Ann Inst Stat Math (2009) 61:355–370
DOI 10.1007/s10463-007-0148-y

Bayesian isotonic changepoint analysis

Enrique E. Alvarez · Dipak K. Dey

Received: 30 March 2005 / Revised: 26 April 2007 / Published online: 18 August 2007

Abstract A general approach to Bayesian isotonic changepoint problems is developed. Such isotonic changepoint analysis includes trends and other constraint problems and it captures linear, non-smooth as well as abrupt changes. Desired marginal posterior densities are obtained using a Markov chain Monte Carlo method. The methodology is exemplified using one simulated and two real data examples, where it is shown that our proposed Bayesian approach captures the qualitative conclusion about the shape of the trend change.

Keywords Bayesian inference · Change point problem · Isotonic regression · Order restricted inference

1 Introduction

Changepoint problems often arise in many statistical analysis problems. The literature on changepoint problems is by now enormous, with applications in virtually every branch of science. The classical *parametric, one dimensional* change point problem

E. E. Alvarez (✉)
Departamento de Matemática y CONICET, Universidad Nacional de La Plata, Calle 50 e/115,
1900 La Plata, Argentina
e-mail: ealvarez@mate.unlp.edu.ar

D. K. Dey
Department of Statistics, University of Connecticut,
215 Glenbrook Rd., Storrs, CT 06269-4120, USA

considers testing for the competing hypotheses:

$$
\begin{aligned}
&H_0 : \theta_1 = \cdots = \theta_n, \\
&H_1 : \theta_1 = \cdots = \theta_\tau \neq \theta_{\tau+1} = \cdots = \theta_n, \quad \exists 1 \leq \tau \leq n,
\end{aligned}
\tag{1}
$$

from data $x_1, \ldots, x_n$ i.i.d. with distribution $F_\theta(\cdot)$ indexed by $\theta \in \Theta \subset \mathbb{R}$. A vast literature exists on this classical problem, both from the frequentist or Bayesian approach, in an *off-line* or sequential treatment, parametric or non-parametric. Excellent reviews can be found in Hinkley et al. (1980), Zacks (1983), Siegmund (1986), Bhattacharya (1994) and many others.

Here we consider only the so called nonsequential problems. Our focus is on fully Bayesian parametric approach for inferences on the isotonic changepoint. Bayesian framework for inferences on changepoint dates back to work by Chernoff and Zacks (1964) and Shiryayev (1963). Smith (1975), Carlin et al. (1992) and Dey and Purkayastha (1997) developed different Bayesian approaches including hierarchichal Bayesian methods for inferences related to changepoints. However, none of those papers consider isotonic structure. The rest of the paper is organized as follows. In Sect. 2, we develop the formulation of the problem using both frequentist and Bayesian approaches. Sect. 3 is devoted to the study of posterior distributions. In the last paragraphs of Sect. 3 and in Sects. 4 and 5 we consider one simulated example and two real datasets to demonstrate our proposed methodology. The description of the computing algorithm used to fit the models is postponed to the Appendix. For a general reference, the problem of fitting Bayesian models in constrained parameter settings is considered extensively in the pioneering article of Gelfand et al. (1992).

In this paper we consider Bayesian estimation of $\theta_1, \ldots, \theta_n$ as well as testing for H_0 in (1) in a situation where the parameters satisfy monotonicity but are otherwise arbitrary. A similar problem was studied from a frequentist approach by Wu et al. (2001), which motivated in part our present study. Next we provide a brief review of the frequentist and Bayesian approaches to the isotonic change-point problem.

2 Formulation of the model

2.1 Frequentist approach

Suppose data $\mathcal{X} := \{X_1, \ldots, X_n\}$ are independent with $X_i \overset{\text{i.i.d.}}{\sim} N(\mu_i, \sigma^2)$, $i = 1, \ldots, n$. Then the log-likelihood function is

$$
l(\mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_i)^2 + C, \tag{2}
$$

where C is a generic constant, independent of μ and the parameter space is $\Omega = \{\mu \in \mathbb{R}^n : \mu_1 \leq \cdots \leq \mu_n\}$. Then the maximum likelihood estimator (MLE)

$\hat{\mu}$ =argmax$_{\mu \in \Omega}$ $l(\mu)$ is given by

$$\hat{\mu}_k = \max_{i \le k} \min_{j \ge k} \frac{X_i + \cdots + X_j}{j - i + 1}$$

(e.g., Robertson et al. 1988, p. 24). This estimator is affected by the so called *spiking problem* in large samples, in that $\hat{\mu_1}$ is too small while $\hat{\mu_n}$ is too large. Instead of (2), therefore it is customary to consider the corresponding *penalized log likelihood function*

$$l^*(\mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_i)^2 + \frac{r\sqrt{n}}{\sigma^2}(\mu_n - \mu_1) + C, \tag{3}$$

where the term $r\sqrt{n}(\mu_n - \mu_1)$ penalizes the difference $\delta := (\mu_n - \mu_1)$. Letting $Y_1 := X_1 + r\sqrt{n}$, $Y_n = X_1 - r\sqrt{n}$ and $Y_i = X_i$ for $1 < i < n$, the maximum penalized likelihood estimator (MPLE) is

$$\tilde{\mu}_k = \max_{i \le k} \min_{j \ge k} \frac{Y_i + \cdots + Y_j}{j - i + 1}. \tag{4}$$

Under H_0, the MPLE is $\tilde{\mu} = \bar{X}$ which suggests a test statistic of the form

$$\Lambda_{n,r} = \frac{1}{\hat{\sigma}_n^2} \sum_{k=1}^{n} (\tilde{\mu} - \bar{X})^2,$$

where σ_n^2 is a consistent estimator of σ^2.

This estimator, even though it was developed for independent and Gaussian data, could be applied to many other scenarios. In particular, for time series data with short range dependence, the asymptotic null distribution of $\Lambda_{n,r}$ is developed in Wu et al. (2001).

2.2 Bayesian approach

Now we start by providing the estimator (4) with a Bayesian interpretation, which provides a natural way to propose useful generalizations. As in the frequentist approach, suppose the data $\mathcal{X} := \{X_1, \ldots, X_n\}$ are independent with $X_i \overset{\text{i.i.d.}}{\sim} N(\mu_i, \sigma^2)$, where it is known that $\mu_1 \le \cdots \le \mu_n$. Let σ have Lebesgue measure on $\mathbb{R}^+$ as improper prior and construct a prior on $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ in three steps: (*i*) given σ, let μ_1 have Lebesgue measure on $\mathbb{R}$ as improper prior; (*ii*) given σ^2 and μ_1, let $\delta := (\mu_n - \mu_1) \sim$ Exponential(γ/σ^2); and (*iii*) given σ^2, μ_1 and δ, let $\mu_2, \ldots, \mu_{n-1}$ be distributed as the order statistics from a uniform distribution on (μ_1, μ_n). Under this specification, the log of the posterior likelihood is

$$l(\boldsymbol{\mu}, \sigma^2|\mathcal{X}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_i)^2 + \frac{\gamma}{\sigma^2}(\mu_n - \mu_1) + C \tag{5}$$

on the set Ω (or $-\infty$ otherwise) where C is a constant.

The resemblance between the frequentist *penalized log-likelihood function* (3) and the Bayesian *posterior distribution* is noteworthy, with just $r = \gamma/\sqrt{n}$. This yields immediately the following proposition.

Proposition 1 *The mode of the posterior distribution* (5) *is the frequentist estimator* (4) *with* $r = \gamma/\sqrt{n}$.

In light of Proposition 1, the frequentist estimator can be interpreted as a special case of a Bayes estimator. It is the mode of a posterior density with suitably chosen (uninformative) priors. Instead, in a fully Bayesian treatment, we will allow for wider prior specifications, more in tune with the objectives of change point detection under monotonicity. Our methodology thus represents a generalization and extension of the frequentist change point inference under a non-decreasing constraint. It is also noteworthy that we believe a Bayesian approach is more natural for this problem. This is because if we are willing to allow the statistician to choose a penalty constant according to how large (s)he thinks the difference δ is, we should as well allow her/him to provide her/his full set of priors. Of course, it could be objected that in a frequentist treatment, the penalty constant r need not be subjective as it could, for instance, be chosen via *cross-validation*. That idea is alluring, but we ignore of any way to make it implementable, for what is known is the asymptotic distribution of $\Lambda_{n,r}$ when r is a constant, but not when r is chosen in a *data dependent* way.

In a Bayesian set-up, the frequentist *penalty* $\gamma/\sqrt{n} = r$ takes the form of a *hyperparameter* in the distribution of $(\mu_n - \mu_1)$. Thus, in practical applications it is natural to choose its value using techniques of *prior elicitation* whenever possible. We will illustrate this point with the Argentina rainfall data in Sect. 5. For applications where it is not possible to adopt a sensible choice for γ, the practitioner is advised to experiment with a range of values and perform appropriate sensitivity analysis. This approach will be illustrated with the Global Warming dataset in Sect. 4

2.2.1 The model

We assume our data $\mathcal{X} = (X_1, \ldots, X_n)$ satisfies

$$X_i = \mu + d\phi_i + \varepsilon_i,$$

where given μ, δ, σ^2, and ϕ_i, $1 \le i \le n$, the disturbances $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$. The conditional distribution of the data is thus $X_i \overset{\text{i.i.d.}}{\sim} N(\mu_i, \sigma^2)$, where $\mu_i := \mu + \delta\phi_i$ represents the trend and it is subject to a monotonicity constraint specified by $\delta \in \mathbb{R}$ and $-1 = \phi_1 = \phi_2 \le \cdots \le \phi_{n-1} = \phi_n = 1$.

It is important at this point to interpret the parameters. We note that (i) μ is a location parameter for the probability distribution of $\mathcal{X}$, representing the overall mean of the data; (ii) d provides two types of information, its sign indicates whether the trend is increasing or decreasing and its magnitude indicates by how much; finally (iii) the multivariate parameter $\boldsymbol{\phi} := (\phi_1, \ldots, \phi_n)$ indicates the way the trend changes, some possibilities are smooth and convex, smooth and concave, abrupt at some isolated points or any combination of all those. It is noteworthy that in the classical change point problem referred at the beginning of Sect. 1, we would have $-1 = \phi_1 = \cdots = \phi_\tau < \phi_{\tau+1} = \cdots = \phi_n = 1$ and δ could be positive or negative.

For our prior specification we propose

$$\mu \sim N(\mu_0, \sigma_\mu^2) \tag{6}$$

$$\delta \sim N(\delta_0, \sigma^2/r^2) \tag{7}$$

$$(1/\sigma^2) =: \lambda \sim \text{Gamma}(a, b), \tag{8}$$

where in absence of good prior information we would adopt a noninformative version provided by some choice of the parameters $\mu_0 = 0$, σ_μ "large", $\delta_0 = 0$, r "small", a "small" and b "large". This prior specification also conveys a new meaning to the frequentist penalty constant $r = [\text{Var}(\varepsilon)/\text{Var}(\delta)]^{\frac{1}{2}}$ as the variability of the observations about its trend, relative to the anticipated variability of δ given by the prior specification. Intuitively, this is saying that there are two situations which deserve a large penalty: i) when the "true" $\delta = (\mu_n - \mu_1)$ is small; and ii) when there is large variability around the trend, because otherwise there is a danger of spuriously estimating a large d when in fact δ may be almost nil. This observation is useful for the practicioner when it is aimed at choosing the hyperparameter r via prior elicitation. Note also that it is straightforward to modify the choice (7) to accommodate for non-decreasing (or non-increasing) trends: e.g., simply take δ (or $-\delta$) $\sim$Exponential$\big(r/2\sigma^2\big)$.

We construct now a prior for the parameter $\boldsymbol{\phi}$ using the following steps. First we "tie it down" at the endpoints by letting $\phi_1 = \phi_2 = -1$ and $\phi_{n-1} = \phi_n = 1$; next in order that the intermediate ϕ's are nondecreasing and bounded in $[-1, 1]$, we proceed in two stages: (i) let the $(n-3)$-dimensional parameter $\mathbf{p} := (p_3, \ldots, p_{n-1})$ have a Dirichlet distribution with parameter $\boldsymbol{\alpha} := (\alpha_3, \ldots, \alpha_{n-1})$, i.e., $\mathbf{p}$ has an absolutely continuous density

$$f_{\mathbf{p}}(p_3, \ldots, p_{n-1}) = \frac{\Gamma\left(\sum_{j=3}^{n-1} \alpha_j\right)}{\prod_{j=3}^{n-1} \Gamma(\alpha_j)} \prod_{j=3}^{n-1} p_j^{\alpha_j - 1} \tag{9}$$

for $\mathbf{p}$ in the unit $(n-3)$-dimensional simplex $\Delta := \left\{\mathbf{p} \in [0, 1]^{n-3} : \sum_{i=3}^{n-1} p_j = 1\right\}$; next (ii) construct the intermediate ϕs by

$$\phi_i = 2\left(\sum_{j=3}^{i} \alpha_j\right) - 1, \quad \text{for } 2 < i < n. \tag{10}$$

The above construction, which ensures that $\boldsymbol{\phi}$ is non-decreasing and *tied-in* at the end points, is appealing on both practical and theoretical grounds. On the practical side, it is paramount that the specification for $\boldsymbol{\phi}$ provided in (9) and (10) allows to set a prior on the form of the trend, which is very useful for prior elicitation. Some possibilities are:

Linear: Suppose that the statistician has no prior information on the type of trend, then it is natural to propose $\alpha_1 = \cdots = \alpha_{n-1} = \alpha$ under which $\boldsymbol{\phi}$ will be on average of the line connecting the points $(1, -1)$ and $(n, 1)$, a noninformative specification.

Convex: If instead it is suspected that the trend has been changing in a smooth concave fashion we could take $\alpha_i := i^{1.2}$.

Concave: Take for instance $\alpha_i := i^{0.8}$.

Sigmoidal: In some circumstances there are reasons to believe that the rate of the change in trend was accelerating up to a certain time, giving way to a deceleration later. Examples of this phenomenon abound in marketing; one such example is the market penetration of novel electronic items. This type of trend could be obtained in our context by taking $\alpha_i = i$ for $1 \le i \le \tau$ and $\alpha_i = n + 1 - i$ for $\tau < i \le n$.

Linear, non-smooth: Taking $\alpha_i := 1, i \neq \tau$ and $\alpha_\tau := 10$.

Abrupt change: The $\boldsymbol{\phi}$ profiles for an abrupt change at τ which can be obtained by making the extreme value of α_τ in the previous case, i.e., taking $\alpha_i = 1, i \neq \tau$ and $\alpha_\tau = 50$.

The Bayesian formulation is also appealing because it enables estimation of $(\mu + \delta\boldsymbol{\phi})$ by *credible intervals*. Those provide not only an indication of the way the trend has changed and its magnitude, but it also makes possible to carry out tests against more general alternative hypotheses than H_1 in (1). This will be illustrated for the Global Warming dataset in Sect. 4, where we use credible bands around $\boldsymbol{\phi}$ to assess whether there has been acceleration in the process of global warming, as represented by a convex form of trend.

3 Posterior densities

Under the full model specification (6),(7), and (8) with $\mu_0 = 0$ and $\delta_0 = 0$ we obtain the following posteriors:

$$[\mu|\mathcal{X}, \delta, \sigma^2, \boldsymbol{\phi}] \sim N\left(\frac{\sum_{i=1}^n (X_i - \delta\phi_i)}{n + \sigma^2/\sigma_\mu^2}; \frac{\sigma^2}{n + \sigma^2/\sigma_\mu^2}\right), \tag{11}$$

$$[\delta|\mathcal{X}, \mu, \sigma^2, \boldsymbol{\phi}] \sim N\left(\frac{\sum_{i=1}^n \phi_i (X_i - \mu)}{\sum_{i=1}^n \phi_i^2 + r^2}; \frac{\sigma^2}{\sum_{i=1}^n \phi_i^2 + r^2}\right), \tag{12}$$

$$[\lambda|\mathcal{X}, \mu, \delta, \boldsymbol{\phi}] \sim \text{Gamma}\left(\frac{n}{2} + a; \frac{1}{2}\sum_{i=1}^n (X_i - \mu - \delta\phi_i)^2 + b\right), \tag{13}$$

where λ is the *precision parameter* $\lambda := (1/\sigma)^2$. As for the remaining posterior distribution $[\boldsymbol{\phi}|\mathcal{X}, \mu, \delta, \sigma^2]$, we show the development next. First, a standard Jacobian calculation shows that the (joint) density of $\boldsymbol{\phi}$ is given by

$$f_{\boldsymbol{\phi}}(\phi_3, \ldots, \phi_{n-1}) = \frac{\Gamma\left(\sum_{i=3}^{n-1} \alpha_i\right)}{\prod_{i=3}^{n-1} \Gamma(\alpha_i)} \prod_{i=3}^{n-1} \left(\frac{\phi_i - \phi_{i-1}}{2}\right)^{\alpha_i - 1} \tag{14}$$

on the set $\Omega_{\boldsymbol{\phi}}$, given by $-1 \leq \phi_3 \leq \cdots \leq \phi_{n-1} = 1$. With this, the posterior distribution of $\boldsymbol{\phi}$ given the data and all the other parameters can be stated up to proportionality by

$$f_{\boldsymbol{\phi}|(\mathcal{X},\mu,\delta,\sigma)}(\phi_3, \ldots, \phi_{n-1}) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\phi_i - \frac{X_i - \mu}{\delta}}{\sigma/\delta}\right)^2\right\} \prod_{i=3}^{n-1}\left(\frac{\phi_i - \phi_{i-1}}{2}\right)^{\alpha_i - 1}$$

on the set $\Omega_{\boldsymbol{\phi}}$. In particular, under the *improper prior* with parameters $\alpha_i = 1$ for all i,

$$[\boldsymbol{\phi}|\mathcal{X}, \mu, \delta, \sigma] \sim N\left(\frac{X - \mu}{\delta}; \frac{\sigma^2}{\delta^2}\right) 1_{\Omega_{\boldsymbol{\phi}}}(\phi_1, \ldots, \phi_n), \tag{15}$$

i.e., an $(n - 4)$ dimensional normal distribution truncated to $-1 \leq \phi_3 \leq \cdots \leq \phi_{n-2} \leq 1$.

Credible sets will be based on the joint posterior distribution of the parameters given the sample. This joint conditional distribution $[\mu, \delta, \sigma, \boldsymbol{\phi}|\mathcal{X}]$ will be obtained via the Gibbs sampler. We present at this point some motivating examples, leaving the description of the algorithm which implements the Gibbs sampler as given in the Appendix. R code is available from the corresponding author upon request.

Our first illustration is a simulated example created to verify the algorithm. We consider our data $X_1, \ldots, X_n$ to be equal to a deterministic trend $T(t)$ for $t = 1, \ldots, 50$ plus a normal noise with variance $(0.8)^2$. Suppose the deterministic trend has the from

$$T(t) = \begin{cases} 25 - \sqrt{25^2 - t^2} & 1 \leq t \leq 25 \\ 35 + \log(t - 25) & 25 < t \leq 50 \end{cases},$$

so that it is convex on [0, 25], has a jump between $t = 25$ and $t = 26$ and is concave and fairly flat thereafter. The performance of the Bayes estimator and a comparison with the frequentist estimator is shown in Fig. 1.

With a Bayesian approach, not only can we obtain different types of point estimators but also provide confidence intervals about the trend. Figure 2 shows the histograms of the posterior distributions of μ, σ and δ, as well as the point estimate of the function $\boldsymbol{\phi}$ with bands around it.

Finally, Figure 3 shows the effects of increasing the hyper-parameter r, which is the analogue of the frequentist *penalty constant*. While the posterior distributions of μ and σ^2 remain unaffected as r increases, for the other parameters we notice: (i) the posterior of δ shrinks towards zero, as expected; and (ii) the function $\boldsymbol{\phi}$ becomes

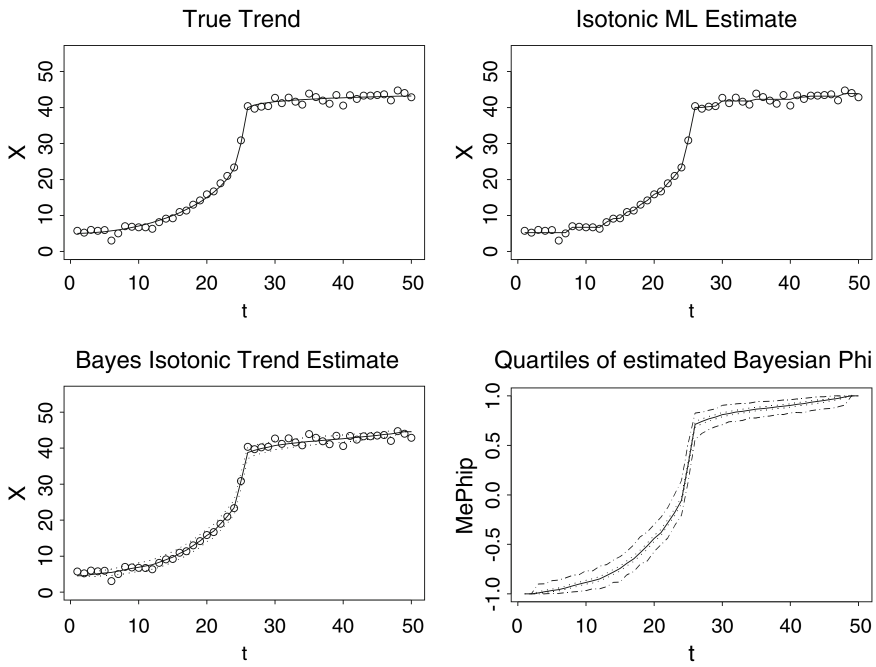

Fig. 1 Bayesian and frequentist estimators on a simulated example

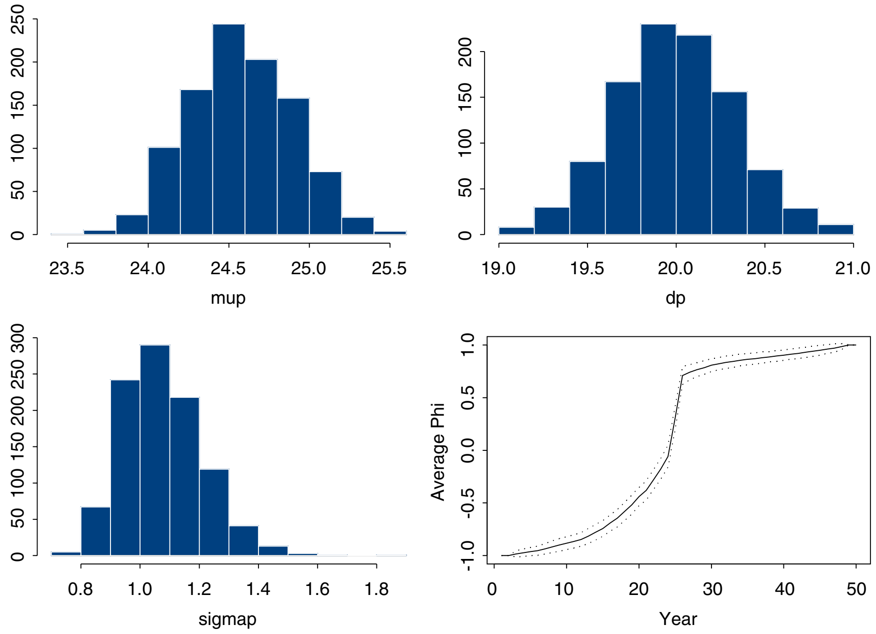

Fig. 2 Posterior densities and ϕ with confidence bands

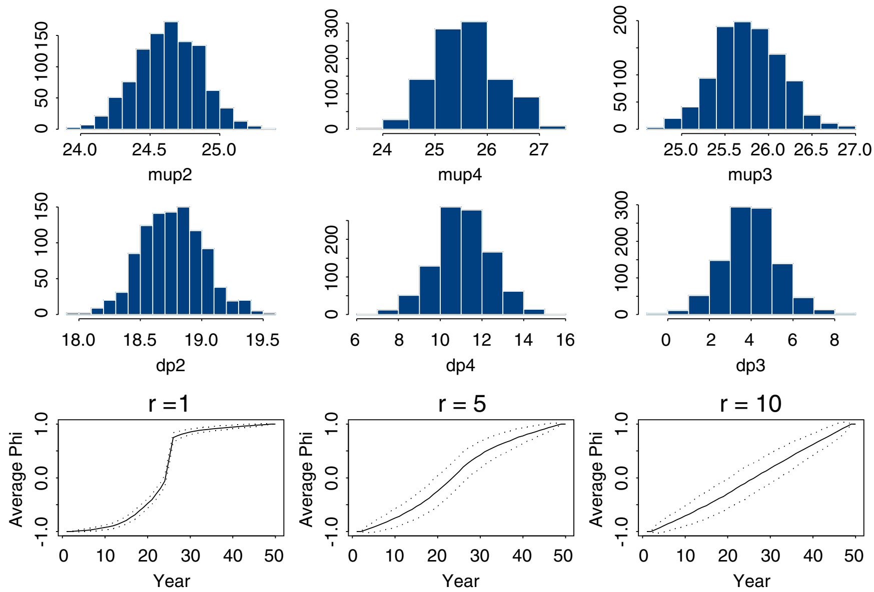

Fig. 3 Bayesian estimators with different "penalties" r

closer to a straight line. We also notice that increasing r has the effect of widening the confidence bands about $\hat{\boldsymbol{\phi}}$. This is natural and inconsequential. To see why, consider the limiting case of $\delta = 0$. In such a case, there is an issue of identifiability because the distribution of $[\mathcal{X}, \mu, \sigma^2, \boldsymbol{\phi}]$ is independent of $\boldsymbol{\phi}$, i.e., any pattern of trend change ($\boldsymbol{\phi} = \phi_1, \ldots, \phi_n$) is consistent with the data. Intuitively speaking, as δ gets smaller the experiment becomes "less identifiable" and this is reflected in wider confidence bands for $\boldsymbol{\phi}$. Equivalently, we notice that the *shape* of the trend change is going to be easier to estimate when there was a large change between endpoints of the observation window.

4 Global warming

We consider the time-honored global warming dataset provided by Jones et. al. (see http://cdiac.esd.ornl.gov/trends/temp/jonescru/jones.html) containing annual temperature anomalies from 1858 to 2000, expressed in degrees Celsius and are relative to the 1961–1990 mean. These data are presented in Fig. 4, together with the Bayesian estimate. We chose a noninformative prior on the form of the trend and we fit a model with parameters $\sigma_\mu = 100$ and $a = b = 1$, and $r = 1$. Even though the global warming data, being a time series, might be affected by serial correlation e.g., Fomby and Vogelsang (2002), we opted in this paper for simplicity to ignore that aspect of the data and model it as a sequence of i.i.d. observations. Extensions of the Bayesian isotonic method to non-i.i.d. setups are being presently studied and it will be published in follow up articles.

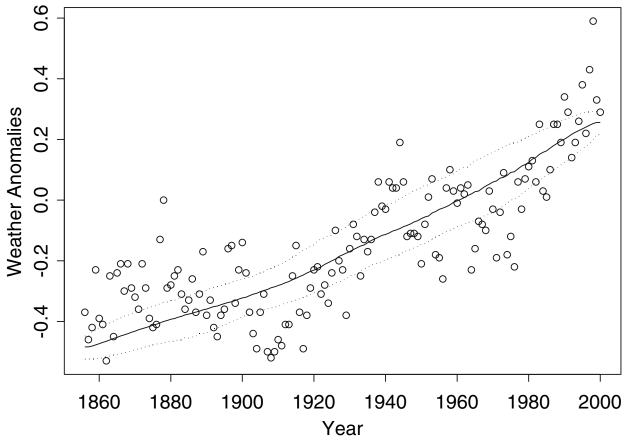

Fig. 4 World annual weather anomalies 1958–2000

Table 1 Posterior inference for global warming data

Parameter	Mean	Median	SD	95% credible interval	
μ	−0.1145	−0.1141	0.0196	−0.1543	−0.0783
σ	0.1761	0.1756	0.0109	0.1563	0.1989
δ	0.3132	0.3115	0.0279	0.2621	0.3700
ϕ_1	−1.0000	−1.0000	0.0000	−1.0000	−1.0000
ϕ_{10}	−0.9125	−0.9146	0.0273	−0.9584	−0.8505
ϕ_{20}	−0.8077	−0.8070	0.0360	−0.8732	−0.7371
ϕ_{30}	−0.7112	−0.7079	0.0395	−0.8133	−0.6412
ϕ_{40}	−0.6138	−0.6135	0.0397	−0.6996	−0.5389
ϕ_{50}	−0.5072	−0.5085	0.0368	−0.5697	−0.4384
ϕ_{60}	−0.3848	−0.3879	0.0483	−0.4718	−0.2809
ϕ_{70}	−0.2268	−0.2324	0.0626	−0.3395	−0.1001
ϕ_{80}	−0.0607	−0.0713	0.0676	−0.1851	0.0894
ϕ_{90}	0.0931	0.0715	0.0756	−0.0412	0.2369
ϕ_{100}	0.2354	0.2261	0.0849	0.1025	0.3896
ϕ_{110}	0.3875	0.3830	0.0897	0.2357	0.5638
ϕ_{120}	0.5542	0.5551	0.0818	0.4167	0.7151
ϕ_{130}	0.7473	0.7468	0.0577	0.6379	0.8481
ϕ_{140}	0.9387	0.9426	0.0293	0.8729	0.9839
ϕ_{145}	1.0000	1.0000	0.0000	1.0000	1.0000

Note: $\boldsymbol{\phi}$ shown only partially, one every 10

Table 1 contains the inferences from the posterior distributions, which confirm that there is a trend increase of about 0.3°C Celsius in global annual temperature between 1858 and 2000. Figure 5 exhibits the histograms of the posterior distributions for μ, σ

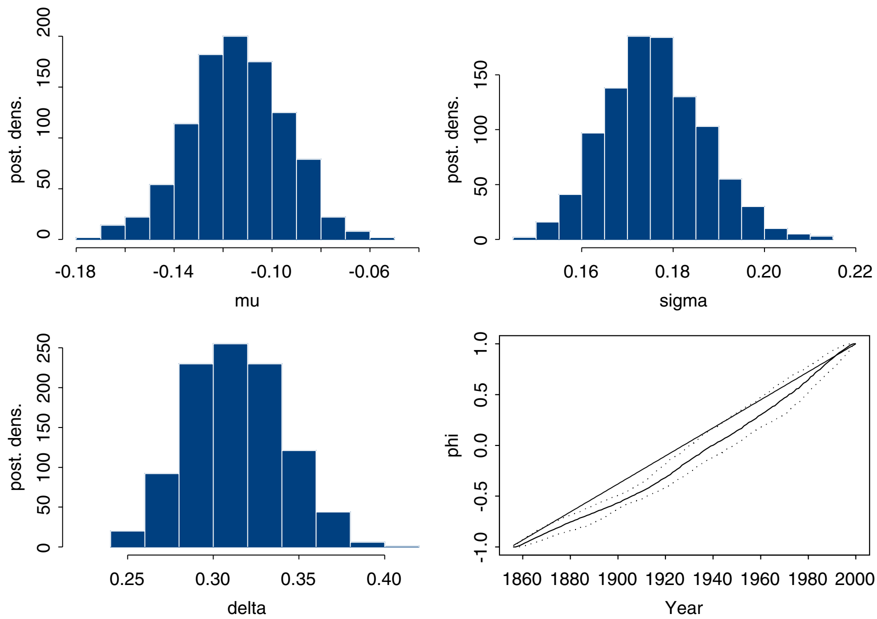

Fig. 5 Posterior estimates for world annual weather anomalies 1958–2000

Table 2 Sensitivity to hyper-parameters for the global warming data

	Global warming			
r	0.00001	1	0.00001	0.00001
a	10	1	1	20
b	2	1	1	4
τ	2	100	100	10
$\hat{\mu}_{.05}$	−0.1627	−0.1300	−0.1537	−0.1609
$\hat{\mu}_{.50}$	−0.1323	−0.1118	−0.1271	−0.1268
$\hat{\mu}_{.95}$	−0.1019	−0.0940	−0.1077	−0.0930
$\hat{d}_{.05}$	0.2664	0.2725	0.2695	0.2587
$\hat{d}_{.50}$	0.3150	0.3173	0.3159	0.3216
$\hat{d}_{.95}$	0.3645	0.3664	0.3599	0.3838
Convex	No	Yes	No	Yes

and δ, together with the estimated function $\boldsymbol{\phi}$. Interestingly, the 45° line is above the confidence bands for $\boldsymbol{\phi}$. This indicates that the change-point pattern is convex, i.e., not only the trend but also the *rate of trend-change* in weather anomalies is increasing in time. As already discussed in the introduction, the availability of such a qualitative conclusion about the shape of the trend change is the major strength of the Bayesian method. In Table 2 we perform a sensitivity analysis by choosing a few diferent sets of hyperparameters. The results confirm the increase of global temperature by about

Fig. 6 Argentina rainfall data

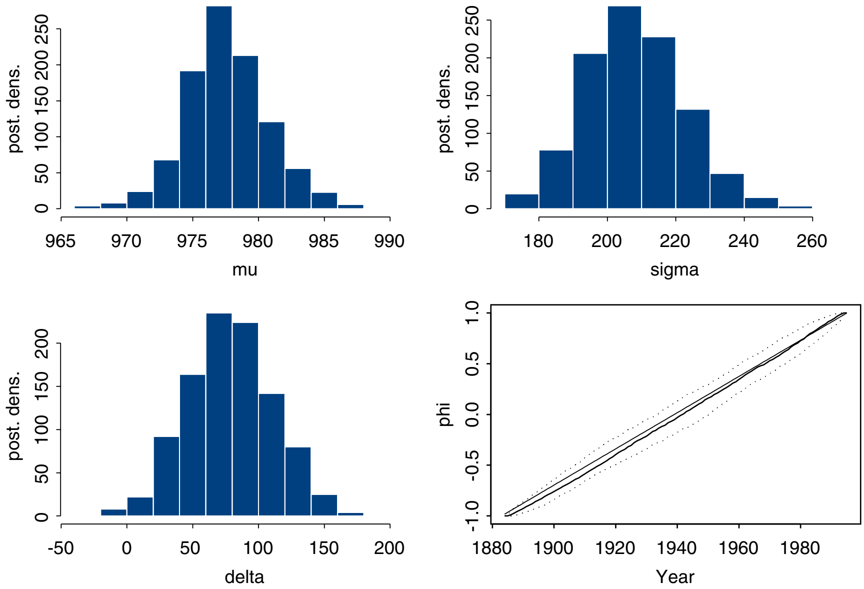

Fig. 7 Posterior estimates for the Argentina rainfall dataset

0.30°C throughout the series, as $\hat{d}_{0.50}$ remains fairly stable for the different choices of hyper-parameters. However, it is not possible to assert that there was acceleration in the pattern of global warming. That is because for some choices of the hyper-parameters the estimated $\boldsymbol{\phi}$ is not convex.

Table 3 Posterior inference for argentina rainfall dataset

Parameter	Mean	Median	SD	95% credible interval	
μ	977.6547	977.4000	3.1913	971.5213	984.5020
σ	208.0292	207.3307	14.4468	181.9659	237.7382
δ	78.4810	78.2309	33.1194	17.1106	141.1022
ϕ_1	−1.0000	−1.0000	0.0000	−1.0000	−1.0000
ϕ_{10}	−0.8755	−0.8781	0.0372	−0.9399	−0.7932
ϕ_{20}	−0.7003	−0.7082	0.0524	−0.7823	−0.5665
ϕ_{30}	−0.5173	−0.5310	0.0619	−0.6023	−0.3611
ϕ_{40}	−0.3327	−0.3420	0.0733	−0.4482	−0.1672
ϕ_{50}	−0.1542	−0.1572	0.0733	−0.2878	−0.0101
ϕ_{60}	0.0279	0.0288	0.0802	−0.1373	0.1921
ϕ_{70}	0.2143	0.2116	0.0740	0.0684	0.3500
ϕ_{80}	0.4057	0.4079	0.0692	0.2733	0.5495
ϕ_{90}	0.5915	0.5807	0.0678	0.4684	0.7317
ϕ_{100}	0.7917	0.7949	0.0585	0.6740	0.8996
ϕ_{110}	0.9806	0.9866	0.0193	0.9280	0.9996
ϕ_{112}	1.0000	1.0000	0.0000	1.0000	1.0000

Note: $\boldsymbol{\phi}$ shown only partially, one every 10

5 Argentina rainfall data

The so called *Argentina rainfall data* presented in Fig. 6 consists of yearly rainfall volume in the northwestern argentine province of Tucumán, from 1884 to 1995 (e.g., Wu et al. 2001). To carry out Bayesian Inference we have selected the following values of the hyperparameters: $r = 1$, $a = 0.5$, $b = 1,000$, $\sigma_\mu^2 = 10^6$ and $\alpha_i = 1$ for all i; this is, except for the *precision parameter* $\lambda = 1/\sigma^2$, the prior specification is noninformative. The values of a and b are chosen so that the average standard deviation of the rainfall is about 200 and its standard deviation is about 300 (i.e., it is believed the total annual rainfall cannot differ for much more than 500 ml between years).

The posterior distributions of μ, δ, and σ are shown in Fig. 7, where the first three plots are the histograms, while the last panel shows the average of the sampled $\boldsymbol{\phi}$s. The descriptive values for posterior inference are presented in Table 3.

We can immediately state the following conclusions:

1. There is indeed an increase in the trend somewhere along the series so that the value of the trend at the end is estimated to be about 80 ml above its level at the beginning of the series.
2. There is no indication of nonlinearity or discontinuity in the pattern of trend change. In contrast with global warming, the 45° line here is very close to the estimated $\boldsymbol{\phi}$ and lies well within the confidence bands. This is as expected since, as seen in Fig. 6, no trend change at all is apparent.

Table 4 Sensitivity to hyper-parameters for the Argentina rainfall data

	Argentina rainfall			
r	1	1	0.00001	0.00001
a	0.5	1	1	0.5
b	1000	1	1	1000
τ	1000000	100	100	1000000
$\hat{\mu}_{.05}$	968.31	928.11	928.02	974.72
$\hat{\mu}_{.50}$	976.48	938.33	937.92	980.56
$\hat{\mu}_{.95}$	979.92	946.58	946.23	987.47
$\hat{d}_{.05}$	21.06	20.55	29.50	30.11
$\hat{d}_{.50}$	75.53	75.56	85.94	87.66
$\hat{d}_{.95}$	130.84	133.09	139.83	140.51

In Table 4 we show the corresponding sensitivity analysis for a few choices of the hyper-parameters, which confirms the conclusions.

Appendix: Numerical algorithm for the Gibbs sampler

Preliminary estimators: In order to propose initial values of the parameters μ^*, δ^*, $(\sigma^2)^*$, $\boldsymbol{\phi}^*$ to initialize the Markov chain Monte Carlo (MCMC), consider maximizing the log-likelihood function (i.e., the log of the density of $[\mathcal{X}|\mu, \delta, \sigma, \boldsymbol{\phi}]$) given by

$$l(\mu, \delta, \sigma^2, \boldsymbol{\phi}) = \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu - \delta\phi_i)^2\right\} \tag{16}$$

for $\boldsymbol{\phi} \in \Omega_\phi$. It is maximized simultaneously at

$$\mu^* = \frac{1}{n} \sum_{i=1}^{n} (X_i - \delta^* \phi_i^*) \tag{17}$$

$$\delta^* = \frac{\sum_{i=1}^{n} \phi_i^* (X_i - \mu^*)}{\sum_{i=1}^{n} (\phi_i^*)^2 + r^2} \tag{18}$$

$$(\sigma^2)^* = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu^* - \delta^* \phi_i^*)^2, \tag{19}$$

and with $\boldsymbol{\phi}^*$ being the isotonization of the vector with components $(X_i - \mu^*)/\delta^*$, re-scaled so that $\phi_1^* = -1$ and $\phi_n^* = 1$. Because the system of equations provided by (17–19) and $\boldsymbol{\phi}^*$ is not fully recursive, we obtain them numerically in two steps:

1. Set as initial estimators $\mu^0 = (\sum_{i=1}^n X_i)/n$ and $\phi_i^0 = 2(i/n) - 1$.
2. Obtain the next iterate by plugging current estimates into the MLE system.
3. Iterate until the value of $l(\mathcal{X}|\mu, \delta, \sigma, \boldsymbol{\phi})$ fails to increase above some small threshold. Take those final estimators as μ^*, δ^*, $(\sigma^2)^*$ and $\boldsymbol{\phi}^*$.

Gibbs Sampler: It consists of the following steps:

1. Take μ^*, δ^*, $(\sigma^2)^*$ and $\boldsymbol{\phi}^*$ as initial draws.
2. Obtain the draws of μ, δ and σ^2 by sampling from (11)–(13) plugging in the current draws as parameters.
3. To sample from the posterior of $\boldsymbol{\phi}$, not that each of its components is, conditional on the rest, a truncated normal variate, i.e.,

$$[\phi_i|\mathcal{X}, \mu, \delta, \sigma^2, (\phi_j, j \neq i)] \sim N\left(\frac{x_i - \mu}{\delta}, \frac{\sigma^2}{\delta^2}\right)\bigg|_{(\phi_{i-1}, \phi_{i+1})}. \tag{20}$$

   Thus we sample $\boldsymbol{\phi}$ by iterating draws from (20) in a random order.
4. Iterate for 10,000 times and discard the first 5,000 observations (burn it). Select for each parameter an independent sample of size 1,000 from the remaining 5,000 observations.

MCMC diagnostics. In order to check the convergence of the Gibbs sampler, two alternative Markov chains are run, based on the following two different sets of starting points:

1. Chain 2:

$$\mu^{**} = \tfrac{1}{n}\sum_{i=1}^n X_i, \qquad \phi_i^{**} = 2\frac{i}{n} - 1, \quad 1 < i < n,$$

$$\delta^{**} = \frac{\sum_{i=1}^n \phi_i^{**}(X_i - \mu^{**})}{\sum_{i=1}^n (\phi_i^{**})^2 + r^2}, \quad (\sigma^2)^{**} = \frac{1}{n}\sum_{i=1}^n (X_i - \mu^* - \delta\phi_i^*)^2.$$

2. Chain 3: μ^{***}, δ^{***}, $(\sigma^2)^{***}$ and $\boldsymbol{\phi}^{***}$ are just random draws from the priors (6)–(10).

Acknowledgments The authors wish to thank the anonymous referees for carefully reading the paper and suggesting important contributions for improvement.

References

Bhattacharya, P. K. (1994). Some aspects of change-point analysis. In: Carlstein, E., Müller, H., Siegmund, D. (Eds.), *Change-point problems*, (pp. 28–56), Hayward, CA: Institute of Mathematical Statistics.

Carlin, B. P., Gelfand, A. E., Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics*, *2*, 389–405.

Chernoff, H., Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Annals of Mathematical Statistics*, *35*, 999–1018.

Dey, D.K., Purkayastha, S. (1997). Bayesian approach to change-point problems. *Communications in Statistics, Theory and Methods*, *26*, 2035–2047.

Gelfand, A. E., Smith, A. F., Lee, T. M. (1992). Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, *87*(418), 523–532.

Fomby, T. B., Vogelsang, T. J. (2002). The application of size-robust trend statistics to global-warming temperature series. *Journal of Climate*, *15*(1), 117–123.

Hinkley, D. V., Chapman, P., Rungel, G. (1980). Changepoint problems, Technical Report 382. University of Minessota, Minneapolis, MN.

Robertson, T., Wright, F. T., Dyskra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley

Shiryayev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, *8*, 22–46.

Siegmund, D. (1986). Boundary crossing probabilities and classical applications. *Annals of Statistics*, *14*, 361–404.

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

Wu, W. B., Woodroofe, M., Mentz, G. B. (2001). Isotonic regression: another look at the change-point problem. *Biometrika*, *88*, 793–804.

Zacks, S. (1983). Survey of classical and Bayesian approaches to the changepoint problem: fixed sample and sequential procedures or testing and estimation. In *Recent advances in Statistics, Herman Chernoff Festschrift*, (pp. 245–269). New York: Academic.