Bayesian Segmentation of Piecewise Constant Autoregressive Processes using MCMC methods

O. Punska - C. Andrieu - A.Doucet
W.J. Fitzgerald

CUED/F-INFENG/TR.344

Bayesian Segmentation of Piecewise Constant Autoregressive Processes using MCMC methods

O. Punska - C. Andrieu*- A. Doucet†- W.J. Fitzgerald

Technical Report CUED/F-INFENG/TR.344
Signal Processing Laboratory, Department of Engineering
University of Cambridge, Trumpington St., CB2 1PZ Cambridge, UK.
Email: {op205, ca226, ad2, wjf}@eng.cam.ac.uk

Abstract

In this paper we address the problem of the Bayesian segmentation of signals which are modelled as piecewise constant autoregressive (AR) processes excited by white Gaussian noise. We define a probabilistic model that takes into account the fact that the number of AR processes needed to represent the data, the model orders, the values of the parameters and the noise variances are unknown. The resulting posterior probability distributions and Bayesian estimators of the model parameters do not admit closed-form analytical expression. A stochastic algorithm based on a reversible jump Markov chain Monte Carlo (MCMC) method [11] is derived to estimate these quantities. Results are obtained for both synthetic and real data (a speech signal which has been examined in the literature previously [1], [4] and [14]) and confirm the applicability of both the model and the algorithm.

*C Andrieu is sponsored by AT&T Laboratories, Cambridge, UK.
†A.Doucet is sponsored by EPSRC, UK.

Contents

1 Introduction

The problem of segmentation is fundamental to many areas of data and image analysis. The process involves dividing a large sequence of data into small homogeneous segments, the boundaries of which may be interpreted as changes in the physical system. This approach has proved extremely useful for different practical problems arising in recognition-oriented signal processing, such as continuous speech processing, biomedical and seismic signal processing, monitoring of industrial processes, etc. Not surprisingly, the task is of great practical and theoretical interest, which is reflected in a large number of surveys. For example, the problem of automatic analysis of continuous speech signals is addressed in [1]; segmentation algorithms for recognition-oriented geophysical signals are described in [2]; and an application of the changepoint detection method to an electroencephalogram (EEG) is presented in [15].

Of course, different authors propose various approaches to the problem of detection of abrupt changes and, in particular, segmentation. This issue is thoroughly surveyed in [4], where different methods are proposed and an exhaustive list of references is given. Since then, several contributions have been made to the field of changepoint theory. For example, the General Piecewise Linear Model and its extension to study multiple changepoints in non-Gaussian impulsive noise environments is introduced in [16], segmentation in a linear regression framework is investigated in [12] and [14], and a general segmentation method suitable for both parametric and nonparametric models is described in [15]. The main goal of these last approaches and, indeed, [8] is the use of the maximum a posteriori (MAP), or maximum-likelihood (ML), estimate. According to [13], this technique eliminates some shortcomings of the Generalised Likelihood Ratio (GLR) test (see [13], [14] for discussion), introduced in [24] and widely used in segmentation in the 1980s (see [1], [3], [4]). Some approaches to solve the problem of multiple changepoint detection in a Bayesian framework, using Markov Chain Monte Carlo (MCMC) [19], are also presented in [2] and [21].

In [1], [4], [15] it is also shown that the algorithms designed for signals modelled as piecewise constant autoregressive (AR) processes excited by white Gaussian noise, have proved useful for processing real signals, such as speech, seismic and EEG data. In all these cases the order of AR model was the same for different segments and was chosen by the user. However, in practice, there are numerous applications (speech processing, for example) where different model orders should be considered for different segments. Thus, not only the number of segments, but the correct model orders for each of them should be estimated. To the best of our knowledge, this joint detection/estimation problem has never been addressed before and in this paper a new methodology to solve it is proposed.

We refer here to the problem of retrospective changepoint detection, assuming that all the data is available at a same time; and follow a Bayesian approach whereby the unknown parameters, including the number of AR processes needed to represent the data, the model orders, the values of the parameters, and noise variances for each segment are regarded as random quantities with known prior distributions. Moreover, some of the hyperparameters are considered random as well and drawn from the appropriate hyperprior distribution [18], whereas they are usually tuned heuristically by the user (see [14], [15]). In general, the main difficulty of the Bayesian approach is that the resulting posterior distribution appears highly non-linear in its parameters, thus precluding analytical calculations. The case treated here is even more complex. Indeed, since the number of changepoints and the orders of the models are assumed random, the posterior distribution is defined on a finite disconnected union of subspaces of various dimensions. Each subspace corresponds to a model with a fixed number of changepoints and fixed model order for each segment. This is a complicated problem of "double" model selection and in order to solve it, we propose an efficient stochastic algorithm based on Markov chain Monte Carlo (MCMC) methods, and a reversible jump MCMC method [11] in particular. This algorithm allows evaluation of some of the posterior features of interest such as marginal distributions. Once they are estimated, model selection can be performed using the marginal maximum a posteriori (MMAP) criterion.

The paper is organised as follows: the model of the signal is given in Section 2; in Section 3, we propose a hierarchical Bayesian model and state the estimation objectives. As mentioned above, this model implies that the posterior distribution and the associated Bayesian estimators do not admit any closed-form expression. Therefore, in order to perform estimation, a reversible jump MCMC algorithm (see [11]), is developed in Section 4. The results for both synthetic and real data (a speech signal examined in the literature previously, see [1], [4] and [14]) are presented in Section 5 and confirm the good performance of both the model and the algorithm when used in practice. In Section 7 some conclusions are drawn. The notation used throughout the paper is described in Appendix.

2 Problem Statement

Let $\mathbf{x}_{0:T-1} \triangleq (x_0, x_1, \ldots, x_{T-1})^{\mathrm{T}}$ be a vector of T observations. The elements of $\mathbf{x}_{0:T-1}$ may be represented by one of the models $\mathcal{M}_{k,\mathbf{p}_k}$, corresponding to the case when the signal is modelled as an AR process with piecewise constant parameters and k ($k = 0, \ldots, k_{\max}$) changepoints. More precisely:

$$\mathcal{M}_{k,\mathbf{p}_k}: \quad x_t = \mathbf{a}_{i,k}^{(p_{i,k})\mathrm{T}} \mathbf{x}_{t-1:t-p_{i,k}} + n_t \quad \text{for } \tau_{i,k} \le t < \tau_{i+1,k}, \quad i = 0, \ldots, k, \tag{1}$$

where a set of $p_{i,k}$ model parameters ($p_{i,k} = 0, \ldots, p_{\max}$, $\mathbf{p}_k \triangleq \mathbf{p}_{1:k,k}$) for the i^{th} segment under the assumption of k changepoints in the signal is arranged in the vector $\mathbf{a}_{i,k}^{(p_{i,k})} = \left(a_{i,k,1}^{(p_{i,k})}, \ldots, a_{i,k,p_{i,k}}^{(p_{i,k})}\right)^{\mathrm{T}}$ and n_t is i.i.d. Gaussian noise of variance $\sigma_{i,k}^2$ ($\boldsymbol{\sigma}_k^2 \triangleq \boldsymbol{\sigma}_{1:k,k}^2$) associated with this AR model. The changepoints of the model $\mathcal{M}_{k,\mathbf{p}_k}$ are denoted $\boldsymbol{\tau}_k \triangleq \boldsymbol{\tau}_{1:k,k}$ and we adopt the convention $\tau_{0,k} = 0$ and $\tau_{k+1,k} = T - 1$ for notational convenience.

The models can be rewritten in the following matrix form:

$$\mathcal{M}_{k,\mathbf{p}_k}: \quad \mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1} = \mathbf{X}_{i,k}^{(p_{i,k})} \mathbf{a}_{i,k}^{(p_{i,k})} + \mathbf{n}_{\tau_{i,k}:\tau_{i+1,k}-1}, \quad i = 0, \ldots, k, \tag{2}$$

where $\mathbf{X}_{i,k}^{(p_{i,k})}$ for the i^{th} segment $(i = 0, ..., k)$ is given by:

$$\mathbf{X}_{i,k}^{(p_{i,k})} = \begin{bmatrix} x_{\tau_{i,k}-1} & x_{\tau_{i,k}-2} & \cdots & x_{\tau_{i,k}-p_{i,k}} \\ x_{\tau_{i,k}} & x_{\tau_{i,k}-1} & \cdots & x_{\tau_{i,k}+1-p_{i,k}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\tau_{i+1,k}-2} & x_{\tau_{i+1,k}-3} & \cdots & x_{\tau_{i+1,k}-1-p_{i,k}} \end{bmatrix}. \quad (3)$$

We interpret the first $p_{\max}$ samples as the initial conditions and proceed with analysis on the remaining $T - p_{\max}$ data points.

We assume that the number of changepoints k and the associated parameters $\mathbf{\Psi}_k \triangleq \left(\boldsymbol{\tau}_k, \mathbf{p}_k, \{\mathbf{a}_{i,k}^{(p_{i,k})}\}_{i=0,...,k}, \boldsymbol{\sigma}_k^2\right)$ are unknown. Given $\mathbf{x}_{0:T-1}$, our aim is to estimate k and $\mathbf{\Psi}_k$.

3 Bayesian model and estimation objectives

We follow a Bayesian approach where the unknown parameters $k, \boldsymbol{\tau}_k, \mathbf{p}_k, \{\mathbf{a}_{i,k}^{(p_{i,k})}\}_{i=0,...,k}, \boldsymbol{\sigma}_k^2$ are regarded as random with a known prior that reflects our degree of belief in the different values of these quantities. In order to increase robustness of the prior, an additional level of hyperprior distributions [18] is introduced. Thus, an extended hierarchical Bayesian model is proposed, which allows us to define a posterior distribution on the space of all possible structures of the signal. Subsequently, the detection/estimation aims are specified and, finally, we derive the posterior distribution marginalised with respect to the unknown nuisance parameters.

3.1 Prior distribution

In our case it is natural to introduce a binomial distribution as a prior distribution for the number of changepoints and their positions (see [9], [14], [15] for a similar choice). This implies that:

$$p\left(k, \boldsymbol{\tau}_k \middle| \lambda\right) = \lambda^k \left(1-\lambda\right)^{T-2-k} \mathbb{I}_{\boldsymbol{\Upsilon}_k}(\boldsymbol{\tau}_k), \quad 0 < \lambda < 1, \quad (4)$$

where $\boldsymbol{\Upsilon}_k \triangleq \{\boldsymbol{\tau}_{1:k,k} \in \{1, \ldots, T-2\}^k$ such that $\tau_{1,k} \neq \tau_{2,k} \neq ... \neq \tau_{k,k}\}$. For the model order prior we adopt a truncated Poisson distribution:

$$p\left(p_{i,k} \middle| \theta\right) \propto \frac{\theta^{p_{i,k}}}{p_{i,k}!} \mathbb{I}_{\{0,\ldots,p_{\max}\}}(p_{i,k}), \quad (5)$$

where the mean θ is interpreted as the expected number of poles for the AR model and the normalizing constant is:

$$C_{p_{\max}} = \frac{1}{\sum_{p_{i,k}=0}^{p_{\max}} \frac{\theta^{p_{i,k}}}{p_{i,k}!}}. \quad (6)$$

Furthermore, we assign a normal distribution to the parameters of the AR models

$$\mathbf{a}_{i,k}^{(p_{i.k})} \Big| \sigma_{i,k}^2, \delta_{i,k}^2 \sim \mathcal{N}\left(0, \sigma_{i,k}^2 \delta_{i,k}^2 \mathbf{I}_{p_{i,k}}\right), \quad i = 0, \ldots, k. \quad (7)$$

and a conjugate Inverse-Gamma distribution to the noise variances

$$\sigma_{i,k}^2 \Big| \frac{\nu_0}{2}, \frac{\gamma_0}{2} \sim \mathcal{IG}(\frac{\nu_0}{2}, \frac{\gamma_0}{2}), \quad \nu_0 > 0, \gamma_0 > 0, \quad i = 0, \ldots, k. \quad (8)$$

This choice of prior, given the Gaussian noise model, allows the marginalization of the parameters $\left(\{\mathbf{a}_{i,k}^{(p_{i,k})}\}_{i=0,\dots,k}, \boldsymbol{\sigma}_k^2\right)$ in this case.

The algorithm requires the specification of $\lambda, \theta, \delta_{i,k}^2, \nu_0$ and γ_0. It is clear that these parameters play a fundamental role in the segmentation of signals, and in order to robustify the prior, we propose to estimate $\lambda, \theta, \delta_{i,k}^2, \gamma_0$ from the data (see [17], [18] for a similar approach), i.e. we consider $\lambda, \theta, \delta_{i,k}^2, \gamma_0$ to be random. We assign a vague conjugate Inverse-Gamma distribution to the scale hyperparameter $\delta_{i,k}^2$:

$$\delta_{i,k}^2 \big| \, \alpha_\delta, \beta_\delta \sim \mathcal{IG}(\alpha_\delta, \beta_\delta), \quad i = 0, \dots, k. \tag{9}$$

Moreover, since in our particular case the acceptance ratio for the birth/death of a change-point depends on the hyper-hyperparameter β_δ (see Eq. (30)), we assume that it is also randomly distributed according to a conjugate prior Gamma distribution:

$$\beta_\delta | \, \zeta_\beta, \varkappa_\beta \sim \mathcal{IG}\left(\zeta_\beta, \varkappa_\beta\right). \tag{10}$$

Similarly, we assign a conjugate prior Gamma density to θ :

$$\theta | \, \zeta, \varkappa \sim \mathcal{IG}\left(\zeta, \varkappa\right). \tag{11}$$

We set $\upsilon_0 = 2, \alpha_\delta = 1, \zeta = 1$ and $\zeta_\beta = 1$ to ensure an infinite variance to express ignorance of the value of the parameters and $\varkappa = \epsilon, \varkappa_\beta = \epsilon_\beta$, $(\epsilon, \epsilon_\beta \ll 1)$; and we choose a uniform prior distribution $\lambda \sim \mathcal{U}_{(0,1)}$ and a non-informative improper Jeffreys' prior for γ_0.

Thus, the following hierarchical structure is assumed for the prior of the parameters:

$$\begin{aligned} p\left(k, \boldsymbol{\Psi}_k, \lambda, \theta, \boldsymbol{\delta}_k^2, \gamma_0, \beta_\delta\right) = & \\ & \prod_{i=0}^k \left[p\left(p_{i,k} \middle| \, \theta\right) p\left(\mathbf{a}_{i,k}^{(p_{i,k})} \middle| \, \sigma_{i,k}^2, \delta_{i,k}^2\right) p\left(\sigma_{i,k}^2 \middle| \, \gamma_0\right) p\left(\delta_{i,k}^2 \middle| \, \beta_\delta\right)\right] \\ & \times p\left(k, \boldsymbol{\tau}_k \middle| \, \lambda\right) p\left(\lambda\right) p\left(\theta\right) p\left(\beta_\delta\right) p\left(\gamma_0\right), \end{aligned} \tag{12}$$

which can be visualised with a directed acyclic graph (DAG) as shown in Fig. 1 (for convenience we do not show $\upsilon_0, \alpha_\delta, \zeta, \varkappa, \zeta_\beta, \varkappa_\beta$).

3.2 Bayesian hierarchical model

For our problem, the overall parameter space can be written as a finite union of subspaces $\boldsymbol{\Theta} \triangleq \cup_{k=0}^{k_{\max}} \{k\} \times \boldsymbol{\Upsilon}_k \times \prod_{i=0}^k \boldsymbol{\Phi}_{p_{i,k}} \times \boldsymbol{\Xi}_k$, where $\boldsymbol{\Phi}_{p_{i,k}}$ denotes the space of the parameters $p_{i,k}, \mathbf{a}_{i,k}^{(p_{i,k})}, \sigma_{i,k}^2$ for the i^{th} segment, i.e. $\boldsymbol{\Phi}_0 \triangleq \mathbb{R}^+$, $\boldsymbol{\Phi}_{p_{i,k}} \triangleq \cup_{p_{i,k}=0}^{p_{\max}} \{p_{i,k}\} \times \left(\mathbb{R}^{p_{i,k}} \times \mathbb{R}^+\right)$, $\boldsymbol{\Xi}_k$ denotes the hyperparameter $\left(\lambda, \theta, \boldsymbol{\delta}_k^2, \gamma_0\right)$ and hyper-hyperparameter (β_δ) space, which is given by $\boldsymbol{\Xi}_k \triangleq (0,1) \times \mathbb{R}^+ \times \left(\mathbb{R}^+\right)^k \times \mathbb{R}^+ \times \mathbb{R}^+$, and $k_{\max} = T - 2$.

There is a natural hierarchical structure for this set-up, which we can formalise by modelling the joint distribution of all variables as follows:

$$p\left(k, \boldsymbol{\Psi}_k, \boldsymbol{\xi}_k, \mathbf{x}_{0:T-1}\right) = p\left(k, \boldsymbol{\Psi}_k, \boldsymbol{\xi}_k\right) p\left(\mathbf{x}_{0:T-1} \middle| \, k, \boldsymbol{\Psi}_k\right), \tag{13}$$

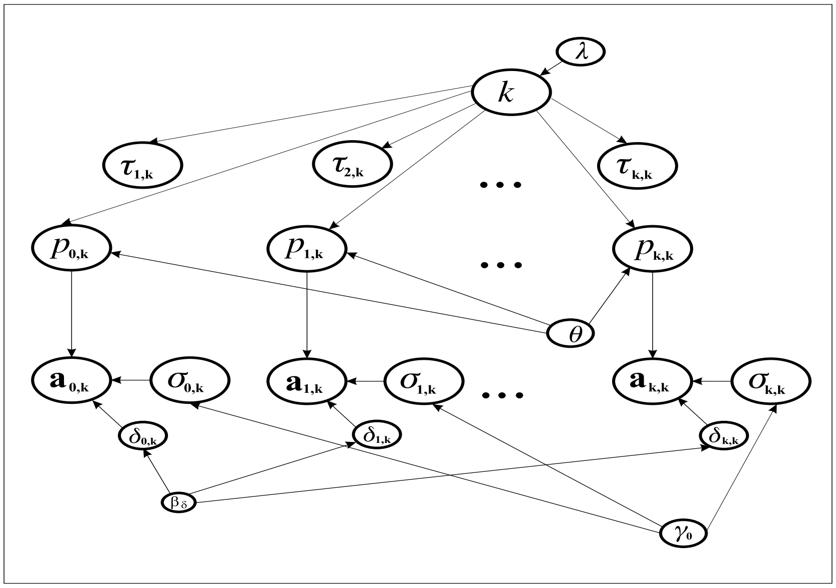

Figure 1: Directed acyclic graph for the prior distribution.

where $\boldsymbol{\xi}_k = \{\lambda, \theta, \boldsymbol{\delta}_k^2, \gamma_0, \beta_\delta\}$. As the excitation is assumed to be i.i.d Gaussian (see Section 2), the likelihood takes the form:

$$p\left(\mathbf{x}_{0:T-1}\right|k, \boldsymbol{\Psi}_k) = \prod_{i=0}^{k} \left(2\pi\sigma_{i,k}^2\right)^{-\frac{\tau_{i+1,k}-\tau_{i,k}}{2}} \exp\left(-\frac{\left(\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1} - \mathbf{X}_{i,k}^{(p_{i,k})}\mathbf{a}_{i,k}^{(p_{i,k})}\right)^{\mathrm{T}}\left(\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1} - \mathbf{X}_{i,k}^{(p_{i,k})}\mathbf{a}_{i,k}^{(p_{i,k})}\right)}{2\sigma_{i,k}^2}\right). \tag{14}$$

This assumption is commonly used one, since the probability distribution that has maximum entropy, subject to knowledge of the first two moments of the noise distribution, is Gaussian, and, therefore, the likelihood of this form is often used unless one has further knowledge concerning the noise statistics. It has been shown to work well in practice and allows the marginalization of the nuisance parameters in our case.

3.3 Bayesian detection and estimation

Any Bayesian inference on k and $\boldsymbol{\Psi}_k, \boldsymbol{\xi}_k$ is based on the following posterior obtained using Bayes' theorem:

$$p\left(k, \boldsymbol{\Psi}_k, \boldsymbol{\xi}_k\right|\mathbf{x}_{0:T-1}) \propto p\left(\mathbf{x}_{0:T-1}\right|k, \boldsymbol{\Psi}_k)\, p\left(k, \boldsymbol{\Psi}_k, \boldsymbol{\xi}_k\right). \tag{15}$$

Our aim is to estimate this posterior distribution, and more specifically some of its features such as the marginal distributions. In our case, however, it is not possible to obtain these quantities analytically, as it requires the evaluation of high-dimensional integrals of nonlinear functions in the parameters (see Section 3.4). Therefore, we apply MCMC methods and a reversible jump MCMC method in particular (see Section 4 for details). The key idea is to build an ergodic Markov chain $\left(k^{(j)}, \mathbf{\Psi}_k^{(j)}, \boldsymbol{\xi}_k^{(j)}\right)_{j\in\mathbb{N}}$ whose equilibrium distribution is the desired posterior distribution. Under weak additional assumptions, the $P \gg 1$ samples generated by the Markov chain are asymptotically distributed according to the posterior distribution and thus allow easy evaluation of all posterior features of interest. For example,

$$\widehat{p}\left(k=l|\,\mathbf{x}_{0:T-1}\right)=\frac{1}{P}\sum_{j=1}^{P}\mathbb{I}_{\{l\}}\left(k^{(j)}\right). \tag{16}$$

In practice, we take the most straightforward approach to obtain marginal densities: the samples $k^{(j)}$ from the joint posterior density $p\left(k,\mathbf{\Psi}_k,\boldsymbol{\xi}_k|\,\mathbf{x}_{0:T-1}\right)$ are collected into frequency bins, ignoring other parameters, and the histogram is plotted directly. Once the estimate of $p\left(k|\,\mathbf{x}_{0:T-1}\right)$ is obtained, the model selection is performed using the MMAP criterion, from which the number of changepoints is estimated as

$$\widehat{k}=\underset{k\in\{0,\ldots,k_{\max}\}}{\arg\max}\ \widehat{p}\left(k|\,\mathbf{x}_{0:T-1}\right). \tag{17}$$

Having fixed $k=\widehat{k}$, we proceed with the estimation of $p\left(\tau_{i,\hat{k}}\middle|\,\widehat{k},\mathbf{x}_{0:T-1}\right)$, $i=1,\ldots,\hat{k}$, and $p\left(p_{i,\hat{k}}\middle|\,\widehat{k},\mathbf{x}_{0:T-1}\right)$, $i=0,\ldots,\hat{k}$, by plotting the corresponding histograms, from which the estimates of the positions of changepoints and the model orders for each segment are obtained in exactly the same way (using MMAP):

$$\widehat{\tau}_{i,\hat{k}}=\underset{\tau_{i,\hat{k}}\in\{1,\ldots,T-1\}}{\arg\max}\ \widehat{p}\left(\tau_{i,\hat{k}}\middle|\,\widehat{k},\mathbf{x}_{0:T-1}\right),\ i=1,\ldots,\hat{k}, \tag{18}$$

$$\widehat{p}_{i,\hat{k}}=\underset{p_{i,\hat{k}}\in\{0,\ldots,p_{\max}\}}{\arg\max}\ \widehat{p}\left(p_{i,\hat{k}}\middle|\,\widehat{k},\mathbf{x}_{0:T-1}\right),\ i=0,\ldots,\hat{k}. \tag{19}$$

In fact, as shown in the next section, the parameters $\left(\{\mathbf{a}_{i,k}^{(p_{i,k})}\}_{i=0,\ldots,k},\boldsymbol{\sigma}_k^2\right)$ can be integrated out analytically due to the Gaussian noise assumption and the choice of prior distribution, and, if necessary, can then be straightforwardly estimated.

3.4 Integration of the nuisance parameters

The proposed Bayesian model allows for the integration of the nuisance parameters $\left(\{\mathbf{a}_{i,k}^{(p_{i,k})}\}_{i=0,\ldots,k},\boldsymbol{\sigma}_k^2\right)$ and subsequently gives us the expression for $p\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\xi}_k|\,\mathbf{x}_{0:T-1}\right)$ up to a normalizing constant:

$$\begin{aligned}&p\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\xi}_k|\,\mathbf{x}_{0:T-1}\right)\propto\\&\qquad\int_{\mathbb{R}^+}\int_{\mathbb{R}^{p_{0,k}}}\ldots\int_{\mathbb{R}^+}\int_{\mathbb{R}^{p_{k,k}}}p\left(k,\mathbf{\Psi}_k,\boldsymbol{\xi}_k|\,\mathbf{x}_{0:T-1}\right)d\mathbf{a}_{0,k}d\sigma_{0,k}^2\ldots d\mathbf{a}_{k,k}d\sigma_{k,k}^2.\end{aligned} \tag{20}$$

Thus, from Eq. (15):

$$
\begin{aligned}
&p\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\xi}_k\right|\mathbf{x}_{0:T-1})\propto\\
&\quad\prod_{i=0}^{k}\left[\left(2\pi\sigma_{i,k}^2\right)^{-\frac{p_{i,k}}{2}}\exp\left(-\frac{\left(\mathbf{a}_{i,k}^{(p_{i,k})}-\mathbf{m}_{i,k}^{(p_{i,k})}\right)^{\mathtt{T}}\left[\mathbf{M}_{i,k}^{(p_{i,k})}\right]^{-1}\left(\mathbf{a}_{i,k}^{(p_{i,k})}-\mathbf{m}_{i,k}^{(p_{i,k})}\right)}{2\sigma_{i,k}^2}\right)\right]\\
&\quad\times\prod_{i=0}^{k}\left[\left(\sigma_{i,k}^2\right)^{-\frac{\nu_0}{2}-\frac{\tau_{i+1,k}-\tau_{i,k}}{2}-1}\exp\left(-\frac{\left(\gamma_0+\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}^{\mathtt{T}}\mathbf{P}_{i,k}^{(p_{i,k})}\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}\right)}{2\sigma_{i,k}^2}\right)\right]\\
&\quad\times\prod_{i=0}^{k}\left[(2\pi)^{-\frac{\tau_{i+1,k}-\tau_{i,k}}{2}}\frac{(\gamma_0)^{\frac{\nu_0}{2}}}{\Gamma(\frac{v_0}{2})}\frac{\beta_\delta^{\alpha_\delta}}{\Gamma(\alpha_\delta)}\left(\delta_{i,k}^2\right)^{-\frac{p_{i,k}}{2}}\left(\delta_{i,k}^2\right)^{-\alpha_\delta-1}\exp\left(-\frac{\beta_\delta}{\delta_{i,k}^2}\right)\right]\\
&\quad\times\prod_{i=0}^{k}\left[\frac{c_{p_{\max}}\theta^{p_{i.k}}}{p_{i,k}!}\right]\lambda^k\left(1-\lambda\right)^{T-k-2}\gamma_0^{-1}\beta_\delta^{\zeta-1}\exp\left(-\varkappa\beta_\delta\right)\mathbb{I}_{\boldsymbol{\Upsilon}_k}(\boldsymbol{\tau}_k)\mathbb{I}_{\digamma_k}(\mathbf{p}_k),
\end{aligned}
\tag{21}
$$

where $\digamma_k\triangleq\{0,\ldots,p_{\max}\}^k$ and

$$
\begin{aligned}
&\mathbf{M}_{i,k}^{(p_{i,k})}=\left[\mathbf{X}_{i,k}^{(p_{i,k})\mathtt{T}}\mathbf{X}_{i,k}^{(p_{i,k})}+\frac{1}{\delta_{i,k}^2}\mathbf{I}_{p_{i,k}}\right]^{-1},\ \ \mathbf{m}_{i,k}^{(p_{i,k})}=\mathbf{M}_{i,k}^{(p_{i,k})}\mathbf{X}_{i,k}^{(p_{i,k})\mathtt{T}}\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1},\\
&\mathbf{P}_{i,k}^{(p_{i,k})}=\mathbf{I}_{\tau_{i,k}:\tau_{i+1,k}-1}-\mathbf{X}_{i,k}^{(p_{i,k})}\mathbf{M}_{i,k}^{(p_{i,k})}\mathbf{X}_{i,k}^{(p_{i,k})\mathtt{T}}.
\end{aligned}
\tag{22}
$$

The marginalised expression becomes:

$$
\begin{aligned}
&p\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\xi}_k\right|\mathbf{x}_{0:T-1})\propto\\
&\quad\prod_{i=0}^{k}\left[\Gamma(\tfrac{v_0+\tau_{i+1,k}-\tau_{i,k}}{2})\left(\gamma_0+\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}^{\mathtt{T}}\mathbf{P}_{i,k}^{(p_{i,k})}\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}\right)^{-\frac{v_0+\tau_{i+1,k}-\tau_{i,k}}{2}}\right]\\
&\quad\times\prod_{i=0}^{k}\left[\left|\mathbf{M}_{i,k}^{(p_{i,k})}\right|^{\frac{1}{2}}\pi^{-\frac{\tau_{i+1,k}-\tau_{i,k}}{2}}\frac{(\gamma_0)^{\frac{\nu_0}{2}}}{\Gamma(\frac{v_0}{2})}\frac{\beta_\delta^{\alpha_\delta}}{\Gamma(\alpha_\delta)}\left(\delta_{i,k}^2\right)^{-\alpha_\delta-\frac{p_{i,k}}{2}-1}\exp\left(-\frac{\beta_\delta}{\delta_{i,k}^2}\right)\right]\\
&\quad\times\prod_{i=0}^{k}\left[\frac{c_{p_{\max}}\theta^{p_{i.k}}}{p_{i,k}!}\right]\times\lambda^k\left(1-\lambda\right)^{T-k-2}\gamma_0^{-1}\beta_\delta^{\zeta-1}\exp\left(-\varkappa\beta_\delta\right)\mathbb{I}_{\boldsymbol{\Upsilon}_k}(\boldsymbol{\tau}_k)\mathbb{I}_{\digamma_k}(\mathbf{p}_k),
\end{aligned}
\tag{23}
$$

As it was already pointed out in Section 3.3, this posterior distribution is complex in the parameters $(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\xi}_k)$ and the posterior model probability $p\left(k,\mathbf{p}_k\right|\mathbf{x}_{0:T-1})$ cannot be determined analytically. In the next section we develop a method to estimate $p\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\xi}_k\right|\mathbf{x}_{0:T-1})$ or, if needed, $p\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\{\mathbf{a}_{i,k}^{(p_{i,k})}\}_{i=0,\ldots,k},\boldsymbol{\sigma}_k^2,\boldsymbol{\xi}_k\middle|\mathbf{x}_{0:T-1}\right)$.

4 MCMC algorithm

The problem addressed in this paper is, in fact, a model uncertainty problem of variable dimensionality in terms of both the number of changepoints and the model order for each segment. It can be treated very efficiently through the use of MCMC methods, and reversible jump MCMC [11] is particularly suitable for this case. This method extends the traditional Metropolis-Hastings algorithm to the case where moves from one dimension to another are proposed with a certain acceptance probability. This probability should be designed in a special way in order to preserve reversibility and thus ensure that $p\left(k,\boldsymbol{\Psi}_k,\boldsymbol{\xi}_k\right|\mathbf{x}_{0:T-1})$ is the invariant distribution of the Markov chain (MC). In general, if we propose a move from model $(k, \mathbf{p}_k)$ with parameters $(\boldsymbol{\tau}_k, \boldsymbol{\xi}_k)$ to model $(k', \mathbf{p}_{k'})$ with parameters $(\boldsymbol{\tau}_{k'}, \boldsymbol{\xi}_{k'})$ using a proposal distribution $q\left(k', \boldsymbol{\tau}_{k'}, \mathbf{p}_{k'}, \boldsymbol{\xi}_{k'} \middle| k, \boldsymbol{\tau}_k, \mathbf{p}_k, \boldsymbol{\xi}_k\right)$, the acceptance probability is given by:

$$\alpha = \min\left\{1, \frac{p\left(k', \boldsymbol{\tau}_{k'}, \mathbf{p}_{k'}, \boldsymbol{\xi}_{k'} \middle| \mathbf{x}_{0:T-1}\right)}{p\left(k, \boldsymbol{\tau}_k, \mathbf{p}_k, \boldsymbol{\xi}_k \middle| \mathbf{x}_{0:T-1}\right)} \frac{q\left(k, \boldsymbol{\tau}_k, \mathbf{p}_k, \boldsymbol{\xi}_k \middle| k', \boldsymbol{\tau}_{k'}, \mathbf{p}_{k'}, \boldsymbol{\xi}_{k'}\right)}{q\left(k', \boldsymbol{\tau}_{k'}, \mathbf{p}_{k'}, \boldsymbol{\xi}_{k'} \middle| k, \boldsymbol{\tau}_k, \mathbf{p}_k, \boldsymbol{\xi}_k\right)}\right\}. \tag{24}$$

Here the proposal is made directly in the new parameter space rather than via "dimensional" matching random variables (see [11]) and the Jacobian term is therefore equal to 1 ([6], [10], [20]).

In fact, the only condition to be fulfilled in selecting different types of moves is to be able to maintain the correct invariant distribution. A particular choice will only affect the convergence rate of the algorithm. To ensure a low level of rejections, we want the proposed "jumps" to be small; therefore the following moves have been selected:

- birth of a changepoint (proposing a new changepoint at random),
- death of a changepoint (removing a changepoint chosen randomly),
- update of the positions of changepoints (proposing a new position for each of the existing changepoints).

At each iteration one of the moves described above is randomly chosen with probabilities b_k, d_k and u_k such that $b_k + d_k + u_k = 1$ for all $0 \leq k \leq k_{\max}$. For $k = 0$ the death of a changepoint is impossible and for $k = k_{\max}$, the birth is impossible, thus $d_0 \triangleq 0$, $b_{k_{\max}} \triangleq 0$. Otherwise we choose $b_k = d_k = u_k$. After each move we perform the update of the number of poles for each AR model. We now describe the main steps of the algorithm:

Reversible Jump MCMC algorithm (main procedure).

1. Initialize $\left(k^{(0)}, \boldsymbol{\tau}_k^{(0)}, \mathbf{p}_k^{(0)}, \lambda^{(0)}, \theta^{(0)}, \boldsymbol{\delta}_k^{2(0)}, \gamma_0^{(0)}, \beta_\delta^{(0)}\right) \in \boldsymbol{\Theta}$. Set $j = 1$.
2. Iteration j.
   If $\left(u \sim \mathcal{U}_{(0,1)}\right) \leq b_{k^{(j)}}$ then birth of a new changepoint (see Section 4.1).
   else if $u \leq b_{k^{(j)}} + d_{k^{(j)}}$ then death of a changepoint (see Section 4.1).
   else update the changepoints positions (see Section 4.2).
3. Update of the number of poles (see Section 4.3).
4. $j \leftarrow j + 1$ and goto 2.

We now detail these different steps of the algorithm. To simplify the notation, we drop the superscript (j) from all variables at iteration j.

4.1 Death/birth of the changepoints

First, let the current state of the MC be $\left(k+1, \boldsymbol{\tau}_{k+1}, \mathbf{p}_{k+1}, \boldsymbol{\delta}^2_{k+1}, \lambda, \theta, \gamma_0, \beta_\delta\right)$ and consider the death move, which implies a modification of the dimension of the model respectively from $k+1$ to k. Our proposal begins by choosing a changepoint to be removed among $k+1$ existing ones. If the move is accepted then two segments $(l-1)^{th}$ and l^{th} will be merged, thus reducing $k+1$ by 1, and a new AR model will be created. We choose the order of the new proposed AR model to be $p_{ol} = p_{1l} + p_{2l}$, where p_{1l}, p_{2l} are the orders of the existing $(l-1)^{th}$ and l^{th} AR models[1] (see Fig. 2). The choice of proposal distribution for the hyperparameter δ^2_{ol} will be described later. The algorithm proceeds as follows:

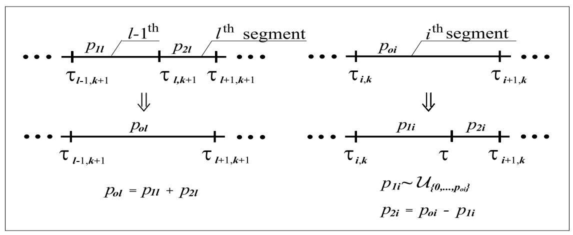

Figure 2: Death (left) and birth (right) moves.

Algorithm for the death move

- Choose a changepoint among the $k+1$ existing ones $l \sim \mathcal{U}_{\{1,\ldots,k+1\}}$.
- The proposed model order is $p_{ol} = p_{1l} + p_{2l}$, where $p_{1l} = p_{l-1,k+1}$, $p_{2l} = p_{l,k+1}$; sample $\delta^2_{ol} \big| \left(\tau_{l-1,k+1}, \tau_{l+1,k+1}, p_{ol}, \mathbf{x}_{\tau_{l-1,k+1}:\tau_{l+1,k+1}-1}\right)$, see Eq. (28)
- Evaluate α_{death}, see Eq. (30).
- If $\left(u_d \sim \mathcal{U}_{(0,1)}\right) \leq \alpha_{death}$ then the new state of the MC becomes $(k, \{\boldsymbol{\tau}_{1:l-1,k+1}, \boldsymbol{\tau}_{l+1:k+1,k+1}\}, \{\mathbf{p}_{1:l-2,k+1}, p_{ol}, \mathbf{p}_{l+1:k+1,k+1}\},$ $\{\boldsymbol{\delta}^2_{1:l-2,k+1}, \delta^2_{ol}, \boldsymbol{\delta}^2_{l+1:k+1,k+1}\}, \lambda, \theta, \gamma_0, \beta_\delta)$, otherwise it stays $\left(k+1, \boldsymbol{\tau}_{k+1}, \mathbf{p}_{k+1}, \boldsymbol{\delta}^2_{k+1}, \lambda, \theta, \gamma_0, \beta_\delta\right)$.

For the birth move ($k \to k+1$), again, first the position of a new changepoint τ is proposed. For $\tau_{i,k} < \tau < \tau_{i+1,k}$ the i^{th} segment should be split into two and the new AR model orders should be $p_{1i} \sim \mathcal{U}_{\{0,\ldots,p_{oi}\}}$ and $p_{2i} = p_{oi} - p_{1i}$, where p_{oi} is the order of the i^{th} model (see Fig. 2). This choice for the number of poles ensures that birth/death moves are reversible ($p_{oi} = p_{1i} + p_{2i}$). Thus, assuming that the current state of the MC is $\left(k, \boldsymbol{\tau}_k, \mathbf{p}_k, \boldsymbol{\delta}^2_k, \lambda, \theta, \gamma_0, \beta_\delta\right)$, we have:

[1]We keep this notation to obtain a general equation for acceptance probabilities.

Algorithm for the birth move

- Propose a new changepoint τ in $\{1,\ldots,T-2\}$: $\tau\sim\mathcal{U}_{\{1,\ldots,T-2\}\backslash\{\boldsymbol{\tau}_k\}}$.
- The proposed model orders are: $p_{1i}=\mathcal{U}_{\{1,\ldots,p_{oi}\}}$, $p_{2i}=p_{oi}-p_{1i}$, where $p_{oi}=p_{i,k}$ for $\tau_{i,k}\leq\tau<\tau_{i+1,k}$;
sample $\delta^2_{1i}\big|\left(\tau_{i,k},\tau,p_{1i},\mathbf{x}_{\tau_{i,k}:\tau-1}\right)$, $\delta^2_{2i}\big|\left(\tau,\tau_{i+1,k},p_{2i},\mathbf{x}_{\tau:\tau_{i+1,k}-1}\right)$ see Eq. (28);
- Evaluate α_{birth}, see Eq. (30).
- If $\left(u_b\sim\mathcal{U}_{(0,1)}\right)\leq\alpha_{birth}$ then the new state of the MC becomes
$(k+1,\{\boldsymbol{\tau}_{1:i,k},\tau,\boldsymbol{\tau}_{i+1:k,k}\},\{\mathbf{p}_{1:i-1,k},p_{1i},p_{2i},\mathbf{p}_{i+1:k,k}\},$
$\{\boldsymbol{\delta}^2_{1:i-1,k},\delta^2_{1i},\delta^2_{2i},\boldsymbol{\delta}^2_{i+1:k,k}\},\lambda,\theta,\gamma_0,\beta_\delta),$
otherwise it stays $\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\delta}^2_k,\lambda,\theta,\gamma_0,\beta_\delta\right)$.

■

To perform these moves in practice, it is necessary to choose a proposal distribution for the elements of $\boldsymbol{\delta}^2_k$ in such a way that we avoid rejecting too many candidates. If the number of changepoints were fixed and we wanted just to update the values of $\boldsymbol{\delta}^2_k$, we would sample the elements of the vector according to a standard Gibbs move (see [19]), i.e. from Eq. (21)

$$\mathcal{IG}\left(\delta^2_{i,k};\,\alpha_\delta+\frac{p_{i,k}}{2},\,\beta_\delta+\frac{\mathbf{a}^{(p_{i,k})\mathrm{T}}_{i,k}\mathbf{a}^{(p_{i,k})}_{i,k}}{2\sigma^2_{i,k}}\right),\tag{25}$$

where $\mathbf{a}^{(p_{i,k})}_{i,k},\sigma^2_{i,k}$ are sampled from the following distributions (see Eq. (21)):

$$\mathcal{IG}\left(\sigma^2_{i,k};\,\frac{\nu_0+\tau_{i+1,k}-\tau_{i,k}}{2},\,\frac{\gamma_0+\mathbf{x}^{\mathrm{T}}_{\tau_{i,k}:\tau_{i+1,k}-1}\mathbf{P}^{(p_{i,k})}_{i,k}\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}}{2}\right),\tag{26}$$

$$\mathcal{N}\left(\mathbf{a}^{(p_{i,k})}_{i,k};\,\mathbf{m}^{(p_{i,k})}_{i,k},\,\sigma^2_{i,k}\mathbf{M}^{(p_{i,k})}_{i,k}\right),\tag{27}$$

with matrices $\mathbf{M}^{(p_{i,k})}_{i,k},\mathbf{P}^{(p_{i,k})}_{i,k}$ and vector $\mathbf{m}^{(p_{i,k})}_{i,k}$ depending on the value of $\delta^2_{i,k}$ before updating. In the case of a birth/death move, we do not have any previous value but instead we can use the mean of the distribution $\mathcal{IG}\left(\alpha_\delta+\frac{p_{i,k}}{2},\ \beta_\delta\right)$: $\delta^{2*}_{i,k}=\frac{\beta_\delta}{\alpha_\delta+\frac{p_{i,k}}{2}-1}$ and sample $\delta^2_{i,k}$ using Metropolis-Hastings steps (see [19]); the corresponding matrices are denoted as $\mathbf{M}^{(p_{i,k})*}_{i,k},\mathbf{P}^{(p_{i,k})*}_{i,k}$ and $\mathbf{m}^{(p_{i,k})*}_{i,k}$. Taking this into account, we construct our proposal distribution in the following way:

$$\mathcal{IG}(\delta^2_{i,k};\,\tilde{\alpha}_\delta,\,\tilde{\beta}_\delta)=\mathcal{IG}\left(\delta^2_{i,k};\,\alpha_\delta+\frac{p_{i,k}}{2},\,\beta_\delta+\frac{\overline{\mathbf{a}}^{(p_{i,k})\mathrm{T}}_{i,k}\overline{\mathbf{a}}^{(p_{i,k})}_{i,k}}{2\overline{\sigma^2_{i,k}}}\right),\tag{28}$$

where $\overline{\mathbf{a}}_{i,k}^{(p_{i,k})}$, $\overline{\sigma_{i,k}^2}$ are the means of the distributions corresponding to Eq. (26), (27) but with $\mathbf{M}_{i,k}^{(p_{i,k})*}$, $\mathbf{P}_{i,k}^{(p_{i,k})*}$ and $\mathbf{m}_{i,k}^{(p_{i,k})*}$:

$$\overline{\mathbf{a}}_{i,k}^{(p_{i,k})} = \mathbf{m}_{i,k}^{(p_{i,k})*}, \ \overline{\sigma_{i,k}^2} = \frac{\gamma_0 + \mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}^{\mathrm{T}} \mathbf{P}_{i,k}^{(p_{i,k})*} \mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}}{\nu_0 + \tau_{i+1,k} - \tau_{i,k} - 2}. \tag{29}$$

The acceptance ratio of the birth and death (of changepoint) moves is evaluated according to the general expression (24). We obtain the acceptance probabilities:

$$\alpha_{birth} = \min\{1, r_{birth}\} \text{ and } \alpha_{death} = \min\{1, r_{birth}^{-1}\} \tag{30}$$

where

$$\begin{aligned} r_{birth} = & \frac{p\left(k+1, \boldsymbol{\tau}_{k+1}, \mathbf{p}_{k+1}, \lambda, \theta, \boldsymbol{\delta}_{k+1}^2, \gamma_0, \beta_\delta \middle| \mathbf{x}_{0:T-1}\right)}{p\left(k, \boldsymbol{\tau}_k, \mathbf{p}_k, \lambda, \theta, \boldsymbol{\delta}_k^2, \gamma_0, \beta_\delta \middle| \mathbf{x}_{0:T-1}\right)} \times \frac{q(k, \boldsymbol{\tau}_k, \mathbf{p}_k | k+1, \boldsymbol{\tau}_{k+1}, \mathbf{p}_{k+1})}{q(k+1, \boldsymbol{\tau}_{k+1}, \mathbf{p}_{k+1} | k, \boldsymbol{\tau}_k, \mathbf{p}_k)} \\ & \times \frac{q\left(\boldsymbol{\delta}_k^2 \middle| k+1, \boldsymbol{\tau}_{k+1}, \mathbf{p}_{k+1}, \boldsymbol{\delta}_{k+1}^2, \gamma_0, \beta_\delta, \mathbf{x}_{0:T-1}\right)}{q\left(\boldsymbol{\delta}_{k+1}^2 \middle| k, \boldsymbol{\tau}_k, \mathbf{p}_k, \boldsymbol{\delta}_k^2, \gamma_0, \beta_\delta, \mathbf{x}_{0:T-1}\right)} \end{aligned} \tag{31}$$

and

$$\begin{aligned} \frac{q(k, \boldsymbol{\tau}_k, \mathbf{p}_k | k+1, \boldsymbol{\tau}_{k+1}, \mathbf{p}_{k+1})}{q(k+1, \boldsymbol{\tau}_{k+1}, \mathbf{p}_{k+1} | k, \boldsymbol{\tau}_k, \mathbf{p}_k)} &= \frac{q(k|k+1)}{q(k+1|k)} \times \frac{q(\boldsymbol{\tau}_k | k+1, \boldsymbol{\tau}_{k+1})}{q(\boldsymbol{\tau}_{k+1} | k, \boldsymbol{\tau}_k)} \times \frac{q(\mathbf{p}_k | k+1, \mathbf{p}_{k+1})}{q(\mathbf{p}_{k+1} | k, \mathbf{p}_k)} \\ &= \frac{d_{k+1}}{b_k} \times \frac{(T-2-k)}{(k+1)} \times \frac{(p_{oi}+1)}{1}. \end{aligned} \tag{32}$$

Finally, from Eq. (23) for the birth of the changepoint τ, $(\tau_{i,k} \leq \tau < \tau_{i+1,k})$ we obtain:

$$r_{birth}^i = \frac{\lambda}{(1-\lambda)} \frac{f(\tau_{i,k}, \tau, p_{1i}, \delta_{1i}^2) f(\tau, \tau_{i+1,k}, p_{2i}, \delta_{2i}^2)}{f(\tau_{i+1,k}, \tau_{i,k}, p_{i,k}, \delta_{i,k}^2)} \frac{d_{k+1}}{b_k} \frac{(T-2-k)(p_{oi}+1)}{(k+1)}, \tag{33}$$

where, for convenience, we denote for the segment between the changepoints $\tau_{i,k}, \tau_{i+1,k}$:

$$\begin{aligned} f(\tau_{i,k}, \tau_{i+1,k}, p_{i,k}, \delta_{i,k}^2) = & \frac{(\gamma_0)^{\frac{\nu_0}{2}}}{\Gamma(\frac{\nu_0}{2})} \frac{c_{p_{\max}} \theta^{p_{i.k}}}{p_{i,k}!} \frac{\beta_\delta^{\alpha_\delta}}{\Gamma(\alpha_\delta)} \left[\frac{\tilde{\beta}_\delta^{\tilde{\alpha}_\delta}}{\Gamma(\tilde{\alpha}_\delta)} \right]^{-1} \exp(-\frac{\beta_\delta - \tilde{\beta}_\delta}{\delta_{i,k}^2}) \\ & \times \Gamma\left(\frac{v_0 + \tau_{i+1,k} - \tau_{i,k}}{2}\right) \left| \mathbf{M}_{i,k}^{(p_{i,k})} \right|^{\frac{1}{2}} \left(\gamma_0 + \mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}^{\mathrm{T}} \mathbf{P}_{i,k}^{(p_{i,k})} \mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}\right)^{-\frac{v_0 + \tau_{i+1,k} - \tau_{i,k}}{2}}. \end{aligned} \tag{34}$$

4.2 Update of the changepoint positions

Although the update of the changepoint positions does not involve the change in dimension k, it is somewhat more complicated than the birth/death moves. In fact, updating the position of changepoint $\tau_{l,k}$ means removing the l^{th} changepoint and proposing instead a new one τ. We determine i such that $\tau_{i,k} < \tau < \tau_{i+1,k}$ and it is worth noticing that if $i \neq l$ the update move may actually be described as a combination of the birth of the changepoint τ and the death of the changepoint $\tau_{l,k}$ (see Fig. 3). Otherwise, we leave the model orders the same and just sample the values of the hyperparameters δ_{1l}^2, δ_{2l}^2. This process is repeated for all existing changepoints, $l = 1, \ldots, k$, and is described below in more detail.

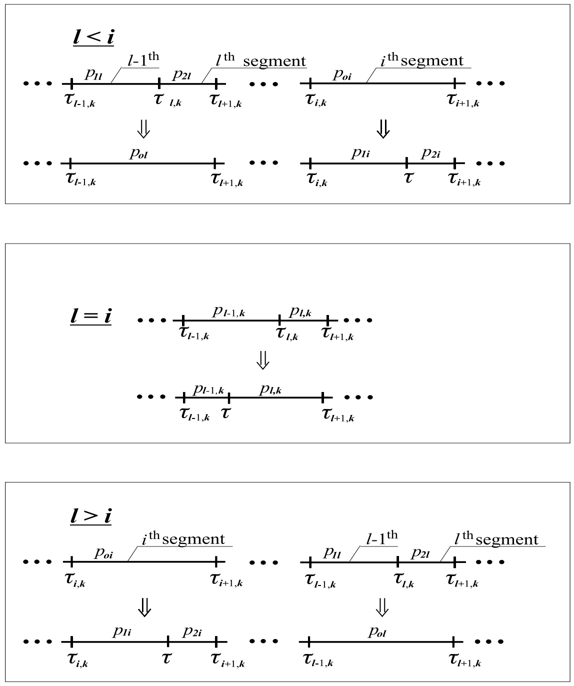

Figure 3: Update of the changepoint positions

Algorithm for the update of the changepoint positions

For $l = 1, \ldots, k$

- Propose a new position for the l^{th} changepoint $\tau \sim \mathcal{U}_{\{1,\ldots,T-2\}\setminus\{\boldsymbol{\tau}_k\}}$ and determine i such that $\tau_{i,k} < \tau < \tau_{i+1,k}$.
- If $l \neq i$ then
  $p_{1i} = \mathcal{U}_{\{1,\ldots,p_{oi}\}}$, $p_{2i} = p_{oi} - p_{1i}$, where $p_{oi} = p_{i,k}$,
  $p_{ol} = p_{1l} + p_{2l}$, where $p_{1l} = p_{l-1,k}$, $p_{2l} = p_{l,k}$;
  sample $\delta^2_{1i} \big| \left(\tau_{i,k}, \tau, p_{1i}, \mathbf{x}_{\tau_{i,k}:\tau-1}\right)$, $\delta^2_{2i} \big| \left(\tau, \tau_{i+1,k}, p_{2i}, \mathbf{x}_{\tau:\tau_{i+1,k}-1}\right)$
  and $\delta^2_{ol} \big| \left(\tau_{l-1,k+1}, \tau_{l+1,k+1}, p_{o,l}, \mathbf{x}_{\tau_{l-1,k+1}:\tau_{l+1,k+1}-1}\right)$, see Eq. (28);
  else sample $\delta^2_{1l} \big| \left(\tau_{l-1,k}, \tau, p_{l-1,k}, \mathbf{x}_{\tau_{l-1,k}:\tau-1}\right)$, $\delta^2_{2l} \big| \left(\tau, \tau_{l+1,k}, p_{l,k}, \mathbf{x}_{\tau:\tau_{l+1,k}-1}\right)$, see Eq. (28);
- Evaluate α_{update}, if $l \neq i$ then see Eq. (35) else see Eq. (36)
- If $\left(u_u \sim \mathcal{U}_{(0,1)}\right) \leq \alpha_{update}$ then the new state of the MC becomes
  - if $l < i$ then
    $(k, \{\boldsymbol{\tau}_{1:l-1,k}, \boldsymbol{\tau}_{l+1:i,k}, \boldsymbol{\tau}, \boldsymbol{\tau}_{i+1:k,k}\}, \{\mathbf{p}_{1:l-2,k}, p_{ol}, \mathbf{p}_{l+1:i-1,k}, p_{1i}, p_{2i}, \mathbf{p}_{i+1:k,k}\},$
    $\{\boldsymbol{\delta}^2_{1:l-2,k}, \delta^2_{ol}, \boldsymbol{\delta}^2_{l+1:i-1,k}\delta^2_{1i}, \delta^2_{2i}, \boldsymbol{\delta}^2_{i+1:k,k}\}, \lambda, \theta, \gamma_0, \beta_\delta);$

– else if $l > i$ then
$(k, \{\boldsymbol{\tau}_{1:i,k}, \boldsymbol{\tau}, \boldsymbol{\tau}_{i+1:l-1,k}, \boldsymbol{\tau}_{l+1:k,k}\}, \{\mathbf{p}_{1:i-1,k}, p_{1i}, p_{2i}, \mathbf{p}_{i+1:l-2,k}, p_{ol}, \mathbf{p}_{l+1:k,k}\},$
$\{\boldsymbol{\delta}^2_{1:i-1,k}, \delta^2_{1i}, \delta^2_{2i}, \boldsymbol{\delta}^2_{i+1:l-2,k}\delta^2_{ol}, \boldsymbol{\delta}^2_{l+1:k,k}\}, \lambda, \theta, \gamma_0, \beta_\delta)$

– else $(k, \{\boldsymbol{\tau}_{1:l-1,k}, \tau, \boldsymbol{\tau}_{l+1:k,k}\}, \mathbf{p}_k, \{\boldsymbol{\delta}^2_{1:l-2,k}, \delta^2_{1l}, \delta^2_{2l}, \boldsymbol{\delta}^2_{l+1:k,k}\}, \lambda, \theta, \gamma_0, \beta_\delta)$.

otherwise it stays $\left(k, \boldsymbol{\tau}_k, \mathbf{p}_k, \boldsymbol{\delta}^2_k, \lambda, \theta, \gamma_0, \beta_\delta\right)$.

Since for $l \neq i$ the update of the positions of changepoints combines the birth of the i^{th} changepoint and death of the l^{th} changepoint at the same time, the acceptance ratio for the proposed move is given by:

$$\alpha_{update} = \min\left\{1, r^i_{birth} r^l_{death}\right\}. \tag{35}$$

If $l = i$, it becomes:

$$r_{update} = \frac{f(\tau_{l-1,k}, \tau, p_{l-1,k}, \delta^2_{1l}) f(\tau, \tau_{l+1,k}, p_{l,k}, \delta^2_{2l})}{f(\tau_{l-1,k}, \tau_{l,k}, p_{l-1,k}, \delta^2_{l-1,k}) f(\tau_{l,k}, \tau_{l+1,k}, p_{l,k}, \delta^2_{l,k})}. \tag{36}$$

where $f(\cdot)$ is defined in (34).

4.3 Update of the number of poles

The update of the number of poles for each segment does not involve changing the number of changepoints and their positions. However, we still have to perform "jumps" between the subspaces of different dimensions $p_{i,k}$ and will therefore continue using the reversible jump MCMC method, though it is formulated now in a less complicated form. Similarly, the moves are chosen to be: (1) birth of the pole ($p_{i,k} \to p_{i,k} + 1$), (2) death of the pole ($p_{i,k} \to p_{i,k} - 1$) and (3) just the update of the hyperparameter $\delta^2_{i,k}$. The probabilities for choosing these moves are defined in exactly the same way: $b_{p_{i,k}} + d_{p_{i,k}} + u_{p_{i,k}} = 1$; $d_0 \triangleq 0$, $b_{p_{\max}} \triangleq 0$, otherwise $b_{p_{i,k}} = d_{p_{i,k}} = u_{p_{i,k}}$ for $i = 0, ..., k$. The procedure is performed for each segment and the main steps are described as follows.

Algorithm for the update of the number of poles.

1. For $i = 1, \ldots, k$

   (a) If $\left(u_p \sim \mathcal{U}_{(0,1)}\right) \leq b_{p_{i,k}}$ then propose $p'_{i,k} = p_{i,k} + 1$;
   else if $u_p \leq b_{p_{i,k}} + d_{p_{i,k}}$ then propose $p'_{i,k} = p_{i,k} - 1$;
   else goto (d).

   (b) If $\left(u_{p_d} \sim \mathcal{U}_{(0,1)}\right) \leq \alpha_{(p_{i,k} \to p'_{i,k})}$ (see Eq. (37)) then the new state of the MC becomes
   $\left(k, \boldsymbol{\tau}_k, \{\mathbf{p}_{1:i-1,k}, p'_{i,k}, \mathbf{p}_{i+1:k,k}\}, \boldsymbol{\delta}^2_k, \lambda, \theta, \gamma_0, \beta_\delta\right)$
   otherwise it stays $\left(k, \boldsymbol{\tau}_k, \mathbf{p}_k, \boldsymbol{\delta}^2_k, \lambda, \theta, \gamma_0, \beta_\delta\right)$

(c) Sample $\sigma^2_{i,k}\Big|\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\delta}^2_k,\mathbf{x}_{0:T-1}\right)$ see Eq. (26);
sample $\mathbf{a}_{i,k}|\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\boldsymbol{\delta}^2_k,\mathbf{x}_{0:T-1},\sigma^2_{i,k}\right)$ see Eq. (27);
sample $\delta^2_{i,k}|\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\mathbf{x}_{0:T-1},\mathbf{a}_{i,k},\sigma^2_{i,k}\right)$ see Eq. (41).

2. Propose $\theta'\big|\left(k,\mathbf{p}_k\right)$ (see Eq. (39))
if $\left(u_\theta\sim\mathcal{U}_{(0,1)}\right)\leq\alpha_\theta$ (see Eq. (40)) then $\theta=\theta'$.
3. Sample $\lambda|\left(k,\boldsymbol{\tau}_k,\mathbf{x}_{0:T-1}\right)$ see Eq. (42).
4. Sample $\gamma_0|\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\mathbf{x}_{0:T-1},\boldsymbol{\sigma}^2_k\right)$ see Eq. (42).
5. Sample $\beta_\delta|\left(k,\boldsymbol{\tau}_k,\mathbf{p}_k,\mathbf{x}_{0:T-1},\{\mathbf{a}^{p_{i,k}}_{i,k}\}_{i=0,\ldots,k},\boldsymbol{\sigma}^2_k,\boldsymbol{\delta}^2_k\right)$ see Eq. (42).

■

The acceptance probability for the different types of moves (in terms of the number of poles) is given by:

$$\alpha_{(p_{i,k}\to p'_{i,k})}=\min\left\{1,r_{(p_{i,k}\to p'_{i,k})}\right\},\tag{37}$$

where from Eq. (24)

$$r_{(p_{i,k}\to p'_{i,k})}=\frac{\left|\mathbf{M}^{(p'_{i,k})}_{i,k}\right|^{\frac{1}{2}}\frac{\theta^{p'_{i,k}}}{p'_{i,k}!}\delta_{i,k}^{-\frac{p'_{i,k}}{2}}\left(\gamma_0+\mathbf{x}^{\mathrm{T}}_{\tau_{i,k}:\tau_{i+1,k}-1}\mathbf{P}^{(p'_{i,k})}_{i,k}\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}\right)^{-\frac{v_0+\tau_{i+1,k}-\tau_{i,k}}{2}}}{\left|\mathbf{M}^{(p_{i,k})}_{i,k}\right|^{\frac{1}{2}}\frac{\theta^{p_{i,k}}}{p_{i,k}!}\delta_{i,k}^{-\frac{p_{i,k}}{2}}\left(\gamma_0+\mathbf{x}^{\mathrm{T}}_{\tau_{i,k}:\tau_{i+1,k}-1}\mathbf{P}^{(p_{i,k})}_{i,k}\mathbf{x}_{\tau_{i,k}:\tau_{i+1,k}-1}\right)^{-\frac{v_0+\tau_{i+1,k}-\tau_{i,k}}{2}}}\tag{38}$$

Thus, for the birth move ($p_{i,k}\to p_{i,k}+1$) the acceptance ratio is $\alpha^p_{birth}=\min\{1,r_{birth}\}$, where $r_{birth}=r_{(p_{i,k}\to p_{i,k}+1)}$. Assuming that the current number of poles is $(p_{i,k}+1)$, one obtains the acceptance ratio for the death move ($p_{i,k}+1\to p_{i,k}$) as $\alpha^p_{death}=\min\left\{1,r^{-1}_{birth}\right\}$. Thus, the birth/death moves are, indeed, reversible.

Taking into account that the series representation of the exponential function is $\exp(\theta)=\sum_{p=0}^{\infty}\frac{\theta^p}{p!}$, we adopt the following proposal distribution for the parameter θ :

$$\mathcal{G}\left(\theta;\,\zeta+\textstyle\sum_{i=0}^{k}p_{i,k},\,\varkappa+(k+1)\right)\tag{39}$$

and sample θ according to a Metropolis-Hastings step with the acceptance probability equal to:

$$\alpha_\theta=\left[\frac{\sum_{p=0}^{p_{\max}}\theta^p}{\sum_{p=0}^{p_{\max}}(\theta')^p}\frac{\exp(-\theta)}{\exp(-\theta')}\right]^{(k+1)}.\tag{40}$$

The hyperparameters $\delta^2_{i,k}$ are sampled using a standard Gibbs move in exactly the same way as described in Section 4.1:

$$\mathcal{IG}\left(\delta^2_{i,k};\,\alpha_\delta+\frac{p_{i,k}}{2},\,\beta_\delta+\frac{\mathbf{a}^{(p_{i,k})\mathrm{T}}_{i,k}\mathbf{a}^{(p_{i,k})}_{i,k}}{2\sigma^2_{i,k}}\right)\tag{41}$$

Similarly, we sample $\beta_\delta, \lambda, \gamma_0$ according to:

$$\begin{array}{c} \mathcal{G}a\left(\beta_\delta;\, \alpha_\delta(k+1),\ \sum_{i=1}^{k} \frac{1}{\delta_{i,k}^2}\right), \\ \mathcal{B}e(\lambda;\, k+1,\, T-k-1), \\ \mathcal{G}a\left(\gamma_0;\, \frac{\nu_0(k+1)}{2},\, \frac{1}{2}\sum_{i=1}^{k} \frac{1}{\sigma_{i,k}^2}\right). \end{array} \tag{42}$$

5 Simulations

We assess the performance of the segmentation method proposed above by applying it to the synthetic data with $T = 500$ and $k = 5$. The parameters of the AR models $\{\mathbf{a}_{i,5}^{(p_{i,5})}\}_{i=0,\ldots,5}$ and noise variances $\boldsymbol{\sigma}_5^2$, drawn at random, are given in Table 1.

i^{th} *segment*	$\sigma_{i,5}$	$\mathbf{a}_{i,5}^{(p_{i,5})}$			
0	1.6	−2.3000	−2.6675	−1.8437	−0.5936
1	0.8	1.3000	−0.9200	0.2600	
2	1.7	0.8000	−0.5200		
3	0.5	2.0000	−1.6350	0.5075	
4	0.6	−1.7000	−0.7450		
5	1.8	−0.5000	0.6100	0.5850	

Table 1: The parameters of the AR model and noise variance for each segment.

The number of iterations of the algorithm was 10000, which seemed to be sufficient since the histograms of the posterior distribution were stabilized. As was described in Section 3, we adopt the MMAP of $\hat{p}(k|\,\mathbf{x}_{0:T-1})$ as a detection criterion and, indeed, find $\hat{k} = 5$ changepoints. Then, for fixed $k = \hat{k}$, the model order for each segment $p_{i,\hat{k}}$ and the positions of changepoints $\tau_{i,\hat{k}}$, $i = 1, \ldots, \hat{k}$ are estimated by MMAP. The results are presented in Table 2. In Fig. 5 and 4 the segmented signal and the estimation of the marginal posterior distributions of the number of changepoints $\hat{p}(k|\,\mathbf{x}_{0:T-1})$ and their positions $\hat{p}\left(\tau_{i,\hat{k}}\middle|\,\hat{k}, \mathbf{x}_{0:T-1}\right)$ are given. Fig. 6 shows the estimates of the marginal posterior distribution of the model order for each signal $\hat{p}\left(p_{i,\hat{k}}\middle|\,\hat{k}, \mathbf{x}_{0:T-1}\right)$.

i^{th}segment	0	1	2	3	4	5
$\tau_{i,5}$ (true value)	-	90	160	250	365	430
$\widehat{\tau_{i,\hat{k}}} = \max \hat{p}\left(\tau_{i,\hat{k}}\middle|\,\hat{k}, \mathbf{x}_{0:T-1}\right)$	-	91	162	249	366	434
$p_{i,5}$ (true value)	4	3	2	3	2	3
$\widehat{p_{i,\hat{k}}} = \max \hat{p}\left(p_{i,\hat{k}}\middle|\,\hat{k}, \mathbf{x}_{0:T-1}\right)$	4	3	2	3	2	3

Table 2: Real and estimated values for changepoint positions and model order for each segment.

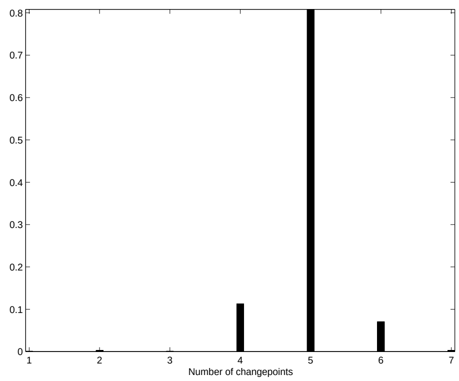

Figure 4: Estimation of the marginal posterior distribution of the number of changepoints $\hat{p}\left(k \mid \mathbf{x}_{0:T-1}\right)$.

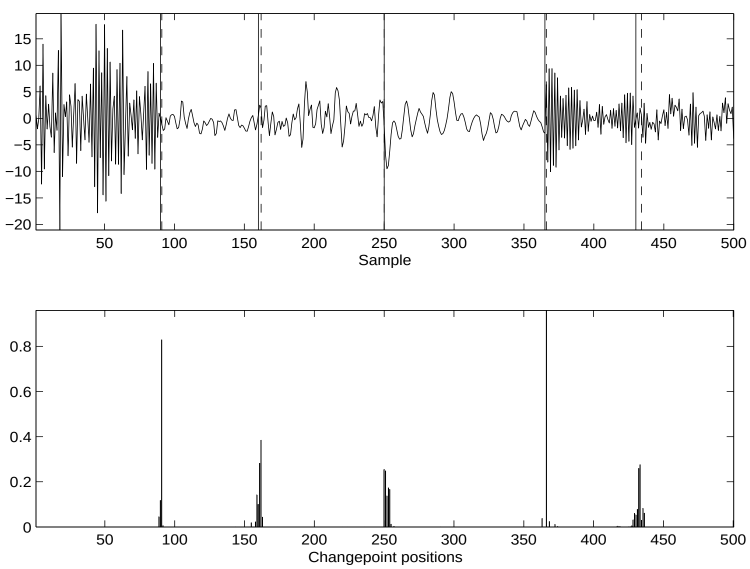

Figure 5: Top: segmented signal (the original changepoints are shown as a solid line, and the estimated changepoints are shown as a dotted line). Bottom: estimation of the marginal posterior distribution of the changepoint positions $\hat{p}\left(\tau_{i,\widehat{k}} \middle| \widehat{k}, \mathbf{x}_{0:T-1}\right)$, $i = 1, \ldots, \hat{k}$.

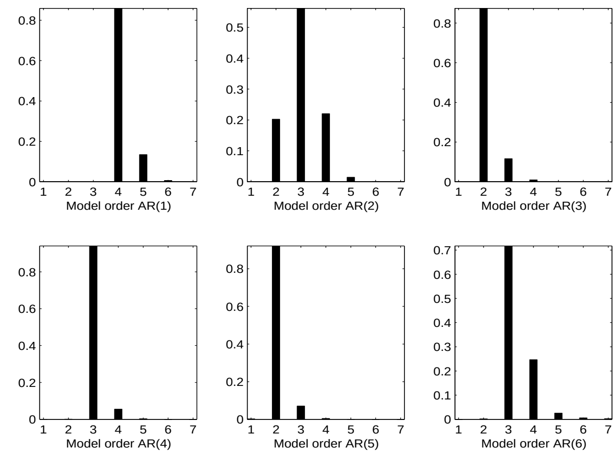

Figure 6: Estimates of the marginal posterior distributions of the number of poles for each segment $\hat{p}\left(p_{i,\widehat{k}} \middle| \widehat{k}, \mathbf{x}_{0:T-1}\right)$, $i = 0, \ldots, \hat{k}$.

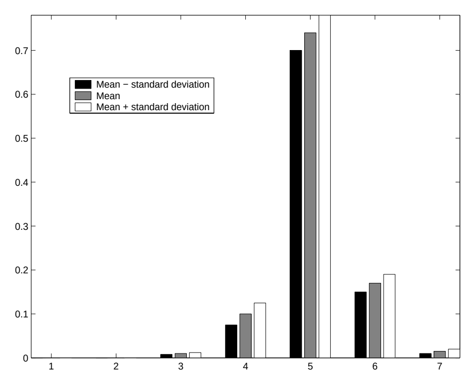

Figure 7: Mean and standard deviation for 50 realizations of the posterior distribution $p\left(k \middle| \mathbf{x}_{0:T-1}^{(i)}\right)$.

Then we estimated the mean and the associated standard deviation of the marginal posterior distributions $\left(p\left(k|\,\mathbf{x}_{0:T-1}^{(i)}\right)\right)_{i=1,\ldots,50}$ for 50 realisations of the experiment with fixed model parameters and changepoint positions. The results are presented in Fig. 7 and it is worth noticing that they are very stable regarding the fluctuations of the realization of the excitation noise.

6 Speech Segmentation

In this section we implemented the proposed algorithm for processing a real speech signal which was examined in the literature before (see [1], [4] and [14]). It was recorded inside a car by the French National Agency for Telecommunications for testing and evaluating speech recognition algorithms as described in [4]. According to [14], the sampling frequency was 12 kHz, and a high-pass filtered version of the signal with cut-off frequency 150Hz and the resolution equal to 16 bits is presented in Fig. 8.

Different segmentation methods (see [1], [3], [4], and [14]) were applied to the signal and the summary of the results can be found in [14]. We show these results in Table 3 in order to compare them to the ones obtained using our proposed method (see also Fig. 8 and 9). The estimated orders of the AR models are presented in Table 4 and as one can see they are quite different from segment to segment. This resulted in the different positions of the changepoints, which is especially crucial in the case of the third changepoint. Its position changed significantly due to the estimated model orders for the second ($\hat{p}_{2,5} = 19$) and third segments ($\hat{p}_{3,5} = 27$). As it is illustrated in Fig. 9, the changepoints obtained by the proposed method visually seem to be more accurate.

Method	AR order	Estimated changepoints							
Divergence	16	445	645	1550	1800	2151	2797	-	3626
Brand's GLR	16	445	645	1550	1800	2151	2797	-	3626
Brand's GLR	2	445	645	1550	1750	2151	2797	3400	3626
Approx. ML ([14])	2	445	626	1609	-	2151	2797	-	3627
Proposed method	estimated	448	624	1377	-	2075	2807	-	3626

Table 3: Changepoint positions for different methods.

Segment	0	1	2	3	4	5	6
Model order	6	5	19	27	16	9	11

Table 4: Estimated model orders.

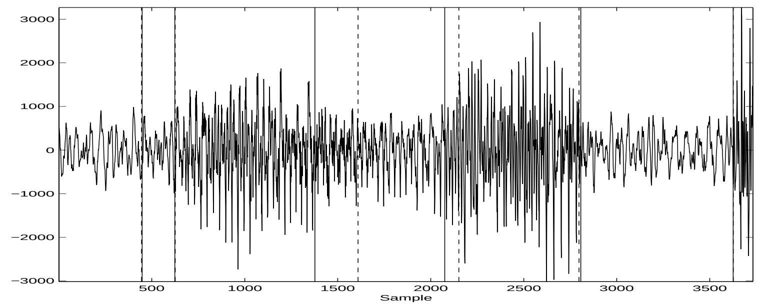

Figure 8: Segmented speech signal (the changepoints estimated by Gustafsson are shown as a dotted line and ones estimated using our proposed method are shown as a solid line).

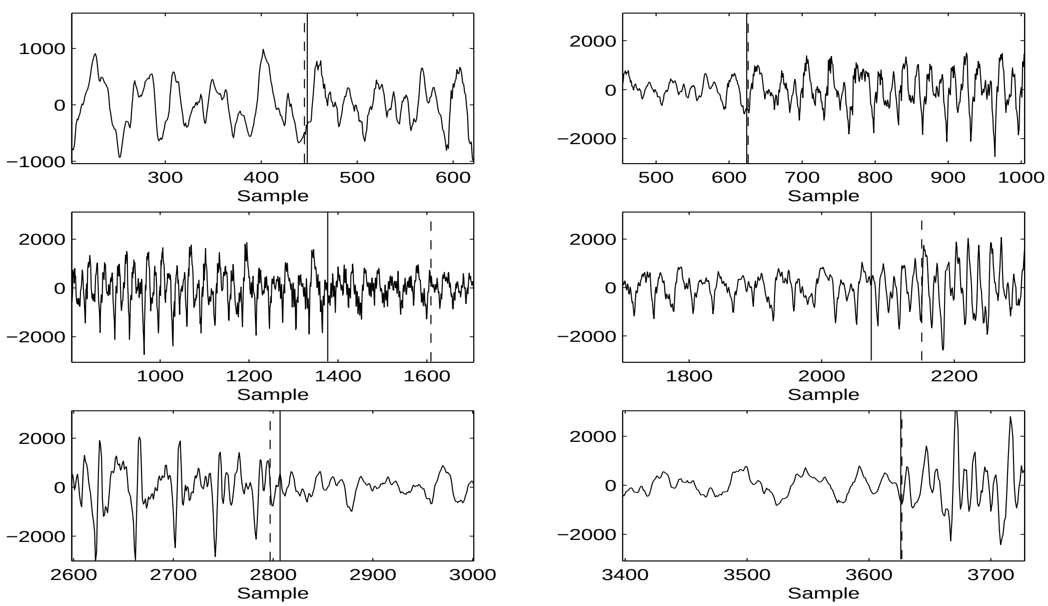

Figure 9: The changepoint positions (the changepoints estimated by Gustafsson are shown as a dotted line and the ones estimated using our proposed method are shown as a solid line).

7 Conclusion

In this paper we have addressed the problem of optimal segmentation of signals modelled as piecewise constant autoregressive (AR) processes excited by white Gaussian noise. An original Bayesian model is proposed, where the number of segments as well as the model orders, parameters and noise variances for each of them are regarded as unknown parameters. Then, an efficient reversible jump MCMC algorithm is developed to overcome the intractability of analytical Bayesian inference. This algorithm also allows us to perform estimation of the hyperparameters whereas they are usually tuned heuristically by the user in other methods [14], [15]. The results for both synthetic and real data demonstrate the efficiency of this method and confirm the good performance of both the model and the algorithm in practice.

In conclusion, we would like to point out that there are several possible extensions to this work. Indeed, the framework we propose is suitable for the segmentation of any data which might be described in terms of a linear combination of basis functions with an additive Gaussian noise component (General Piecewise Linear Model, [7], [16]). It could be also adapted for the case of non-Gaussian noise, or the more challenging task of finding the changepoints from one model type into another. Finally, it is especially important to develop an on-line method to solve the problem of changepoint detection as it is of great interest in many applications. This is the subject of our current research.

Acknowledgment

The authors are very grateful to Prof. Fredrik Gustafsson for providing the speech signal used for segmentation and would like to thank Michael Hazas and Richard Cook for careful proof-reading and useful comments.

References

[1] R. Andre-Obrecht, "A new statistical approach for automatic segmentation of continuous speech signals," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, pp. 29-40, 1988.

[2] D. Barry and J. Hartigan, "A Bayesian analysis for changepoint problems," *J. Am. Stat. Assoc.*, vol. 88, pp. 309-319, 1993.

[3] M. Basseville and A. Benveniste, "Design and comparative study of some sequential jump detection algorithms for digital signals," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 31, pp. 521-535, 1983.

[4] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Application,* Prentice Hall, Information and system science series, 1993

[5] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, Wiley, 1994.

[6] P. Dellaportas, J.J. Forster and I. Ntzoufras, "On Bayesian model and variable selection using MCMC," paper based upon a talk presented at the HSSS Workshop on Variable Dimension MCMC, New Forest, September 1997.

[7] D. Denison, B. Mallick and A.F.M. Smith, "Automatic Bayesian curve fitting," *J. Roy. Stat. Soc. B*, vol. 60, pp. 333-350, 1998.

[8] P. Djurić, "Segmentation of nonstationary signals," in *Proc. Conf. IEEE ICASSP'92*, pp. 161–164, 1992.

[9] P. Djurić, "A MAP solution to off-line segmentation of signals," in *Proc. Conf. IEEE ICASSP'94*, pp. 505-508, 1994.

[10] S.J. Godsill, "On the relationship between MCMC model uncertainty methods," technical report, Univ. Cambridge, Dept. Engineering, CUED-F-INFENG/TR.305, 1997.

[11] P.J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711-732, 1995.

[12] F. Gustafsson, "Optimal segmentation in a linear regression framework," in *1991 Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 1677-1680, Toronto, Canada,1991.

[13] F. Gustafsson, "The marginalized likelihood ratio test for detecting abrupt changes," *IEEE Trans. Automatic control*, vol. 41, pp. 66-77, 1996.

[14] F. Gustafsson, "Segmentation of signals using piecewise constant linear regression models," to appear *IEEE Trans. Signal Processing*, 1999.

[15] M. Lavielle, "Optimal segmentation of random processes," *IEEE Trans. Signal Processing*, vol. 46, pp. 1365-1373, 1998.

[16] J.J.K. O'Ruanaidh and W.J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing,* Springer-Verlag, 1996.

[17] S. Richardson and P.J. Green, "On Bayesian analysis of mixtures with unknown number of components," *J. Roy. Stat. Soc. B*, vol. 59, no. 4, pp. 731-792, 1997.

[18] C.P. Robert, *The Bayesian Choice. A Decision- Theoretic Motivation*, Springer Texts in Statistics, Springer-Verlag, 1994.

[19] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag Series in Statistics, 1999.

[20] J.A. Stark, W.J. Fitzgerald, S.B. Hladky, "Multiple-order Markov chain Monte Carlo sampling methods with application to a changepoint model," technical report, Univ. Cambridge, Dept. Engineering, CUED-F-INFENG/TR.302, 1997.

[21] D.A. Stephens, "Bayesian retrospective multiple-changepoint identification," *Applied Statistics*, vol. 43, pp. 159-178, 1994.

[22] L. Tierney, "Markov chains for exploring posterior distributions (with discussion)," *Ann. Stat.*, pp. 1701-1762, 1994.

[23] P.T. Troughton and S.J. Godsill, "A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves," technical report, Univ. Cambridge, Dept. Engineering, CUED-F-INFENG/TR.304, 1997.

[24] A.S. Willsky, H.L. Jones, "A generalised likelihood ratio approach to the detection and estimation of jumps in linear systems," *IEEE Trans. Automatic Control*, pp. 108-112, 1976.

Notation

z	scalar
$\mathbf{z}$	column vector
z_i	ith element of $\mathbf{z}$
$\mathbf{z}_{0:n}$	vector $\mathbf{z}_{0:n} \triangleq (z_0, z_1, ..., z_n)^{\mathtt{T}}$
$\mathcal{Z}_n = \{1, \ldots, n\} \backslash \{\boldsymbol{\tau}_k\}$	$\mathcal{Z}_n = \bigcup_{j=0}^{k} \{\tau_j + 1, \ldots, \tau_{j+1} - 1\}$, where $\tau_0 = 0$ and $\tau_{k+1} = n$
$\mathbf{I}_n$	identity matrix of dimension $n \times n$
$\mathbf{A}$	matrix
$\mathbf{A}^{\mathtt{T}}$	transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of matrix $\mathbf{A}$
$\lvert\mathbf{A}\rvert$	determinant of matrix $\mathbf{A}$
$\mathbb{I}_E(\mathbf{z})$	indicator function of the set E (1 if $\mathbf{z} \in E$, 0 otherwise)
$\mathbf{z} \sim p(\mathbf{z})$	$\mathbf{z}$ is distributed according to distribution $p(\mathbf{z})$
$\mathbf{z} \vert \mathbf{y} \sim p(\mathbf{z})$	the conditional distribution of $\mathbf{z}$ given $\mathbf{y}$ is $p(\mathbf{z})$

Probability distribution	$\mathcal{F}$	$f_{\mathcal{F}}(\cdot)$	
Inverse Gamma	$\mathcal{IG}(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \exp(-\beta/z) \mathbb{I}_{(0,+\infty)}(z)$,	$\alpha > 0, \beta > 0.$
Gamma	$\mathcal{G}a(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp(-\beta z) \mathbb{I}_{(0,+\infty)}(z)$,	$\alpha > 0, \beta > 0.$
Gaussian	$\mathcal{N}(\mathbf{m}, \Sigma)$	$\lvert 2\pi \boldsymbol{\Sigma} \rvert^{-1/2} \exp\left(-\frac{1}{2} (\mathbf{z} - \mathbf{m})^t \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \mathbf{m})\right)$.	
Beta	$\mathcal{B}e(\alpha, \beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1} (1-z)^{\beta-1} \mathbb{I}_{(0,1)}(z)$,	$\alpha > 0, \beta > 0.$
Uniform	$\mathcal{U}_A$	$\left[\int_A d\mathbf{z}\right]^{-1} \mathbb{I}_A(\mathbf{z})$.	
Binomial	$\mathcal{B}i(\lambda, n)$	$\binom{n}{z} \lambda^z (1-\lambda)^{n-z} \mathbb{I}_{\mathbb{N}}(z)$,	$0 < \lambda < 1$, $n \in \mathbb{N}$.