Jaromír Antoch; David Legát

Application of MCMC to change point detection

Applications of Mathematics, Vol. 53 (2008), No. 4, 281–296

Persistent URL: `http://dml.cz/dmlcz/140321`

53 (2008) APPLICATIONS OF MATHEMATICS No. 4, 281–296

APPLICATION OF MCMC TO CHANGE POINT DETECTION*

Jaromír Antoch, Praha, David Legát, Praha

(Supplement to the special issue of Appl. Math. *53* (2008), No. 3)

Abstract. A nonstandard approach to change point estimation is presented in this paper. Three models with random coefficients and Bayesian approach are used for modelling the year average temperatures measured in Prague Klementinum. The posterior distribution of the change point and other parameters are estimated from the random samples generated by the combination of the Metropolis-Hastings algorithm and the Gibbs sampler.

Keywords: change point estimation, Markov chain Monte Carlo (MCMC), Metropolis-Hastings algorithm, Gibbs sampler, Bayesian statistics, Klementinum temperature series

MSC 2010: 65C40, 62F40, 65C05

1. Introduction

Change point detection is a topic of interest of many applied and theoretical statisticians since the seventies. This paper presents a less usual approach to the change point estimation, i.e., we use Bayesian statistics and MCMC approach assuming statistical models with random parameters for modelling the data. Construction of the models is illustrated on the analysis of well known temperature series from Klementinum. Markov chains are generated from the posterior distribution of these parameters using MCMC and desired results are derived from the chains obtained.

MCMC simulation methods cover several powerful algorithms for generation of Markov chains with desired stationary distribution. Metropolis-Hastings algorithm and Gibbs sampler belong among the best known and most often used methods for this purpose. Statistical background of these two methods is presented in the monograph [14], while the monograph [6] emphasizes the practice and MCMC applications.

* The work is a part of the research project MSM 0021620839 financed by MŠMT. It was partially supported by grant GAČR 201/06/0186 (for the first author) and grant GAUK 135007-B (for the second author).

Concerning the monographs and survey papers on the change point problem, we refer to [1]–[5] among others.

2. Analyzed data

Klementinum is a historical building in the central part of Prague. Meteorological station, where the temperature has been measured since 1775, is placed in one of its towers. For our analysis we had available year averages of these measurements during

Astronomical tower of Klementinum (P. Příhoda, pen-and-ink drawing, 2008).

the period from 1775 till 1992. This sequence, consisting of $N = 218$ observations, is analyzed with the goal to detect a possible change in the model, e.g. shift in the mean etc. Members of the analyzed sequence, which are displayed in Fig. 1, will be denoted by $Z_1, \ldots, Z_N$ in the sequel.

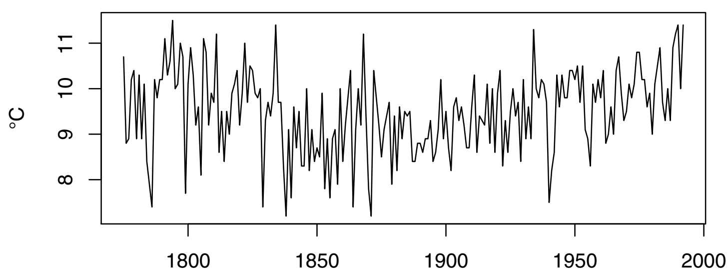

Figure 1. Klementinum mean year temperatures.

Klementinum data were analyzed by many authors. Relevant to our work are, among others, the papers [9] and [10].

3. Models

Throughout the paper we consider three models with random parameters suitable for description of our observations. Moreover, we assume for simplicity that there is only one change in each of them. These models differ in the shape of expected value of the sequence $Z_1, \ldots, Z_N$. In the first model the mean is constant both before and after the change point. In the second and third models the form is linear, the difference being that the second model allows for a jump in the change point while in the third model we consider a gradual change. Moreover, in all models we assume that $Z_1, \ldots, Z_N$ are independent normally distributed random variables with a constant variance σ^2.

In our application to the Klementinum data the assumption about the normal distribution of Z_i's is not violated, which, unfortunately, is not the case concerning the independence of the observations. This dependence is mainly caused by the trend which is inherited in the nature of the data, not so much by the dependence between the consecutive years as can be seen when the data are detrended and only then the autocorrelation function is calculated.

3.1. Model 1: Piecewise constant expected value

In this model we suppose that the expected value of Z_i is constant, being the same for the first r observations but different for the last $N - r$ observations. Exact form of the model is

$$Z_i \sim \begin{cases} N(\mu_1, \sigma^2), & 1 \leqslant i \leqslant r, \\ N(\mu_2, \sigma^2), & r < i \leqslant N, \end{cases} \tag{3.1}$$

where r, μ_1, μ_2 and σ^2 are unknown random parameters to be estimated. As said above, Bayesian approach and MCMC are used to estimate them. Notice that the parametrization $\gamma = 1/\sigma^2$ simplifies the form of the results. Therefore, instead of (3.1) we will use the notation $Z_i \sim N(\mu_i, 1/\gamma)$, $i = 1, 2$.

The first problem is to choose a prior distribution for the vector of unknown parameters. It is supposed that r, μ_1, μ_2 and γ are independent random variables. To increase the chance of finding more change points and not to roam over a single possibility, as often happens when random walk algorithm is used, we used for the parameter r a discrete uniform distribution on the set $\{1, \ldots, 217\}$, denoted throughout by $R\{1, \ldots, 217\}$, and

$$\mathcal{L}(\mu_1) \sim N(\nu_1, \xi_1^2), \quad \mathcal{L}(\mu_2) \sim N(\nu_2, \xi_2^2) \quad \text{and} \quad \mathcal{L}(\gamma) \sim Ga(1, 1).$$

Concerning the parameters μ_1 and μ_2, in simulations we put $\nu_1 = \nu_2 = 9.5$ and $\xi_1^2 = \xi_2^2 = 1$, i.e. we used the values corresponding to the data under the (null) hypothesis of no change. As concerns the distribution of γ, it is usual to use gamma distribution, for details see [13]. Then the density of the prior distribution is $f(\mu_1, \mu_2, \gamma, r) \propto \exp\{-\frac{1}{2}(\mu_1 - \nu_1)^2/\xi_1^2 - \frac{1}{2}(\mu_2 - \nu_2)^2/\xi_2^2 - \gamma\}$, where the notation $f(x) \propto g(x)$ expresses the fact that there exists a constant K such that $f(x) = Kg(x)$. Such choice of the prior distribution is common for linear models with normal residuals and leads to a "nice" form of the posterior distribution in this case. The meaning of the word "nice" is explained in Section 4. The likelihood function of the sequence $Z_1, \dots, Z_N$ for given values of parameters is

$$f(z_1, \dots, z_N \mid \mu_1, \mu_2, \gamma, r) \propto \gamma^{N/2} \exp\Bigg(-\frac{\gamma}{2}\Bigg[\sum_{i=1}^{r}(z_i - \mu_1)^2 + \sum_{i=r+1}^{N}(z_i - \mu_2)^2\Bigg]\Bigg),$$

and finally, the Bayes theorem provides us with the posterior distribution

$$f_{\mu_1,\mu_2,\gamma,r \mid \boldsymbol{z}}$$
$$\propto \gamma^{N/2} \exp\Bigg(-\frac{(\mu_1 - \nu_1)^2}{2\xi_1^2} - \frac{(\mu_2 - \nu_2)^2}{2\xi_2^2} - \gamma - \frac{\gamma}{2}\Bigg[\sum_{i=1}^{r}(z_i - \mu_1)^2 + \sum_{i=r+1}^{N}(z_i - \mu_2)^2\Bigg]\Bigg),$$

where $\boldsymbol{z}$ stands for $z_1, \dots, z_N$. Notice that we use this shorthand throughout the rest of the paper.

3.2. Model 2: Two-phase linear model with a jump

The second model is a natural generalization of the first, i.e., linear trend of the expected value is allowed in both parts of the sequence of observations, leading to

$$Z_i \sim \begin{cases} N(\alpha_1 + \beta_1 i, 1/\gamma), & 1 \leqslant i \leqslant r, \\ N(\alpha_2 + \beta_2(i - r), 1/\gamma), & r < i \leqslant N, \end{cases} \tag{3.2}$$

where α_1, α_2, β_1, β_2, γ and r are unknown random parameters to be estimated. We fix the prior distribution analogously to the first model, i.e.

$$\begin{aligned} &\mathcal{L}(\alpha_1) \sim N(\nu_1, \xi_1^2), \qquad &&\mathcal{L}(\alpha_2) \sim N(\nu_2, \xi_2^2), \qquad &&\mathcal{L}(r) \sim R\{1, \dots, 217\}, \\ &\mathcal{L}(\beta_1) \sim N(\eta_1, \zeta_1^2), \qquad &&\mathcal{L}(\beta_2) \sim N(\eta_2, \zeta_2^2), \qquad &&\mathcal{L}(\gamma) \sim Ga(1, 1). \end{aligned}$$

Concerning the parameters α_1, α_2, β_1 and β_2, in simulations we put $\nu_1 = \nu_2 = 9.5$, $\xi_1^2 = \xi_2^2 = 1$, $\eta_1 = \eta_2 = 0$ and $\zeta_1^2 = \zeta_2^2 = 0.1$, i.e. we used the values corresponding to the data under the (null) hypothesis of no change.

Analogously as in Model 1 we can derive the density of the posterior distribution, which has the form

$$
\begin{aligned}
&f_{\alpha_1,\alpha_2,\beta_1,\beta_2,\gamma,r\,|\,\boldsymbol{z}}\\
&\quad\propto \gamma^{N/2}\exp\Big(-\Big[\frac{(\alpha_1-\nu_1)^2}{2\xi_1^2}+\frac{(\alpha_2-\nu_2)^2}{2\xi_2^2}+\frac{(\beta_1-\eta_1)^2}{2\zeta_1^2}+\frac{(\beta_2-\eta_2)^2}{2\zeta_2^2}+\gamma\Big]\Big)\\
&\qquad\qquad\times\exp\bigg(-\frac{\gamma}{2}\bigg[\sum_{i=1}^{r}(z_i-\alpha_1-\beta_1 i)^2+\sum_{i=r+1}^{N}(z_i-\alpha_2-\beta_2(i-r))^2\bigg]\bigg).
\end{aligned}
$$

3.3. Model 3: Two-phase linear model with a gradual change

The third model is similar to the second. We just replace the parameter α_2 by $\alpha_1+\beta_1 r$. This constraint ensures that the assumed evolution of expected values is "continuous" at the change point r. This model has the form

$$
Z_i \sim \begin{cases} N(\alpha_1+\beta_1 i, 1/\gamma), & 1 \leqslant i \leqslant r,\\ N(\alpha_1+\beta_1 r+\beta_2(i-r), 1/\gamma), & r < i \leqslant N. \end{cases} \tag{3.3}
$$

Very interesting parametric approach concerning the estimation of the change point in this model is described in [8].

The prior distributions for the parameters α_1, β_1, β_2, γ and r are the same as in Model 2, parameter α_2 has been "cancelled". Analogously as in Model 1 we can derive the density of the posterior distribution, which has the form

$$
\begin{aligned}
&f_{\alpha_1,\beta_1,\beta_2,\gamma,r\,|\,\boldsymbol{z}}\\
&\quad\propto \gamma^{N/2}\exp\Big(-\Big[\frac{(\alpha_1-\nu_1)^2}{2\xi_1^2}+\frac{(\beta_1-\eta_1)^2}{2\zeta_1^2}+\frac{(\beta_2-\eta_2)^2}{2\zeta_2^2}+\gamma\Big]\Big)\\
&\quad\times\exp\bigg(-\frac{\gamma}{2}\bigg[\sum_{i=1}^{r}(z_i-\alpha_1-\beta_1 i)^2+\sum_{i=r+1}^{N}(z_i-\alpha_1-\beta_1 r-\beta_2(i-r))^2\bigg]\bigg).
\end{aligned}
$$

4. MCMC simulations

In the previous sections we outlined three models, fixed the prior distribution for their parameters and derived the density of the appropriate posterior distribution for each of them. These posterior distributions are examined closely in this section. The goal is to retrieve answers to the following questions:

- What is the shape of the marginal density of the change point r?
- Is it unimodal or multimodal?

- Where are local maxima of this density located?
- What is the conditional distribution of other parameters when the change point r is expected to fall between some prescribed values r_1 and r_2?

As mentioned above, MCMC simulation methods belong to possible approaches enabling us to solve these tasks. They are based on generation of random samples from a posterior distribution and estimation of desired probabilities and characteristics from these samples.

Each of the following subsections is dedicated to one model. Readers will find a description of the Gibbs sampler combined with Metropolis-Hastings algorithm, which were used to generate the samples. Aside from that, we summarize forms of necessary conditional distributions and acceptance probabilities as well.

4.1. Model with a piecewise constant expected value

The MCMC algorithm for the first model generates each observation of the desired random sample in four steps. Suppose we have already generated the sample $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}$, where $\boldsymbol{x}^{(i)}$ denotes the vector $(\mu_1^{(i)}, \mu_2^{(i)}, \gamma^{(i)}, r^{(i)})$. Then the vector $\boldsymbol{x}^{(n+1)}$ is generated as follows:

(1) Generate a candidate r' for new value of the parameter r from $R\{1, \ldots, 217\}$.

(2) Accept the candidate r' from step 1 with an appropriate probability, which will be specified later; i.e. $r^{(n+1)} = r'$ if accepted else $r^{(n+1)} = r^{(n)}$.

(3) Generate new values $\mu_1^{(n+1)}$ and $\mu_2^{(n+1)}$ from the conditional distribution $f_{\mu_1,\mu_2\,|\,\gamma,r,\boldsymbol{z}}$, where the values $\gamma = \gamma^{(n)}$ and $r = r^{(n+1)}$ are given.

(4) Generate a new value $\gamma^{(n+1)}$ from the conditional distribution $f_{\gamma\,|\,\mu_1,\mu_2,r,\boldsymbol{z}}$, where the values $\mu_1 = \mu_1^{(n+1)}$, $\mu_2 = \mu_2^{(n+1)}$ and $r = r^{(n+1)}$ are given.

Several computations have to be carried out before launching this algorithm. First, we must derive the conditional distributions for steps 3 and 4. Second, we must determine an appropriate acceptance probability in step 2. The density $f_{\mu_1,\mu_2\,|\,\gamma,r,\boldsymbol{z}}$ has the form

$$f_{\mu_1,\mu_2\,|\,\gamma,r,\boldsymbol{z}} \propto \exp\left(-\left[\frac{(\mu_1-\nu_1)^2}{2\xi_1^2} + \frac{\gamma}{2}\sum_{i=1}^{r}(z_i-\mu_1)^2\right]\right) \times \exp\left(-\left[\frac{(\mu_2-\nu_2)^2}{2\xi_2^2} + \frac{\gamma}{2}\sum_{i=r+1}^{N}(z_i-\mu_2)^2\right]\right).$$

Elementary calculations and substitutions $M_1(r) = \sum\limits_{i=1}^{r} z_i$ and $M_2(r) = \sum\limits_{i=r+1}^{N} z_i$ show that $f_{\mu_1,\mu_2\,|\,\gamma,r,\boldsymbol{z}}$ is the density of a two-dimensional normal distribution with the mean $\boldsymbol{\mu}$ and the variance matrix $\boldsymbol{\Sigma}$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \dfrac{\nu_1 + \gamma\xi_1^2 M_1(r)}{1+\gamma\xi_1^2 r} \\ \dfrac{\nu_2 + \gamma\xi_2^2 M_2(r)}{1+\gamma\xi_2^2 (N-r)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \dfrac{\xi_1^2}{1+\gamma\xi_1^2 r} & 0 \\ 0 & \dfrac{\xi_2^2}{1+\gamma\xi_2^2 (N-r)} \end{pmatrix}.$$

The density $f_{\gamma\,|\,\mu_1,\mu_2,r,\boldsymbol{z}}$ can be calculated analogously as above and we get

$$f_{\gamma\,|\,\mu_1,\mu_2,r,\boldsymbol{z}} \propto \gamma^{N/2} \exp\biggl(-\gamma\biggl[1 + \frac{1}{2}\sum_{i=1}^{r}(z_i-\mu_1)^2 + \frac{1}{2}\sum_{i=r+1}^{N}(z_i-\mu_2)^2\biggr]\biggr),$$

i.e. $f_{\gamma\,|\,\mu_1,\mu_2,r,\boldsymbol{z}}$ is the density of a gamma distribution with the shape parameter $\frac{1}{2}N+1$ and the scale parameter $\Bigl(1 + \frac{1}{2}\Bigl[\sum_{i=1}^{r}(z_i-\mu_1)^2 + \sum_{i=r+1}^{N}(z_i-\mu_2)^2\Bigr]\Bigr)^{-1}$.

The only parameter which we generate using the Metropolis-Hastings algorithm is the change point r. Notice that we do not use the Gibbs sampler due to the fact that the conditional distribution $f_{r\,|\,\mu_1,\mu_2,\gamma,\boldsymbol{z}}$ is too complex and none of the software available to us provides sampling from such a distribution. Therefore, we always generate a candidate for r and this is accepted with some probability in the next step. The formal way how to fix this acceptance probability has been derived in [7], so that we just state the right form of it. More precisely, we use $\alpha(\boldsymbol{x}^{(n)}, \boldsymbol{x}') = \min(1, \beta(\boldsymbol{x}^{(n)}, \boldsymbol{x}'))$, where

$$\beta(\boldsymbol{x}^{(n)}, \boldsymbol{x}') = \frac{f_{\mu_1,\mu_2,\gamma,r\,|\,\boldsymbol{z}}(\boldsymbol{x}')}{f_{\mu_1,\mu_2,\gamma,r\,|\,\boldsymbol{z}}(\boldsymbol{x}^{(n)})}, \tag{4.1}$$

$\boldsymbol{x}^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \gamma^{(n)}, r^{(n)})$ denotes the previous solution and $\boldsymbol{x}' = (\mu_1^{(n)}, \mu_2^{(n)}, \gamma^{(n)}, r')$ denotes the new candidate solution. It means that the vectors $\boldsymbol{x}^{(n)}$ and $\boldsymbol{x}'$ differ in the last component only, where the new candidate r' replaced the change point $r^{(n)}$ from the last iteration.

Substituting the posterior density to the equation (4.1), we obtain the final form of the acceptance probability $\beta(\boldsymbol{x}^{(n)}, \boldsymbol{x}')$ equal to

$$\begin{aligned}\exp\biggl(\frac{\gamma^{(n)}}{2}\biggl[\sum_{i=1}^{r^{(n)}}(z_i-\mu_1^{(n)})^2 + \sum_{i=r^{(n)}+1}^{N}(z_i-\mu_2^{(n)})^2 - \sum_{i=1}^{r'}(z_i-\mu_1^{(n)})^2 \\ - \sum_{i=r'+1}^{N}(z_i-\mu_2^{(n)})^2\biggr]\biggr),\end{aligned}$$

which can be equivalently written in the form

$$\beta(\boldsymbol{x}^{(n)}, \boldsymbol{x}') = \begin{cases} \exp\left(\dfrac{\gamma^{(n)}}{2}\left[\displaystyle\sum_{i=r^{(n)}+1}^{r'} (z_i - \mu_2^{(n)})^2 - \sum_{i=r^{(n)}+1}^{r'} (z_i - \mu_1^{(n)})^2\right]\right), \\ \quad r' > r^{(n)}, \\ \exp\left(\dfrac{\gamma^{(n)}}{2}\left[\displaystyle\sum_{i=r'+1}^{r^{(n)}} (z_i - \mu_1^{(n)})^2 - \sum_{i=r'+1}^{r^{(n)}} (z_i - \mu_2^{(n)})^2\right]\right), \\ \quad r' \leqslant r^{(n)}. \end{cases}$$

4.2. Two-phase linear model with a jump

The second model is characterized by six parameters, which results in a bit more complicated MCMC algorithm. Each new observation is generated in five steps:

(1) Generate a candidate r' for the new value of the parameter r from $R\{1, \ldots, 217\}$.
(2) Accept the candidate r' from step 1 with a probability $\alpha(\boldsymbol{x}^{(n)}, \boldsymbol{x}')$ that will be specified later; i.e. $r^{(n+1)} = r'$ if accepted else $r^{(n+1)} = r^{(n)}$.
(3) Generate $\alpha_1^{(n+1)}$ and $\alpha_2^{(n+1)}$ from the conditional distribution $f_{\alpha_1,\alpha_2 \mid \beta_1,\beta_2,\gamma,r,\boldsymbol{z}}$, where the values $\beta_1 = \beta_1^{(n)}$, $\beta_2 = \beta_2^{(n)}$, $\gamma = \gamma^{(n)}$ and $r = r^{(n+1)}$ are given.
(4) Generate $\beta_1^{(n+1)}$ and $\beta_2^{(n+1)}$ from the conditional distribution $f_{\beta_1,\beta_2 \mid \alpha_1,\alpha_2,\gamma,r,\boldsymbol{z}}$, where the values $\alpha_1 = \alpha_1^{(n+1)}$, $\alpha_2 = \alpha_2^{(n+1)}$, $\gamma = \gamma^{(n)}$ and $r = r^{(n+1)}$ are given.
(5) Generate $\gamma^{(n+1)}$ from the conditional distribution $f_{\gamma \mid \alpha_1,\alpha_2,\beta_1,\beta_2,r,\boldsymbol{z}}$, where the values $\alpha_1 = \alpha_1^{(n+1)}$, $\alpha_2 = \alpha_2^{(n+1)}$, $\beta_1 = \beta_1^{(n+1)}$, $\beta_2 = \beta_2^{(n+1)}$ and $r = r^{(n+1)}$ are given.

It is clear that we must calculate the conditional densities

$$f_{\alpha_1,\alpha_2 \mid \beta_1,\beta_2,\gamma,r,\boldsymbol{z}}, \quad f_{\beta_1,\beta_2 \mid \alpha_1,\alpha_2,\gamma,r,\boldsymbol{z}}, \quad f_{\gamma \mid \alpha_1,\alpha_2,\beta_1,\beta_2,r,\boldsymbol{z}},$$

and the acceptance probability $\alpha(\boldsymbol{x}^{(n)}, \boldsymbol{x}')$ for step 2. As the calculations are similar to those presented in Subsection 4.1, we present only the results. For better transparency of the resulting formulas, the following notation is used:

$$M_{1\alpha}(r, \beta_1) = \sum_{i=1}^{r} (z_i - \beta_1 i), \qquad M_{1\beta}(r, \alpha_1) = \sum_{i=1}^{r} i(z_i - \alpha_1),$$
$$M_{2\alpha}(r, \beta_2) = \sum_{i=r+1}^{N} (z_i - \beta_2(i - r)), \quad M_{2\beta}(r, \alpha_2) = \sum_{i=r+1}^{N} (i - r)(z_i - \alpha_2).$$

The function $f_{\alpha_1,\alpha_2 \mid \beta_1,\beta_2,\gamma,r,\boldsymbol{z}}$ is the density of a two-dimensional normal distribution with the mean $\boldsymbol{\mu}_\alpha$ and the variance matrix $\boldsymbol{\Sigma}_\alpha$, where

$$\boldsymbol{\mu}_\alpha = \begin{pmatrix} \dfrac{\nu_1 + \gamma\xi_1^2 M_{1\alpha}(r,\beta_1)}{1+\gamma\xi_1^2 r} \\ \dfrac{\nu_2 + \gamma\xi_2^2 M_{2\alpha}(r,\beta_1)}{1+\gamma\xi_2^2(N-r)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_\alpha = \begin{pmatrix} \dfrac{\xi_1^2}{1+\gamma\xi_1^2 r} & 0 \\ 0 & \dfrac{\xi_2^2}{1+\gamma\xi_2^2(N-r)} \end{pmatrix}.$$

The function $f_{\beta_1,\beta_2 \mid \alpha_1,\alpha_2,\gamma,r,\boldsymbol{z}}$ is the density of a two-dimensional normal distribution with the mean $\boldsymbol{\mu}_\beta$ and the variance matrix $\boldsymbol{\Sigma}_\beta$, where

$$\boldsymbol{\mu}_\beta = \begin{pmatrix} \dfrac{\eta_1 + \gamma\zeta_1^2 M_{1\beta}(r,\alpha_1)}{1+\gamma\zeta_1^2 G(r)} \\ \dfrac{\eta_2 + \gamma\zeta_2^2 M_{2\beta}(r,\alpha_2)}{1+\gamma\zeta_2^2 G(N-r)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_\beta = \begin{pmatrix} \dfrac{\zeta_1^2}{1+\gamma\zeta_1^2 G(r)} & 0 \\ 0 & \dfrac{\zeta_2^2}{1+\gamma\zeta_2^2 G(N-r)} \end{pmatrix},$$

where $G(k) = \sum\limits_{i=1}^{k} i^2$.

Finally, the function $f_{\gamma \mid \alpha_1,\alpha_2,\beta_1,\beta_2,r,\boldsymbol{z}}$ is the density of a Gamma distribution with the shape parameter $\frac{1}{2}N+1$ and the scale parameter

$$\left(1 + \frac{1}{2}\left[\sum_{i=1}^{r}(z_i - \alpha_1 - \beta_1 i)^2 + \sum_{i=r+1}^{N}(z_i - \alpha_2 - \beta_2(i-r))^2\right]\right)^{-1}.$$

The acceptance probability for a new candidate change point r', or $\boldsymbol{x}'$, is $\alpha(\boldsymbol{x}^{(n)},\boldsymbol{x}') = \min(1, \beta(\boldsymbol{x}^{(n)},\boldsymbol{x}'))$, where

$$\beta(\boldsymbol{x}^{(n)},\boldsymbol{x}') = \begin{cases} \exp\Bigg(\dfrac{\gamma}{2}\Bigg[\sum\limits_{i=r^{(n)}+1}^{r'}(z_i - \alpha_2 - \beta_2(i-r^{(n)}))^2 \\ \qquad - \sum\limits_{i=r^{(n)}+1}^{r'}(z_i - \alpha_1 - \beta_1 i)^2\Bigg]\Bigg), \quad r' > r^{(n)}, \\ \exp\Bigg(\dfrac{\gamma}{2}\Bigg[\sum\limits_{i=r'+1}^{r^{(n)}}(z_i - \alpha_1 - \beta_1 i)^2 \\ \qquad - \sum\limits_{i=r'+1}^{r^{(n)}}(z_i - \alpha_2 - \beta_2(i-r^{(n)}))^2\Bigg]\Bigg), \quad r' \leqslant r^{(n)}. \end{cases}$$

4.3. Two-phase linear model with gradual change

The third model involves parameters α_1, β_1, β_2, γ and r. The MCMC algorithm, which generates a random sample from the posterior distribution in this model, repeats the following six steps:

(1) Generate a candidate r' for the new value of the parameter r from $R\{1,\ldots,217\}$.

(2) Accept a candidate r' from step 1 with probability $\alpha(\boldsymbol{x}^{(n)},\boldsymbol{x}')$, that will be specified later; i.e. $r^{(n+1)} = r'$ if accepted else $r^{(n+1)} = r^{(n)}$.

(3) Generate a new value $\alpha_1^{(n+1)}$ from the conditional distribution $f_{\alpha_1\,|\,\beta_1,\beta_2,\gamma,r,\boldsymbol{z}}$, where the values $\beta_1 = \beta_1^{(n)}$, $\beta_2 = \beta_2^{(n)}$, $\gamma = \gamma^{(n)}$ and $r = r^{(n+1)}$ are given.

(4) Generate a new value $\beta_1^{(n+1)}$ from the conditional distribution $f_{\beta_1\,|\,\alpha_1,\beta_2,\gamma,r,\boldsymbol{z}}$, where the values $\alpha_1 = \alpha_1^{(n+1)}$, $\beta_2 = \beta_2^{(n)}$, $\gamma = \gamma^{(n)}$ and $r = r^{(n+1)}$ are given.

(5) Generate a new value $\beta_2^{(n+1)}$ from the conditional distribution $f_{\beta_2\,|\,\alpha_1,\beta_1,\gamma,r,\boldsymbol{z}}$, where the values $\alpha_1 = \alpha_1^{(n+1)}$, $\beta_1 = \beta_1^{(n+1)}$, $\gamma = \gamma^{(n)}$ and $r = r^{(n+1)}$ are given.

(6) Generate a new value $\gamma^{(n+1)}$ from the conditional distribution $f_{\gamma\,|\,\alpha_1,\beta_1,\beta_2,r,\boldsymbol{z}}$, where the values $\alpha_1 = \alpha_1^{(n+1)}$, $\beta_1 = \beta_1^{(n+1)}$, $\beta_2 = \beta_2^{(n+1)}$ and $r = r^{(n+1)}$ are given.

The function $f_{\alpha_1\,|\,\beta_1,\beta_2,\gamma,r,\boldsymbol{z}}$ is the density of a normal distribution with the mean μ_α and the variance σ_α^2, where

$$\mu_\alpha = \frac{\nu_1 + \gamma\xi_1^2\Big[\sum\limits_{i=1}^{r}(z_i - \beta_1 i) + \sum\limits_{i=r+1}^{N}\big(z_i - \beta_1 r - \beta_2(i-r)\big)\Big]}{1+\gamma\xi_1^2 N}, \quad \sigma_\alpha^2 = \frac{\xi_1^2}{1+\gamma\xi_1^2 N}.$$

The function $f_{\beta_1\,|\,\alpha_1,\beta_2,\gamma,r,\boldsymbol{z}}$ is the density of a normal distribution with the mean μ_{β_1} and the variance $\sigma_{\beta_1}^2$, where

$$\mu_{\beta_1} = \frac{\eta_1 + \gamma\zeta_1^2\Big[\sum\limits_{i=1}^{r} i(z_i - \alpha_1) + r\sum\limits_{i=r+1}^{N}(z_i - \alpha_1 - \beta_2(i-r))\Big]}{1+\gamma\zeta_1^2 G(r) + \gamma\zeta_1^2 r^2(N-r)},$$

$$\sigma_{\beta_1}^2 = \frac{\zeta_1^2}{1+\gamma\zeta_1^2 G(r) + \gamma\zeta_1^2 r^2(N-r)}, \quad \text{where again} \quad G(k) = \sum_{i=1}^{k} i^2.$$

The function $f_{\beta_2\,|\,\alpha_1,\beta_1,\gamma,r,\boldsymbol{z}}$ is the density of a normal distribution with the mean μ_{β_2} and the variance $\sigma_{\beta_2}^2$, where

$$\mu_{\beta_2} = \frac{\eta_2 + \gamma\zeta_2^2\sum\limits_{i=r+1}^{N}(i-r)(z_i - \alpha_1 - \beta_1 r)}{1+\gamma\zeta_2^2 G(N-r)} \quad \text{and} \quad \sigma_{\beta_2}^2 = \frac{\zeta_2^2}{1+\gamma\zeta_2^2 G(N-r)}.$$

Finally, the function $f_{\gamma\,|\,\alpha_1,\beta_1,\beta_2,r,\boldsymbol{z}}$ is the density of a Gamma distribution with the shape parameter $\frac{1}{2}N+1$ and the scale parameter

$$\left(1 + \frac{1}{2}\left[\sum_{i=1}^{r}(z_i - \alpha_1 - \beta_1 i)^2 + \sum_{i=r+1}^{N}(z_i - \alpha_1 - \beta_1 r - \beta_2(i-r))^2\right]\right)^{-1}.$$

The acceptance probability for a new candidate change point r', or $\boldsymbol{x}'$, is $\alpha(\boldsymbol{x}^{(n)}, \boldsymbol{x}') = \min(1, \beta(\boldsymbol{x}^{(n)}, \boldsymbol{x}'))$, where

$$\beta(\boldsymbol{x}^{(n)}, \boldsymbol{x}') = \begin{cases} \exp\Bigg(\dfrac{\gamma^{(n)}}{2}\Bigg[\displaystyle\sum_{i=r^{(n)}+1}^{r'} (z_i - \alpha_1 - \beta_1 r^{(n)} - \beta_2(i - r^{(n)}))^2 \\ \qquad - \displaystyle\sum_{i=r^{(n)}+1}^{r'} (z_i - \alpha_1 - \beta_1 i)^2\Bigg]\Bigg), \quad r' > r^{(n)}, \\ \exp\Bigg(\dfrac{\gamma^{(n)}}{2}\Bigg[\displaystyle\sum_{i=r'+1}^{r^{(n)}} (z_i - \alpha_1 - \beta_1 i)^2 \\ \qquad - \displaystyle\sum_{i=r'+1}^{r^{(n)}} (z_i - \alpha_1 - \beta_1 r^{(n)} - \beta_2(i - r^{(n)}))^2\Bigg]\Bigg), \quad r' \leqslant r^{(n)}. \end{cases}$$

5. Results

All algorithms described in Section 4 were implemented using the system Mathematica and Matlab. For each algorithm we generated several Markov chains, starting points being chosen from an over-dispersed distribution on the set of possible values of parameters. All these chains seemed to be stationary from their 500th member. Therefore, for each model a Markov chain with 10^4 members has been generated with the starting point in the expected value of the parameters (under the prior). Desired characteristics of a posterior distribution were estimated from the last 9000 members of such a chain. Cutting off the first 1000 members ensures that we are using the chain which is close to the stationary process. This ensures that our estimates are independent of the starting point. The proportion of the accepted changes of the parameter r is 11.5% in Model 1, 2% in Model 2 and 17.5% in Model 3. Concerning Model 2, it is necessary to take into account that the new r selected from $R\{1, \ldots, 217\}$ is most typically rejected. Notice that the increase of the number of members of the MCMC chain did not change the results. Conclusions derived from the generated chains are presented in the following subsections.

5.1. Model with a piecewise constant expected value

First of all, we were interested in the posterior distribution of the parameter r. You can find the kernel estimates of its density in Fig. 2. It is evident that the mass of this density is concentrated between the years 1915 and 1992. The most probable year for a change point in this model is 1943. The probability that the

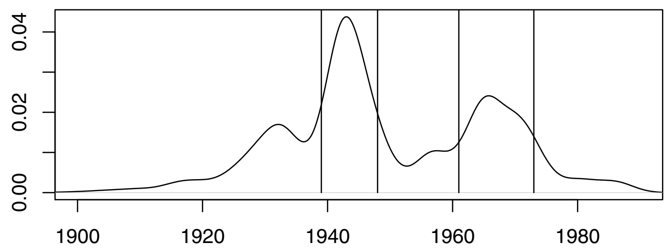

Figure 2. Kernel density estimator of the parameter $r + 1774$.

change point lies between 1939 and 1948 is approximately 0.36. Another important period is 1961–1973 with the probability approximately 0.27.

Suppose, moreover, that the change occurred within the period 1939–1948, which corresponds to $165 \leqslant r \leqslant 174$. We were interested in the values of other parameters, especially μ_1 and μ_2. If we estimate the conditional expected value $\mathbf{E}[\mu_1 | 165 \leqslant r \leqslant 174]$ by the average $\sum_{i \in C} \mu_1^{(i)} / \#C$, $\mathbf{E}[\mu_2 | 165 \leqslant r \leqslant 174]$ by $\sum_{i \in C} \mu_2^{(i)} / \#C$, where C is a set of the indexes $1000 < i \leqslant 10000$ such that $165 \leqslant r^{(i)} \leqslant 174$, then the estimated values are $\hat{\mu}_1 = 9.36$ and $\hat{\mu}_2 = 9.98$. Corresponding model is displayed in Fig. 3.

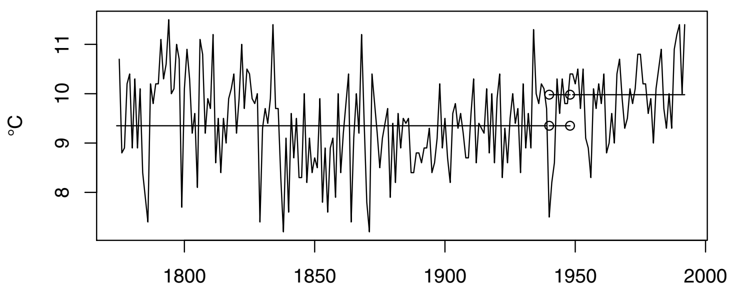

Figure 3. Data and estimated Model 1.

5.2. Two-phase linear model with a jump

The posterior distribution of the parameter r in the second model is concentrated close to the value 62. The frequencies of some particular values in the sequence $\{r^{(i)}, 1000 < i \leqslant 10000\}$ are shown in Tab. 1.

r	60	61	62	63	64	65	else
year	1834	1835	1836	1837	1838	1839	
frequency	566	2868	4870	338	0	75	283
relative frequency	0.0629	0.319	0.541	0.038	0	0.008	0.031

Table 1. Frequency table of sequence $r^{(i)}$.

Contrary to Model 1, MCMC was used to estimate the change point r only, leading to $\hat{r} = 62$ that corresponds to the year 1836. Parameters α_1, α_2, β_1, β_2 and γ were estimated after splitting the data in two parts, i.e. $Z_1, \ldots, Z_{\hat{r}}$ and $Z_{\hat{r}+1}, \ldots, Z_N$, separately in each part by the linear regression, giving

$$\hat{\alpha}_1 = 9.717, \quad \hat{\alpha}_2 = 8.58, \quad \hat{\beta}_1 = 0.0022, \quad \hat{\beta}_2 = 0.0102 \quad \text{and} \quad \hat{\gamma} = 1.539.$$

Corresponding model, i.e. the regression lines corresponding to these estimators, are displayed in Fig. 4.

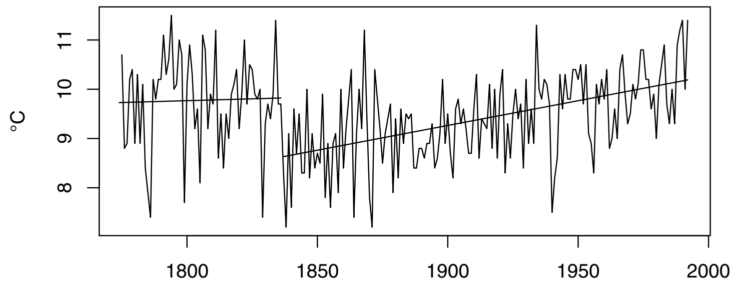

Figure 4. Data and estimated Model 2.

We obtained the model in which the first part is practically constant, the growth in the first 61 years is only 0.15 °C which corresponds to the increase 0.0022 °C per year. After a dramatic decrease of temperature in the year 1836 we observe a steady growth of temperature from 8.7 °C to 10.2 °C in the second part of our model, which corresponds to the increase 0.01 °C per year.

This model seems to be, at least by naked eyes, more appropriate than the first model. Indeed, notice that there exist several periods in Model 1, e.g. 1840–1860, where almost all observations lie below (or above) the fitted curve, being not the case when fitting Model 2. The change point r seems to be located more reasonably than in Model 1. Notice, moreover, that we have obtained practically the same result, i.e. the same placement of the change point r, as if we would use classical approach of minimizing the residual sum of squares (RSS).

5.3. Two-phase linear model with gradual change

The distribution of the change point r in Model 3 is displayed in Fig. 5. We see the maximum between the years 1850–1860. Assuming that $r = 1855$ is the true change point, the estimated parameters of the model (3.3) and RSS values are given in Tab. 2 and the fitted model is plotted in Fig. 7.

Aside that, we assumed that the true change point r is located in any of the years 1780–1990. For each of these locations of r we estimated parameters of the

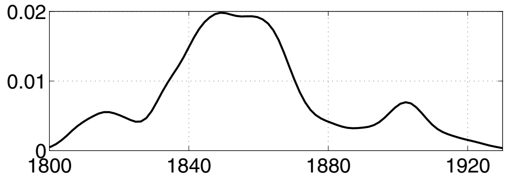

Figure 5. Kernel density estimator of the parameter $r + 1774$.

year	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	RSS
1855	10.15	-0.016	0.009	151.96
1882	9.97	-0.010	0.012	150.99

Table 2. Estimates of $\alpha_1, \beta_1, \beta_2$ and corresponding RSS.

model (3.3) and calculated corresponding RSS. Plot of the values of RSS in Fig. 6 shows that the models for the years 1850–1890 are practically equivalent and that our Bayesian solution coincides with the beginning of this period. Finally, we calculated the "best model" minimizing RSS for $r \in (1780, 1990)$ according to [8]. This "optimal" model corresponds to the year 1882. The estimated parameters and the corresponding RSS value are given in Tab. 2, the fitted model is plotted in Fig. 7.

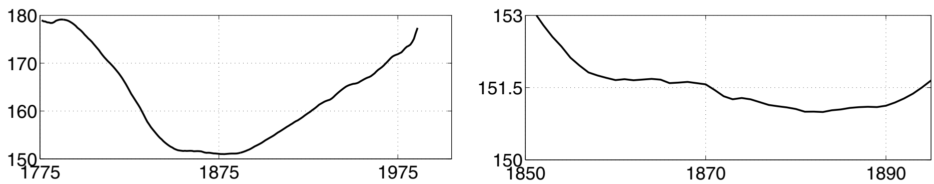

Figure 6. RSS for all possible changes and its detail for the years 1850–1895.

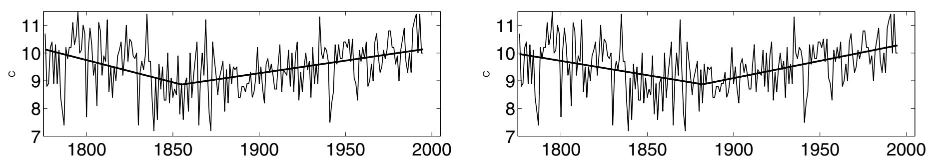

Figure 7. Data and estimated Model 3 for $r = 1855$ (left) and $r = 1882$ (right).

6. Conclusions

The sequence of average year temperatures measured in Klementinum has already been analyzed by many other statisticians, see [2], [9] or [10] among others for different approaches. It is evident that different authors used different methods. We can say that we obtained similar results as the authors of these papers, of course, using different methodology. Personally, from the models described in this paper we prefer Model 2.

The MCMC simulation methodology, which was used here, is a quite general tool. On the other hand, one disadvantage, which was revealed in our (simple) models, should be mentioned, namely, it is necessary to carry out many (routine on one hand but often tedious and not necessarily simple on the other one) computations before MCMC algorithms can be applied.

Notice, finally, that we have developed also the MCMC algorithms comprising more than one change. Evidently, the algorithms are more complicated and the resulting densities are more "ugly". However, they were developed along the same lines as the algorithms presented in this paper, so that from the methodological point of view they do not bring something surprisingly new.

References

[1] *J. Antoch, M. Hušková, D. Jarušková*: Off-line statistical process control. In: Multivariate Total Quality Control, Chapter 1. Physica-Verlag/Springer, Heidelberg, 2002, pp. 1–86.

[2] *J. Antoch, M. Hušková*: Estimators of changes. Asymptotics, Nonparametrics, and Time Series. Marcel Dekker, Basel, 1999, pp. 533–577.

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

[4] *B. P. Carlin, A. E. Gelfand, A. F. M. Smith*: Hierarchical Bayesian analysis of change point problems. Appl. Stat. *41* (1992), 389–405.

[5] *M. Csörgő, L. Horváth*: Limit Theorems in Change-Point Analysis. J. Wiley & Sons, New York, 1997.

[6] *W. R. Gilks, S. Richardson, D. J. Spiegelhalter* (eds.): Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, London, 1995.

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

[8] *D. V. Hinkley*: Inference about the intersection in two-phase regression. Biometrika *56* (1969), 495–504.

[9] *M. Janžura, J. Nielsen*: Segmentation method and change-point problem. ROBUST'02 (J. Antoch, G. Dohnal, J. Klaschka, eds.). JČMF, Praha, pp. 163–177. (In Czech.)

[10] *D. Jarušková*: Some problems with application of change point detection methods to enviromental data. Environmetrics *8* (1997), 469–483.

[11] *D. Legát*: MCMC methods. Master thesis. Charles University, Praha, 2004. (In Czech.)

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

[13] *A. O'Hogan, J. Foster*: Kendall's Advanced Theory of Statistics, Bayesian Inference. Arnold, London, 1999.

[14] *Ch. P. Robert, G. Casella*: Monte Carlo Statistical Methods, 2nd ed. Springer, Heidelberg, 2005.

Authors' address: *J. Antoch*, *D. Legát*, Charles University in Prague, Faculty of Mathematics and Physics, Department of Statistics, Sokolovská 83, 186 75 Praha 8-Karlín, Czech Republic, e-mail: `jaromir.antoch@mff.cuni.cz`, `david.legat@mff.cuni.cz`.