's procedure leads to an increase in the number of iterations, it does not require new programming and thus is straightforward to implement.

5. Applications assuming normal change-point models

This new perspective for change-point analysis is illustrated by applying the methods presented in the previous sections to eight series of annual energy inflows. The four normal models discussed in Section

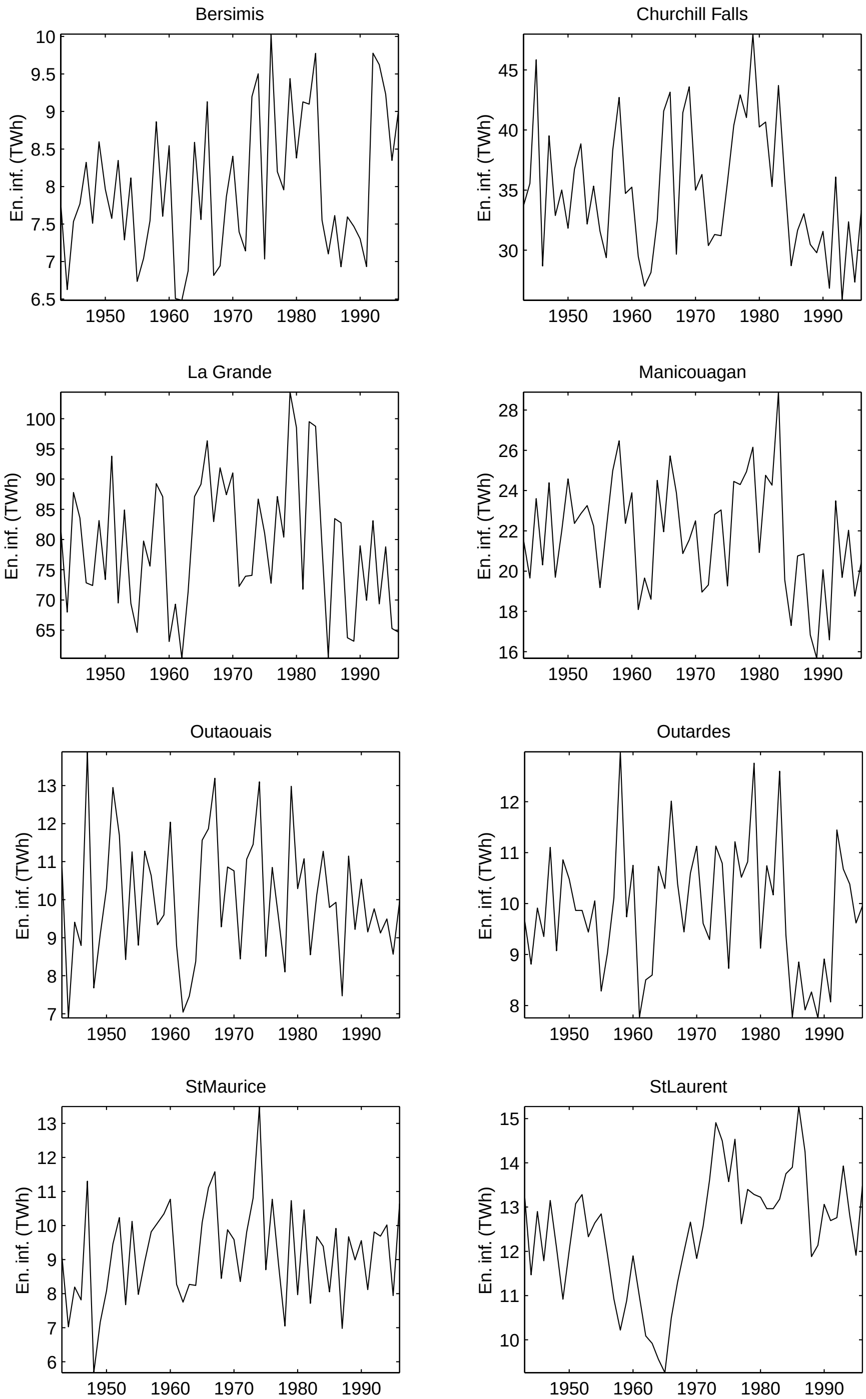

Fig. 2. Annual energy inflows for eight hydropower systems in TWh (1943–1996).

2 are considered: the no change hypothesis, a single change in the mean level only, a single change in the variance only, and a simultaneous change in both the mean and the variance. Perreault et al. (2000) have already applied models M_1 and M_2 to annual energy inflows for two sites. However, their analysis was restricted to the estimation of the change-point and its intensity, and did not account for model uncertainty.

All calculations that require Gibbs sampling (evaluation of prior and posterior predictive distributions for model M_2) were based on $t = 1000$ iterations and $m = 1000$ replicates. To ensure maximum accuracy of Chib's approach, the prior predictive estimate was evaluated at a high density point $\boldsymbol{\theta}^*$, namely the posterior mode.

5.1. The data

The series of particular importance for energy planning in Québec are the annual energy inflows for the eight major hydropower systems managed by Hydro-Québec: St-Laurent, Outaouais, La Grande, St-Maurice, Bersimis, Manicouagan, Outardes and Churchill Falls. For each of these systems, the energy inflows are evaluated by multiplying the natural inflow of each reservoir in the system by a factor based on the production capacity of the corresponding power plant. The historical annual energy inflows expressed in terawatt-hour (TWh) for all hydropower systems (1943–1996) are shown in Fig. 2.

Examining these time series, an abrupt change in the mean level and/or the variance for some hydropower systems may be suspected. However, statistical characteristics for these series, such as the mean and the variance, are calculated over the entire period of observation, and then used as inputs to design scenarios or for forecasting future energy availability. Stationarity is therefore assumed, even when in fact this hypothesis could be violated. This might lead to wrong decisions regarding policies about future energy-related development for the company. Therefore, in order to develop management rules for energy planning, procedures for analyzing changes in energy inflow series are needed.

5.2. Specifying prior distributions

Bayesian model selection using standard Bayes factors cannot be made using non-informative prior distributions. Prior knowledge must thus be elicited from expert opinion and/or from external information about the studied phenomena in order to obtain proper prior distributions. In what follows, we briefly describe how proper prior distributions were specified for our particular problem.

As mentioned in the previous section, the annual energy inflows are generally used by Hydro-Québec for energy planning, under the assumption that all historical observations were obtained under the same conditions. Model M_0, which represents the no change hypothesis, can then be considered as a proxy to the true model in a prior state of belief. Therefore, following the discussion in Section 3.2, prior probabilities for each model being true were chosen such that change and no change alternatives have equal weight, i.e.

$$p(M_0) = 1/2$$

and

$$p(M_1) = p(M_2) = p(M_3) = 1/6.$$

Faced with such vague knowledge and lack of conviction about the existence of a change, one cannot reasonably favor any epoch of change for models M_1, M_2 and M_3. The year τ at which a change may have occurred was therefore assumed to follow a discrete uniform distribution for each of these models. Note that with this density, the prior expected change-point is 1970, i.e. the mean of a discrete uniform probability distribution on the interval [1943, 1995].

Finally, for the other parameters, the prior degrees of belief were assumed to be represented by normal-inverted gamma type of distributions (see Appendix A). The complete specification of prior knowledge about these quantities required the choice of hyperparameters in these expressions. This was done by eliciting regional information, as described in Perreault et al. (2000). For each site, a regression model was used to predict the average and the variance of the energy inflows, before and after 1970. These predictions were then used, along with their standard error, to estimate the mean and the variance of the normal and inverted gamma distributions. Finally, simple systems of equations for the first two moments were solved, yielding estimated hyperparameter values for all sites. A simple regression

Table 2
Bayes factors and model posterior probabilities for all series

	B_{jk}	M_0	M_1	M_2	M_3	$p(M_k\|\mathbf{x})$
Bersimis	M_0	1	0.3865	0.2642	0.2364	0.2205
	M_1	2.5870	1	0.6835	0.6115	0.1902
	M_2	3.7850	1.4631	1	0.8947	0.2783
	M_3	4.2307	1.6354	1.1177	1	0.3110
Churchill Falls	M_0	1	0.1659	0.7848	0.0929	0.1424
	M_1	6.0284	1	4.7313	0.5603	0.2862
	M_2	1.2742	0.2114	1	0.1184	0.0605
	M_3	10.7591	1.7847	8.4441	1	0.5108
La Grande	M_0	1	0.7049	0.6445	0.9349	0.4261
	M_1	1.4187	1	0.9143	1.3264	0.2015
	M_2	1.5517	1.0937	1	1.4507	0.2204
	M_3	1.0696	0.7539	0.6893	1	0.1519
Manicouagan	M_0	1	3.3276	0.4086	1.5035	0.4678
	M_1	0.3005	1	0.1228	0.4518	0.0469
	M_2	2.4475	8.1445	1	3.6800	0.3816
	M_3	0.6651	2.2132	0.2717	1	0.1037
Outaouais	M_0	1	3.0306	0.2644	0.7530	0.3554
	M_1	0.3300	1	0.0872	0.2485	0.0391
	M_2	3.7827	11.4638	1	2.8482	0.4481
	M_3	1.3281	4.0249	0.3511	1	0.1573
Outardes	M_0	1	1.7533	0.1761	1.1560	0.2966
	M_1	0.5703	1	0.1004	0.6593	0.0564
	M_2	5.6798	9.9585	1	6.5660	0.5615
	M_3	0.8650	1.5167	0.1523	1	0.0855
St-Maurice	M_0	1	1.1804	0.7005	1.7396	0.5129
	M_1	0.8472	1	0.5934	1.4737	0.1448
	M_2	1.4276	1.6852	1	2.4835	0.2441
	M_3	0.5748	0.6785	0.4027	1	0.0983
St-Laurent	M_0	1	<0.0001	0.0727	0.0001	0.0001
	M_1	>150	1	>150	1.6854	0.6274
	M_2	13.7496	0.0005	1	0.0009	0.0003
	M_3	>150	0.5933	>150	1	0.3722

considering only the generating capacity appeared to be the best regional model for predicting the average and the variance of the energy inflows for each hydropower system. For details, see Perreault et al. (2000).

5.3. Inference about the existence and the configuration of a change

We are now concerned with the problem of verifying the existence of a change in the annual energy inflow and, if any, the identification of its type among the configurations assumed by M_1, M_2 and M_3. As explained in Section 3.2, this is essentially the problem of choosing a model in $\mathcal{M} = \{M_0, M_1, M_2, M_3\}$, without any subsequent decision. To perform a Bayesian model selection in $\mathcal{M}$, each of the four prior predictive densities, $p(\mathbf{x}|M_k)$ $i = 0, \ldots, 3$, has to be evaluated. This is easily done for models M_0, M_1 and M_3. The density $p(\mathbf{x}|M_0)$, assuming conjugate prior distributions, is a well-known result and is expressible in closed form (Berger, 1985). The prior predictive distributions $p(\mathbf{x}|M_1)$ and $p(\mathbf{x}|M_3)$ can be calculated using Eq. (6). Since $p(\mathbf{x}|\tau, M_1)$ and $p(\mathbf{x}|\tau, M_3)$ appearing in Eq. (6) can be expressed in

Table 3
Bayes factors for change against no change and for identifying the type of the change, if any

System	B_{c0}	$B_{1(2,3)}$	$B_{2(1,3)}$	$B_{3(1,2)}$
Bersimis	3.53	0.65	1.11	1.33
Churchill Falls	6.02	1.00	0.16	2.95
La Grande	1.35	–	–	–
Manicouagan	1.14	–	–	–
Outaouais	1.81	–	–	–
Outardes	2.37	0.10	14.25	0.16
St-Maurice	0.95	–	–	–
St-Laurent	>150	3.37	0.00	1.19

closed form (see Appendix A), the evaluation of the prior predictive densities for models M_1 and M_3 is straightforward. More precisely, we have

$$p(\mathbf{x}|M_0) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{1+n\lambda}} \frac{\beta^{\alpha}}{(\beta')^{\alpha'}} \frac{\Gamma(\alpha')}{\Gamma(\alpha)} \tag{21}$$

$$p(\mathbf{x}|M_1) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{\beta^{\alpha}}{\sqrt{\lambda_1\lambda_2}\Gamma(\alpha)} \sum_{\tau=1}^{n-1} p(\tau|M_1) \times(\lambda_1'\lambda_2')^{1/2} \frac{\Gamma(\alpha')}{(\beta')^{\alpha'}} \tag{22}$$

$$p(\mathbf{x}|M_3) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{\beta_1^{\alpha_1}\beta_2^{\alpha_2}}{\sqrt{\lambda_1\lambda_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} \sum_{\tau=1}^{n-1} p(\tau|M_3) \times(\lambda_1'\lambda_2')^{1/2} \frac{\Gamma(\alpha_1')\Gamma(\alpha_2')}{(\beta_1')^{\alpha_1'}(\beta_2')^{\alpha_2'}} \tag{23}$$

where the updated hyperparameters for these models are given in Appendix A. To simplify, the same notation for the hyperparameters of each model was used. But, to stay rigorous, different notations should have been used to express the fact that hyperparameter values depend upon their specific model.

Now, since informative prior distributions were specified for each parameter, the Bayes factors B_{10}, B_{30} and B_{31} and the model posterior probabilities $p(M_k|\mathbf{x})$, $k = 0, 1, 3$, can be evaluated analytically using Eqs. (11) and (9) with the above expressions. However, the integral needed to evaluate $p(\mathbf{x}|M_2)$ is intractable (Perreault et al., 2000). Therefore, the Bayes factors B_{20}, B_{21} and B_{32} cannot be evaluated analytically. This problem can be solved via Gibbs sampling by using Chib's method as presented in Section 4.2. The full conditional distributions needed for Chib's approach were derived in Perreault et al. (2000), and are recalled in Appendix A.

All Bayes factors for each hydropower system are presented in Table 3. These can be used to compare, two by two, all models. The posterior probabilities (9) of each model being true have also been evaluated, and are given in the last column of Table 2. They can be compared to the prior probabilities specified in Section 5.2 in order to see how the prior state of belief was updated by the data.

It can be seen in particular that the odds for model M_0 have decreased for every series except St-Maurice. Therefore, the data tend to contradict Hydro-Québec's prior belief that the annual energy inflows were generated from the same process. We can be particularly confident about the existence of a single change for the St-Laurent series, since the posterior probability for M_0 is very small (0.0001). However, for the other systems, it may be appropriate to use the Bayes factor B_{c0} given by Eq. (13) along with a decision rule such as Raftery's scheme (Table 1). These Bayes factors are reported in Table 3 for each series.

According to the scheme of Table 1, the Bayes factors B_{c0} for Bersimis and Churchill Falls provide positive evidence in favor of a single change, while for St-Laurent, as it was anticipated, very strong evidence against no change is provided. Note that the annual energy inflows of Outardes almost exhibit a change ($B_{c0} = 2.37$). For the other systems, it is not possible to reject the no change hypothesis. Since for Bersimis, Churchill Falls, Outardes and St-Laurent the change hypothesis is plausible, Bayes factors to identify its type were calculated as described at the end of Section 3.2. Discarding model M_0 as a credible hypothesis, Eq. (12) was used to evaluate $B_{1(2,3)}$, $B_{2(1,3)}$ and $B_{3(1,2)}$. The Bayes factor $B_{1(2,3)}$ corresponds to the ratio of posterior to prior odds on M_1 against M_2 and M_3, and so on. These pairwise comparisons of plausibility measures may be used to discriminate between the considered change-point models, under the hypothesis that one of these configurations has occurred. It seems that models M_2 and M_3 are equally credible for Bersimis, while for Outardes M_2 is clearly the most plausible configuration of change, if any. The annual energy inflows observed for Churchill Falls are more likely to exhibit a simultaneous change in both

Table 4
Posterior predictive densities conditional on τ (note: $\mathcal{ST}(x|a,b,c) \propto [1 + c^{-1}b(x-a)^2]^{-(c+1)/2}$)

Density	Parameters of the Student *t*-distributions		
	Location *a*	Scale *b*	Degrees of freedom *c*
$p(y\|\mathbf{x}, M_0)$	ϕ'	$\alpha'[(1+\lambda')\beta']^{-1}$	$2\alpha'$
$p(y\|\tau, \mathbf{x}, M_1)$	ϕ'_2	$\alpha'[(1+\lambda'_2)\beta']^{-1}$	$2\alpha'$
$p(y\|\tau, \mathbf{x}, M_3)$	ϕ'_2	$\alpha'_2[(1+\lambda'_2)\beta'_2]^{-1}$	$2\alpha'_2$

the mean and the variance (M_3). Finally, it seems that the most plausible type of change for annual energy inflows for St-Laurent is a single change in the mean only (M_1).

Inference about the parameters of interest, assuming configuration M_1 or M_2, can be made by using the approaches described in Perreault et al. (2000). In particular, assuming a single change in the variance only, 1960 and 1954 are seen to be the most probable years of change (mode of $p(\tau|\mathbf{x}, M_2)$) for Bersimis and Outardes, respectively. In Perreault et al. (2000) a detailed analysis was performed to characterize a change in Churchill Falls annual energy inflows assuming model M_1. This inference may be revised in light of the above results by using the posterior distributions for M_3 derived in Appendix A. It is important to note that if Raftery's scheme has to be strictly used, discrimination between the three change-point models is not possible for Bersimis

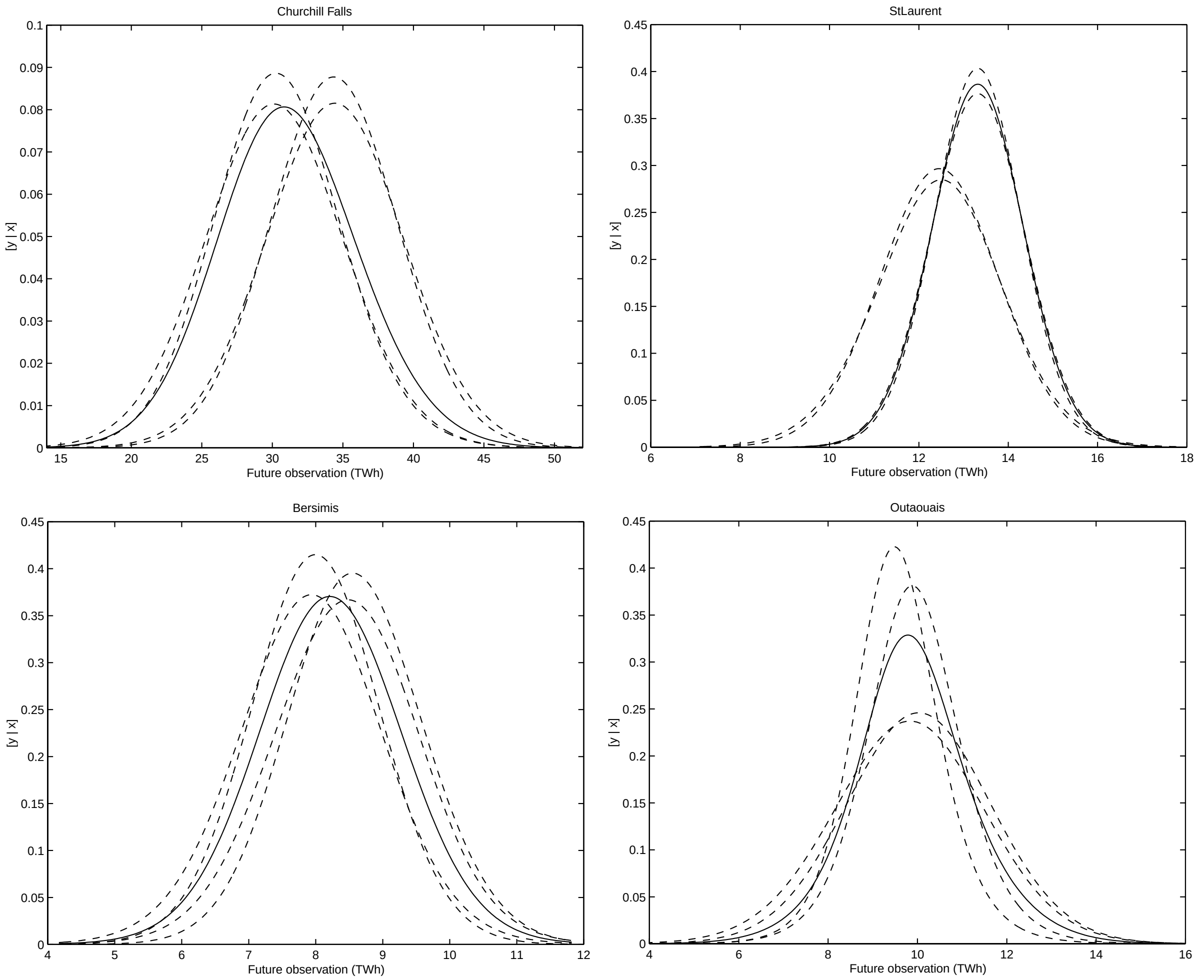

Fig. 3. Conditional and overall 1-year predictive densities.

Table 5
Results of a 1-year prediction for all hydropower systems

System	Conditional prediction				Overall
	M_0	M_1	M_2	M_3	
Bersimis	8.00	8.60	7.93	8.45	8.23
	(0.97)	(0.94)	(1.09)	(1.06)	(1.07)
Churchill Falls	34.43	30.16	34.34	30.44	31.16
	(4.93)	(4.88)	(4.58)	(4.39)	(4.91)
La Grande	79.32	77.02	79.46	77.46	78.61
	(11.55)	(11.64)	(13.02)	(13.37)	(12.24)
Manicouagan	21.91	20.88	22.07	20.91	21.82
	(2.96)	(3.01)	(3.65)	(3.94)	(3.37)
Outaouais	10.03	9.82	9.94	9.60	9.91
	(1.65)	(1.69)	(1.20)	(1.12)	(1.39)
Outardes	9.91	9.51	9.89	9.78	9.87
	(1.27)	(1.26)	(1.36)	(1.35)	(1.33)
St-Maurice	9.19	9.39	9.20	9.36	9.24
	(1.43)	(1.39)	(1.24)	(1.33)	(1.37)
St-Laurent	12.46	13.32	12.50	13.32	13.32
	(1.36)	(1.07)	(1.40)	(1.01)	(1.05)

and Churchill Falls. In fact, none of the three Bayes factors indicate positive evidence ($B_{i(jk)} > 3$).

5.4. Predictive analysis

The detection of a change and the identification of its type are important for Hydro-Québec, but prediction is often the real ultimate goal when analyzing annual energy inflows. In what follows, we derive the posterior predictive density for a future value under each of the models in the set $\mathscr{M} = \{M_0, M_1, M_2, M_3\}$ and use them to evaluate an overall posterior predictive distribution, as developed in Section 3.3. The forecasting of a future observation is then performed for all Hydro-Québec's series of annual energy inflows.

For models M_1 and M_3, integration in Eq. (7) to evaluate $p(y|\tau, \mathbf{x}, M_k)$ is straightforward and leads to an expression in closed form. The resulting density functions $p(y|\tau, \mathbf{x}, M_1)$ and $p(y|\tau, \mathbf{x}, M_3)$ are seen to be Student t-distributions. For model M_0, the predictive density $p(y|\mathbf{x}, M_0)$, which of course does not involve a change-point parameter τ, is also a Student distribution (Berger, 1985). The parameters of the corresponding t-distributions are presented in Table 4.

Hence, for a given change-point model, say M_3, the unconditional density $p(y|\mathbf{x}, M_3)$, is evaluated with Eq. (7), and appears as a finite mixture of the associated conditional distributions weighted by the $n-1$ values of $p(\tau|\mathbf{x}, M_3)$ given by Eq. (A5). As it was the case for the prior predictive density, the posterior predictive distribution $p(y|\mathbf{x}, M_2)$ cannot be calculated analytically. However, it can easily be approximated using the Gibbs sampling outputs (see Section 4.1).

Posterior predictive distributions for a 1-year future value, $p(y|\mathbf{x}, M_k)$, were computed for the eight series of annual energy inflows assuming the four models. The overall posterior predictive density was also evaluated using Eq. (18) and the posterior probabilities given in Table 3. These are shown in Fig. 3 for Bersimis, Churchill Falls, Outaouais and St-Laurent. The dotted lines correspond to the conditional posterior predictive densities $p(y|\mathbf{x}, M_k)$, while the continuous line is the combined predictive mixture. This figure illustrates how the overall posterior predictive distribution behaves, as compared to the conditional ones. When uncertainty about the true model is quite low, i.e. one of the posterior probabilities is near one (St-Laurent for instance), this distribution is very close to the predictive distributions obtained from the optimal model M_k^*. On the other hand, when uncertainty about the true model is considerable, i.e. all posterior probabilities are similar (Bersimis for instance), all conditional posterior predictive distributions contribute to the overall predictive distribution.

Various quantities of interest about future energy availability (quantiles, exceedance probabilities, etc.) can be derived from these densities. Table 5 reports a point prediction for each of the posterior predictive densities along with its standard deviation (in parentheses). A quadratic utility function was assumed, i.e. the 1-year forecast corresponds to the expected value (Eqs. (14) and (16)). The 90% credible interval for the overall prediction is illustrated in Fig. 4 for each hydropower system.

Among others, the main observations drawn from Table 5 and Fig. 4 are as follows:

- For 5 hydropower systems, assuming the usual stationarity hypothesis leads to 1-year higher forecasts of energy inflow than those obtained by maintaining all models in an overall point prediction. In particular, this is the case for Churchill Falls (+10%) and La Grande (+0.9%), the systems which contribute the most to the overall energy

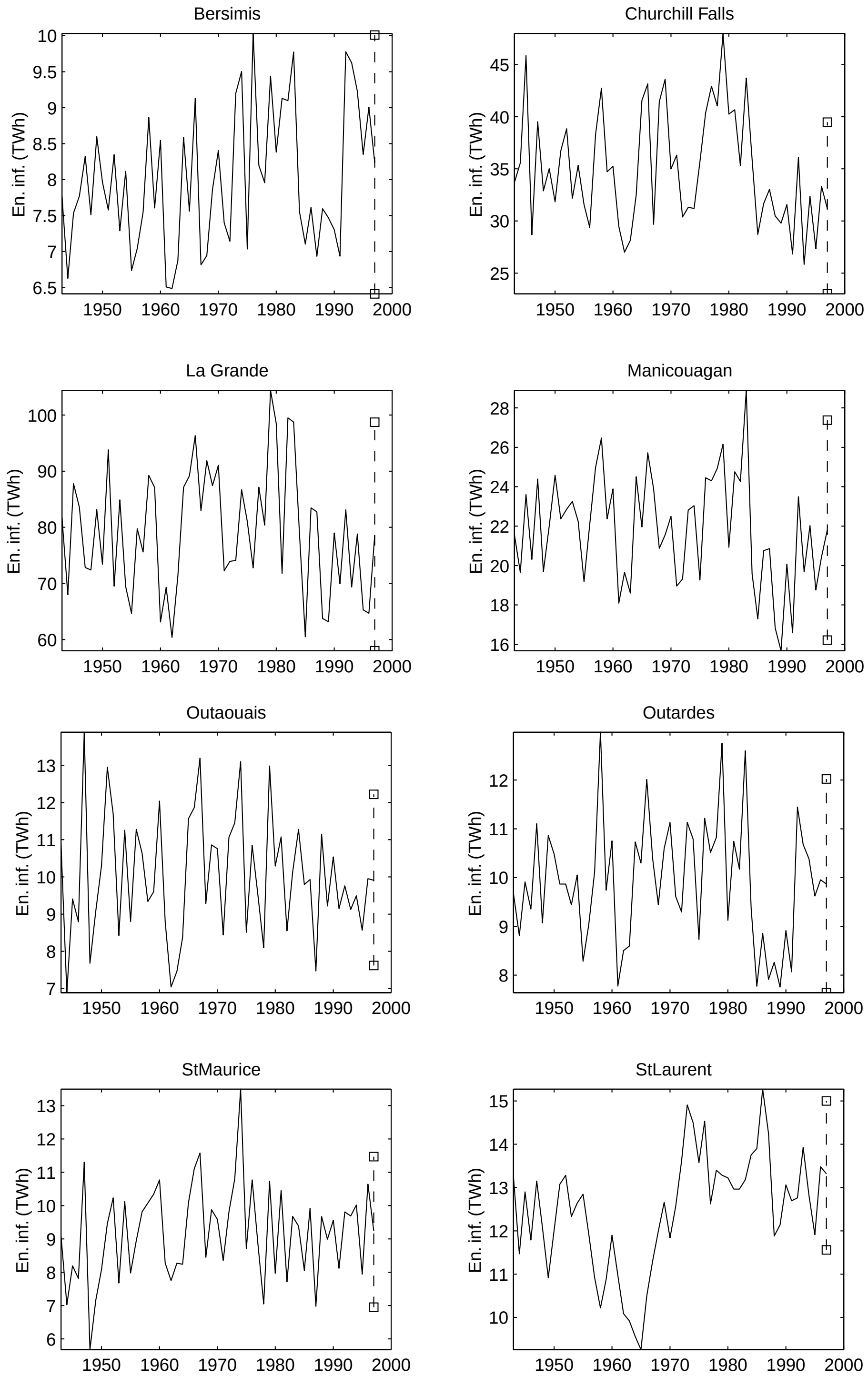

Fig. 4. 90% credibility intervals for a 1-year forecast.

produced by Hydro-Québec. In this particular application, incorporating uncertainty about M_0 provided a more conservative prediction about future energy availability.
- Considering all situations in an overall prediction leads to forecasts with larger standard deviations for five systems. This stems from the fact that, in addition to the sample uncertainty about the parameters, model uncertainty is taken into account.
- Finally, 90% credible intervals can be very large, especially when uncertainty about the true model is important. This is the case for Bersimis, La Grande and Manicouagan. The credible interval for the overall prediction may even cover more than the range of the observed sample data (Bersimis).

It is important to mention that, in contrast with Bayesian point estimates such as the posterior mean, the Bayes factors, and therefore model selection, tend to be sensitive to the choices of priors on the model parameters. Thus, an important issue which was not discussed here is the sensitivity of the approach to prior distributions. Such analysis should be performed in practice. Some guidelines and references can be found in Kass and Raftery (1995).

6. Discussion and conclusions

The hydrologist is often dealing with many models that involve different assumptions about the studied phenomenon. This is the case when changes have to be analyzed in hydrometeorological series. Even if the hydrologist wishes to infer using a single model, there are necessary many choices to be made. For example, one may hesitate between a change in the mean level only and a simultaneous change in both the mean and the variance. Moreover, a hydrologist would generally not consider the no change hypothesis as an impossible situation. In estimating quantities of interest, such as future realizations, it is certainly desirable to provide an assessment of the uncertainty that accounts for the model-building process itself (Kass and Raftery, 1995; Berger and Rios Insua, 1998). As it was mentioned in Section 1, most authors have centered their investigations on the estimation of the change-point and the intensity of the change for a given model. Few have addressed the problem of comparing various alternatives, including the no change situation. Another important issue which has been neglected in the literature is posterior predictive analysis for change-point problems. This paper has provided a new and convenient approach, based on Bayesian model selection: (1) for the overall assessment of change versus no change, and, if necessary, discrimination between different configurations of change; and (2) to maintain consideration of all models when forecasting a future observation. The approach was illustrated by its application on Hydro-Québec annual energy inflows assuming four univariate normal models (one model representing the no change hypothesis, and three change-point models).

The approaches presented herein can be used for other models such as gamma, binomial and Poisson distributions. More precisely, the use of probability distributions that belong to the exponential class of p.d.f.'s allows for exactly the same line of reasoning based on conjugacy for fixed change-point. Moreover, the independence hypothesis may be released, and change-point analysis can be performed for autoregressive models. Indeed, stationary autoregressive series with a strong coefficient of autocorrelation may exhibit the same type of behavior as series with changes in the mean level. With regard to the common use of AR models to represent hydrological persistency, it is therefore important to study which of two hypotheses (AR or M_1) is more credible. The methodology presented in this paper can answer this question by including in the set of models $\mathcal{M}$ an AR model. It is important to note that the models considered are not appropriate to forecast in the future if another change in climate state takes place. These models assume that we will stay indefinitely in the second state. If we think that in the future the state may change again, we should generalize the approach. However, the series of interest in this paper do not indicate any return to the original state since only one change seems to have occurred. Anyhow, if one is interested to generalize the approach, useful models for forecasting if another change in climate state may take place would be the so-called hidden Markov chain models (see Robert et al. (2000) and references therein).

Finally, a complete Bayesian analysis can be achieved to cast the statistical inference of hydrometeorological changes into an operational decision-making framework with terminal actions such as planning new dams. Such investigations could develop Krzysztofowicz's results (Krzysztofowicz, 1994): he embraces the decisional aspects for non-stationary stopping-control processes; or more recent work by Hobbs et al. (1997), and Venkatesh and Hobbs (1999) who studied the problem of investments for managing water levels under climate change uncertainty.

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, which considerably improved this paper. This research is supported by la Direction Général de l'Enseignement et de la Recherche de France (DGER), les fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR), and l'Institut National de la Recherche Scientifique (INRS-Eau).

Appendix A. Bayesian inference for univariate normal change-point models

A.1. Model M_3

Writing $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\boldsymbol{\sigma} = (\sigma_1^2, \sigma_2^2)$, the likelihood function resulting from n observations $\mathbf{x} = (x_1, x_2, \dots, x_n)$ generated by model M_3 is given by

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau, M_3) = \left(\frac{1}{2\pi}\right)^{n/2}\left(\frac{1}{\sigma_1^2}\right)^{\tau/2}\left(\frac{1}{\sigma_2^2}\right)^{(n-\tau)/2}$$

$$\times \exp\left\{-\frac{\tau}{2\sigma_1^2}[s_\tau^2 + (\bar{x}_\tau - \mu_1)^2]\right\}$$

$$\times \exp\left\{-\frac{n-\tau}{2\sigma_2^2}[s_{n-\tau}^2 + (\bar{x}_{n-\tau} - \mu_2)^2]\right\} \quad \text{(A1)}$$

where

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

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

For fixed τ, that expression has the same structure as a product of two normal distributions with two inverted gamma distributions. This suggests a normal-inverted gamma type of distribution to represent prior knowledge about $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. Assuming prior independence between τ and the other parameters, and that $p(\tau|M_3)$ is any discrete distribution on the set $\{1, 2, \dots, n-1\}$, leads to the joint prior parameter p.d.f.:

$$p(\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau|M_3) = \mathcal{N}(\mu_1|\phi_1, \lambda_1\sigma_1^2)\mathcal{N}(\mu_2|\phi_2, \lambda_2\sigma_2^2)$$

$$\times \mathcal{IG}(\sigma_1^2|\alpha_1, \beta_1)\mathcal{IG}(\sigma_2^2|\alpha_2, \beta_2)p(\tau|M_3)$$

$$= \mathcal{NNIGIG}(\boldsymbol{\mu}, \boldsymbol{\sigma}|\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta})p(\tau|M_3) \quad \text{(A2)}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2)$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, $\boldsymbol{\beta} = (\beta_1, \beta_2)$are the hyperparameters, and $\mathcal{IG}(\sigma^2|\alpha, \beta)$ stands for the inverted gamma distribution with parameters α and β. Because of conjugate properties the conditional joint posterior distribution $p(\boldsymbol{\mu}, \boldsymbol{\sigma}|\tau, \mathbf{x}, M_3)$ also belongs to the class of normal-inverted-gamma distributions, but with updated parameters $(\boldsymbol{\phi}', \boldsymbol{\lambda}', \boldsymbol{\alpha}', \boldsymbol{\beta}')$. More precisely,

$$p(\boldsymbol{\mu}, \boldsymbol{\sigma}|\tau, \mathbf{x}, M_3) = \mathcal{NNIGIG}(\boldsymbol{\mu}, \boldsymbol{\sigma}|\boldsymbol{\phi}', \boldsymbol{\lambda}', \boldsymbol{\alpha}', \boldsymbol{\beta}') \quad \text{(A3)}$$

where

$$\lambda_1' = \lambda_1/(1 + \tau\lambda_1), \qquad \lambda_2' = \lambda_2/[1 + (n-\tau)\lambda_2],$$

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

$$\phi_2' = [1 - \lambda_2'(n-\tau)]\phi_2 + \lambda_2'(n-\tau)\bar{x}_{n-\tau},$$

$$\alpha_1' = \alpha_1 + \tau/2, \qquad \alpha_2' = \alpha_2 + (n-\tau)/2,$$

$$\beta_1' = \frac{\tau}{2}[s_\tau^2 + (1 - \lambda_1'\tau)(\phi_1 - \bar{x}_\tau)^2] + \beta_1,$$

$$\beta_2' = \frac{(n-\tau)}{2}[s_{n-\tau}^2 + (1 - \lambda_2'(n-\tau))(\phi_2 - \bar{x}_{n-\tau})^2] + \beta_2$$

The prior predictive distribution $p(\mathbf{x}|\tau, M_3)$ can be determined by dividing the Bayes factor numerator

$p(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{x}|\tau, M_3)$ by Eq. (A3), cancelling factors involving $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$:

$$p(\mathbf{x}|\tau, M_3) = \left(\frac{1}{2\pi}\right)^{n/2} \sqrt{\frac{\lambda'_1\lambda'_2}{\lambda_1\lambda_2}} \frac{\beta_1^{\alpha_1}\beta_2^{\alpha_2}}{(\beta'_1)^{\alpha'_1}(\beta'_2)^{\alpha'_2}} \frac{\Gamma(\alpha'_1)\Gamma(\alpha'_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}. \quad \text{(A4)}$$

Given τ and $\mathbf{x}$, it is then straightforward to show that μ_1 and μ_2 are conditionally distributed as Student t-distributions $\mathcal{ST}(\mu_i|a, b, c)$, σ_1^2 and σ_2^2 as inverted gamma distributions $\mathcal{IG}(\sigma_i^2|\alpha, \beta)$. Integrating the appropriate parameters out of expression (A3), leads directly to these conditional posterior distributions of the parameters before and after the change-point:

$$p(\mu_i|\tau, \mathbf{x}, M_3) = \mathcal{ST}(\mu_i|\phi'_i, \alpha'_i(\lambda'_i\beta'_i)^{-1}, 2\alpha'_i),$$

$$p(\sigma_i^2|\tau, \mathbf{x}, M_3) = \mathcal{IG}(\sigma_i^2|\alpha'_i, \beta'_i),\ i = 1, 2.$$

To draw conclusions regarding the intensity of shifts, it is natural to define the parameters $\delta = \mu_2 - \mu_1$ and $\eta = \sigma_2^2/\sigma_1^2$. Their conditional posterior distributions can be deduced by simple univariate transformations of variable from the conditional distributions of the original parameters. Given τ and $\mathbf{x}$, η has a Beta distribution of the second kind (Menzefricke, 1981). It can also be shown that $(\alpha'_1\beta'_1\eta/\alpha'_2\beta'_2)$ is distributed as a Fisher distribution with $2\alpha'_2$ and $2\alpha'_1$ degrees of freedom. The posterior conditional distribution of δ is a Behrens–Fisher distribution, which cannot be evaluated in closed form. However, it can be efficiently approximated by a Student t-distribution. Details can be found in Box and Tiao (1973).

Since, according to the Bayes theorem, the joint p.d.f. for $(\mathbf{x}, \tau)$ is just $p(\mathbf{x}|\tau, M_3)p(\tau|M_3)$, the marginal posterior density of the change-point $\tau = 1, 2, \ldots, n-1$ under M_3 is readily seen to be

$$p(\tau|\mathbf{x}, M_3) = \frac{p(\mathbf{x}|\tau, M_3)p(\tau|M_3)}{\sum_{\tau=1}^{n-1} p(\mathbf{x}|\tau, M_3)p(\tau|M_3)} \propto p(\tau|M_3)\sqrt{\lambda'_1\lambda'_2}\frac{\Gamma(\alpha'_1)\Gamma(\alpha'_2)}{(\beta'_1)^{\alpha'_1}(\beta'_2)^{\alpha'_2}} \quad \text{(A5)}$$

This discrete distributions gives, at time point τ, the posterior probability of simultaneous shift occurrence in both the mean level and variance. To draw conclusions regarding the parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and the intensity of shifts $\delta = \mu_2 - \mu_1$ and $\eta = \sigma_2^2/\sigma_1^2$, their marginal posterior distributions must be derived. The corresponding marginal distributions are finite mixtures of the associated conditional distributions weighted by the $n-1$ values of $p(\tau|\mathbf{x}, M_3)$ given by Eq. (A5). For instance, we have

$$p(\mu_i|\mathbf{x}, M_3) = \sum_{\tau=1}^{n-1} p(\mu_i|\tau, \mathbf{x}, M_3)p(\tau|\mathbf{x}, M_3),\quad i = 1, 2$$

A.2. Model M_0

- *Prior distribution*

$$p(\mu, \sigma^2|M_0) = \mathcal{N}(\mu|\phi, \lambda\sigma^2)\mathcal{IG}(\sigma^2|\alpha, \beta) = \mathcal{NIG}(\mu, \sigma^2|\phi, \lambda, \alpha, \beta) \quad \text{(A6)}$$

- *Posterior distribution*

$$p(\mu, \sigma^2|\mathbf{x}, M_0) = \mathcal{NIG}(\mu, \sigma^2|\phi', \lambda', \alpha', \beta') \quad \text{(A7)}$$

where

$$\lambda' = \frac{\lambda}{(1 + n * \lambda)}$$

$$\phi' = (1 - \lambda' * n) * \phi + \lambda' * n * \bar{x}$$

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

and

$$\beta' = \tfrac{1}{2}[ns_n^2 + (1 - \lambda)^{-1}(\phi - \bar{x}_n)^2] + \beta$$

A.3. Model M_1

- *Prior distribution*

$$\begin{aligned} p(\boldsymbol{\mu}, \sigma^2, \tau|M_1) &= \mathcal{N}(\mu_1|\phi_1, \lambda_1\sigma^2)\mathcal{N}(\mu_2|\phi_2, \lambda_2\sigma^2) \\ &\quad \times \mathcal{IG}(\sigma^2|\alpha, \beta)p(\tau|M_1) \\ &= \mathcal{NNIG}(\boldsymbol{\mu}, \sigma^2|\boldsymbol{\phi}, \boldsymbol{\lambda}, \alpha, \beta)p(\tau|M_1) \end{aligned} \quad \text{(A8)}$$

- *Conditional posterior distribution*

$$p(\boldsymbol{\mu}, \sigma^2|\tau, \mathbf{x}, M_1) = \mathcal{N}\mathcal{N}\mathcal{I}\mathcal{G}(\boldsymbol{\mu}, \sigma^2|\boldsymbol{\phi}', \boldsymbol{\lambda}', \alpha', \beta') \quad \text{(A9)}$$

where

$$\lambda_1' = \lambda_1/(1 + \tau\lambda_1),$$

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

$$\lambda_2' = \lambda_2/[1 + (n - \tau)\lambda_2],$$

$$\phi_2' = [1 - \lambda_2'(n - \tau)]\phi_2 + \lambda_2'(n - \tau)\bar{x}_{n-\tau,}$$

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

$$+ \frac{n-\tau}{2}[s_{n-\tau}^2 + (1 - \lambda_2'(n - \tau))(\phi_2 - \bar{x}_{n-\tau})^2] + \beta,$$

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

- *Prior predictive distribution conditional on* τ

$$p(\mathbf{x}|\tau, M_1) = \left(\frac{1}{2\pi}\right)^{n/2}\sqrt{\frac{\lambda_1'\lambda_2'}{\lambda_1\lambda_2}}\frac{\beta^\alpha}{(\beta')^{\alpha'}}\frac{\Gamma(\alpha')}{\Gamma(\alpha)} \quad \text{(A10)}$$

- *Marginal posterior distribution of* τ

$$p(\tau|\mathbf{x}, M_1) \propto p(\tau|M_1)\sqrt{\lambda_1'\lambda_2'}(\beta')^{-\alpha'} \quad \text{(A11)}$$

A.4. Model M_2

- *Prior distribution*

$$p(\mu, \sigma_1^2, \sigma_2^2, \tau|M_2)$$

$$= \mathcal{N}(\mu|\phi, \lambda\sigma_1^2)\mathcal{I}\mathcal{G}(\sigma_1^2|\alpha_1, \beta_1)\mathcal{I}\mathcal{G}(\sigma_2^2|\alpha_2, \beta_2)p(\tau|M_2)$$

$$= \mathcal{N}\mathcal{I}\mathcal{G}\mathcal{I}\mathcal{G}(\mu, \sigma_1^2, \sigma_2^2|\phi, \lambda, \boldsymbol{\alpha}, \boldsymbol{\beta})p(\tau|M_2) \quad \text{(A12)}$$

- *Full conditionals*

$$p(\mu|\sigma_1^2, \sigma_2^2, \tau, \mathbf{x}, M_2)$$

$$= \mathcal{N}\left(\mu|\phi'', \frac{\lambda'\sigma_1^2\sigma_2^2}{\lambda'(n-\tau)\sigma_1^2 + \sigma_2^2}\right)$$

$$p(\sigma_1^2|\mu, \sigma_2^2, \tau, \mathbf{x}, M_2)$$

$$= \mathcal{I}\mathcal{G}\left(\sigma_1^2|\alpha_1', \frac{\tau}{2}\left[s_\tau^2 + (\bar{x}_\tau - \mu)^2 + \frac{(\mu - \phi)^2}{\tau\lambda}\right] + \beta_1\right)$$

$$p(\sigma_2^2|\mu, \sigma_1^2, \tau, \mathbf{x}, M_2)$$

$$= \mathcal{I}\mathcal{G}\left(\sigma_2^2|\alpha_2', \frac{n-\tau}{2}[s_{n-\tau}^2 + (\bar{x}_{n-\tau} - \mu)^2] + \beta_2\right)$$

$$p(\tau|\mu, \sigma_2^2, \sigma_1^2, \mathbf{x}, M_2) \propto p(\mathbf{x}|\mu, \sigma_2^2, \sigma_1^2, \tau, M_2)p(\tau|M_2)$$

where

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

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

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

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

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

Appendix B. Implementation of Chib's approach

In the context of the general change-point model (1), the full conditional distributions are $p(\tau|\theta_1, \theta_2, \mathbf{x})$, $p(\theta_1|\theta_2, \tau, \mathbf{x})$ and $p(\theta_2|\theta_1, \tau, \mathbf{x})$, and $\{\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, \theta_2^{(j)}, \tau^{(j)});\ j = 1, \ldots, m\}$ are the m draws from the joint posterior density $p(\boldsymbol{\theta}|\mathbf{x}, M_k)$ obtained using the Gibbs sampler, after the first t values of the chain have been discarded.

Chib's estimate for $p(\mathbf{x}|M_k)$ is given by

$$\hat{p}_{CH}(\mathbf{x}|M_k) = \frac{p(\mathbf{x}|\boldsymbol{\theta}^*, M_k)p(\boldsymbol{\theta}^*|M_k)}{\hat{p}(\boldsymbol{\theta}^*|\mathbf{x}, M_k)} \quad \text{(B1)}$$

where $\hat{p}(\boldsymbol{\theta}^*|\mathbf{x}, M_k)$ is an estimate for $p(\boldsymbol{\theta}|\mathbf{x}, M_k)$ evaluated at $\boldsymbol{\theta}^*$. Using basic probability rules, the posterior density ordinate $p(\boldsymbol{\theta}*|\mathbf{x}, M_k)$, for a given value $\boldsymbol{\theta}^*$, can be expressed as

$$p(\boldsymbol{\theta}^*|\mathbf{x}, M_k) = p(\theta_1^*|\mathbf{x}, M_k)p(\theta_2^*|\theta_1^*, \mathbf{x}, M_k)p(\tau^*|\theta_1^*, \theta_2^*, \mathbf{x}, M_k) \quad \text{(B2)}$$

The last term of the right-hand side is nothing other than the full conditional distribution of τ evaluated at τ^* after setting $(\theta_1, \theta_2, \tau) = (\theta_1^*, \theta_2^*, \tau^*)$. Since all full conditionals are assumed to be expressible in closed form, $p(\tau^*|\theta_1^*, \theta_2^*, \mathbf{x}, M_k)$ can be evaluated exactly. For the same reason, the first term $p(\theta_1^*|\mathbf{x}, M_k)$, which is the marginal ordinate, can be estimated from the m draws of the Gibbs run using the Rao–Blackwellized estimate:

$$\hat{p}(\theta_1^*|\mathbf{x}, M_k) = \frac{1}{m}\sum_{j=1}^{m} p\Big(\theta_1^*|\theta_2^{(j)}, \tau^{(j)}, \mathbf{x}, M_k\Big) \tag{B3}$$

However, the partial conditional ordinate $p(\theta_2^*|\theta_1^*, \mathbf{x}, M_k)$ in Eq. (B2) cannot be obtained directly. In fact, the draws of θ_2 from the Gibbs sampler are from the distribution $p(\theta_2^*|\mathbf{x}, M_k)$ and not from $p(\theta_2|\theta_1, \mathbf{x}, M_k)$, so that an estimate for this density cannot be calculated using these simulated values. The simple solution proposed by Chib (1995) is to continue sampling for an additional m iterations with the conditional densities

$$p(\theta_2|\theta_1^*, \tau, \mathbf{x}, M_k)$$

and

$$p(\tau|\theta_1^*, \theta_2, \mathbf{x}, M_k)$$

where in each of these densities, θ_1 is set equal to θ_1^*. Consequently,

$$\hat{p}(\theta_2^*|\theta_1^*, \mathbf{x}, M_k) = \frac{1}{m}\sum_{j=1}^{m} p(\theta_2^*|\theta_1^*, \tau^{(j)}, \mathbf{x}, M_2) \tag{B4}$$

is a Rao–Blackwellized estimate for the second term in Eq. (B2).

Substituting the density estimates Eqs. (B3) and (B4) into Eq. (B2), and then the obtained estimate $\hat{p}(\boldsymbol{\theta}^*|\mathbf{x}, M_k)$ into Eq. (B1) yields finally to

$$\hat{p}_{CH}(\mathbf{x}|M_k) = \frac{p(\mathbf{x}|\boldsymbol{\theta}^*, M_k)p(\boldsymbol{\theta}^*|M_k)}{\hat{p}(\theta_1^*|\mathbf{x}, M_k)\hat{p}(\theta_2^*|\theta_2^*, \mathbf{x}, M_k)p(\tau^*|\theta_1^*, \theta_2^*, \mathbf{x}, M_k)} \tag{B5}$$

Since expression (19) holds for any value of $\boldsymbol{\theta}$, the choice of $\boldsymbol{\theta}^*$ is not critical. However, Chib (1995) pointed out that efficiency considerations dictate that for a given number of replicates m, the prior predictive is likely to be more accurately estimated at a high density point, where more samples are available, than at a point in the tails. The posterior mode or the maximum likelihood estimate, which can be easily approximated from the initial Gibbs output, should be used. Alternatively, one can also use the posterior mean provided there is no concern that it is a low density point. Finally, this approach can be easily generalized to more than three unknown parameters.

References

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

Berger, J.O., Pericchi, L.R., 1996. The intrinsic Bayes factor for model selection and prediction. J. Am. Stat. Assoc. 91, 109–122.

Berger, J.O., Rios Insua, D., 1998. Recent developments in Bayesian inference with applications in hydrology. In: Parent, E., Hubert, P., Bobée, B., Miquel, J. (Eds.). Statistical and Bayesian Methods in Hydrological Sciences, UNESCO Publishing, pp. 43–62.

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

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, MA (p. 107).

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

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

Chib, S., 1995. Marginal likelihood from the Gibbs output. J. Am. Stat. Assoc. 90, 1313–1321.

Gelfand, A.E., Dey, D.K., 1994. Bayesian model choice: asymptotics and exact calculations. J. R. Stat. Soc. B 56, 501–514.

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

Hobbs, B.F., Chao, P.T., Venkatesh, B.N., 1997. Using decision analysis to include climate change in water resources decision making. Clim. Change 37, 177–202.

Jeffreys, H., 1961. Theory of Probability, Oxford University Press, Oxford, UK.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

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.

Krzysztofowicz, R., 1994. Strategic decisions under nonstationary conditions: a stopping-control paradigm. In: Duckstein, L., Parent, E. (Eds.). Engineering Risk in Natural Resources Management, NATO Advanced Studies Institute Series, Series E, 275. Kluwer, Dordrecht, pp. 359–371.

Lubès-Niel, H., Masson, J.M., Paturel, J.E., Servat, E., 1998.

Variabilité climatique et statistiques. Etude par simulation de la puissance et de la robustesse de quelques tests utilisés pour vérifier l'homogénéité de chroniques. Rev. Sci. Eau 3, 383–408.
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. Stat. 30, 141–146.
O'Hagan, A., 1995. Fractional Bayes factors for model comparisons. J. R. Stat. Soc. B 57, 99–138.
Perreault, L., Bernier, J., Bobée, B., Parent, E., 2000. Bayesian change-point analysis in hydrometeorological time series 1. The normal model revisited. Journal of Hydrology. 235, 221–241.
Raftery, A.E., 1996. Approximate Bayes factors and accounting for model uncertainty in generalised linear models. Biometrika 83, 251–266.
Rao, A.R., Tirtotjondro, W., 1996. Investigation of changes in characteristics of hydrological time series by Bayesian methods. Stochastic Hydrol. Hydraul. 10, 295–317.
Robert, C.P., Rydén, T., Titterington, D.M., 2000. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. J. R. Stat. Soc. B 62, 57–75.
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.
Stephens, D.A., 1994. Bayesian retrospective multiple-changepoint identification. Appl. Stat. 43, 159–178.
Venkatesh, B.N., Hobbs, B.F., 1999. Analysing investments for managing Lake Erie levels under climate change uncertainty. Water Resour. Res. 35, 1671–1683.