Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach

Nicolas Dobigeon[†], Jean-Yves Tourneret[†] and Manuel Davy[*]

E-mail : {Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr, Manuel.Davy@ec-lille.fr

TECHNICAL REPORT – 2006, March

[†]IRIT/ENSEEIHT/TéSA, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

[*]LAGIS, BP 48, Cité Scientifique, 59651 Villeneuve d'Ascq cedex, France

Abstract

We propose a joint segmentation algorithm for piecewise constant AR processes recorded by several independent sensors. The algorithm is based on a hierarchical Bayesian model. Appropriate priors allow to introduce correlations between the change locations of the observed signals. Numerical problems inherent to Bayesian inference are solved by a Gibbs sampling strategy. The proposed joint segmentation methodology yields improved segmentation results when compared to parallel and independent individual signal segmentations. The initial algorithm is derived for piecewise constant autoregressive processes whose orders are fixed on each segment. However, several extensions are discussed allowing to handle models with unknown orders as well as signals with different dynamics. Simulation results conducted with arc-tracking and speech signals illustrate the performance of the algorithm.

I. Introduction

In many practical situations, one tools up some process with a collection of sensors, each of which delivering a time series. When the aim is process monitoring, an important task is to detect abrupt changes that occur in the sensor signals, and that may be related to a change in the process itself. Important such cases are in vibration monitoring of gearboxes, segmentation of multiple-track audio, etc. Using several sensors makes the detection more accurate, but a practical difficulty is about the fusion of the detections made on each signal. An alternative solution consists of implementing joint abrupt change detection over all the sensors.

This paper addresses the problem of segmenting correlated signals recorded from several sensors. Of course, signal segmentation has already received much attention in the signal processing literature (see for instance the textbooks [1]–[3] and references therein). Recent advances can be mainly divided into two categories. The first class of methods consists of penalizing a data based criterion in order to avoid over-segmentation. Different approaches have been recently proposed to determine the appropriate penalization for segmentation [4]–[6]. The second class of methods relies on Bayesian inference. The choice of appropriate priors for the unknown parameters induce penalization on the data-driven criterion built from the likelihood of the observations. The standard Bayesian estimators including the maximum *a Posteriori* (MAP) and the minimum mean square error (MMSE) estimators can then be derived. The computational complexity inherent to these change-point estimators is usually bypassed by using Markov chain Monte Carlo (MCMC) methods [7]–[9]. One recurrent problem with this kind of methodology is hyperparameter estimation. There are mainly two directions which can be followed to estimate these hyperparameters. The first approach couples MCMCs with an expectation maximization (EM) algorithm or a stochastic approximation (SAEM) [10], [11]. The second approach defines non-informative prior distributions for the hyperparameters introducing a second level of hierarchy within the Bayesian paradigm. This results in a so-called hierarchical Bayesian model. The hyperparameters are then integrated out from the joint posterior distribution or estimated from the observed data [9].

The main contribution of this paper is to study a joint segmentation procedure which allows one to handle signals recorded from different sensors. The proposed approach introduces correlations between the change-points of the observed signals. More precisely, when a change is detected in

one or several signals at a given time location, the proposed algorithm favors the occurence of a change at this time location in the other signals. This change-point correlation is built within a Bayesian framework by defining appropriate change-point priors. The proposed methodology is very similar to the hierarchical Bayesian curve fitting technique studied in [9]. However, the segmentation procedure studied in this paper allows joint segmentation of signals recorded by different sensors, contrary to the algorithm proposed in [9]. This is to our knowledge the first time a fully Bayesian algorithm is developed for joint segmentation of piecewise constant AR processes.

A. Notations and problem formulation

In this paper, we consider that J sensors deliver J signals (also referred to as *observations*), whose sample size is n. Individual signals are denoted in vector form as $\mathbf{y}_j = \left[y_{j,1}, \ldots, y_{j,n}\right]$ for $j = 1, \ldots, J$, where $y_{j,i}$ is the sample of signal j at time i. Each of the J signals is modeled as a piecewise constant AR process as follows:

$$y_{j,i} = \sum_{l=1}^{p} a_{j,k,l} y_{j,i-l} + e_{j,i}, \tag{1}$$

where $k = 1, \ldots, K_j$ is the segment index which refers to one of the $\mathbf{y}_j$ portions where the AR process is stationary. In each of these K_j segments, for signal $\#j$, the set of AR parameters is denoted in vector form as $\mathbf{a}_{j,k} = \left[a_{j,k,1}, \ldots, a_{j,k,p}\right]^{\mathsf{T}}$. The poles of the AR processes are supposed to be inside the unit circle, ensuring stationarity and causality on each segment. The segment $\#k$ in the signal $\#j$ has boundaries denoted by $[l_{j,k-1} + 1, l_{j,k}]$ where $l_{j,k}$ is the time index immediately after which a change occurs, with the convention that $l_{j,0} = 0$ and $l_{j,K_j} = n$. Finally, $\mathbf{e}_j = \left[e_{j,1}, \ldots, e_{j,n}\right]$ is a vector of i.i.d. zero mean Gaussian noise samples. The noise vectors $\mathbf{e}_1, \ldots, \mathbf{e}_J$ are assumed independent.

Modeling the observations as AR processes can be motivated as follows: for any continuous spectral density $S(f)$, an AR process can be found with a spectral density arbitrary close to $S(f)$ [12, p. 130]. Many authors have followed this viewpoint in change detection algorithms, including [13], [14]. We assume in a first step that the orders of the AR models in Eq. (1) are all equal to p. This assumption is actually only aimed at simplifying the presentation. A more general model, where the (unknown) orders of the AR models on each segments are assumed unrelated from one segment to another, and from one signal to another, is derived later in this

paper. By using the notation $\mathbf{x}_{j,i:i'} = \left[x_{j,i}, \ldots, x_{j,i'}\right]$, the set of equations (1) can be written in the following matrix form:

$$\mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}^{\mathsf{T}} = \mathbf{Y}_{j,k}\mathbf{a}_{j,k} + \mathbf{e}_{j,l_{j,k-1}+1:l_{j,k}}^{\mathsf{T}}, \tag{2}$$

where $\mathbf{Y}_{j,k}$ denotes a matrix of size $(l_{j,k} - l_{j,k-1}) \times p$:

$$\mathbf{Y}_{j,k} = \begin{bmatrix} y_{j,l_{j,k-1}} & y_{j,l_{j,k-1}-1} & \cdots & y_{j,l_{j,k-1}-p+1} \\ y_{j,l_{j,k-1}+1} & y_{j,l_{j,k-1}} & \cdots & y_{j,l_{j,k-1}-p+2} \\ \vdots & \vdots & \vdots & \vdots \\ y_{j,l_{j,k}-1} & y_{j,l_{j,k}-2} & \cdots & y_{j,l_{j,k}-p} \end{bmatrix}. \tag{3}$$

This paper proposes a Bayesian framework as well as an efficient algorithm aimed at estimating the change-point locations $l_{j,k}$ from the J observed time series $\mathbf{y}_j$, $j = 1, \ldots, J$.

B. Paper organization

The Bayesian model used for joint change-point detection is presented in Section II. This model requires to adjust hyperparameters related to the change-point location, AR parameter and noise variance priors. The proposed methodology assigns vague priors to the unknown hyperparameters. The hyperparameters are then integrated out from the joint posterior or estimated from the observed data. This results in a hierarchical Bayesian model described in Section II. An appropriate Gibbs sampler studied in Section III allows one to generate samples distributed according to the change-point posterior. The sampler convergence properties are investigated through simulations presented in Section IV. Some alternative models are introduced in Section V, including AR models whose orders on each signal segment are unknown. Section VI studies the performance of the proposed joint procedure for arc-tracking detection and speech segmentation. Conclusions are reported in Section VII.

II. Hierarchical Bayesian Model

The joint abrupt change detection problem presented in the previous section is based on the estimation of the unknown parameters K_j (numbers of segments), $l_{j,k}$ (change-point locations), $\sigma_{j,k}^2$ (noise variances, with $\boldsymbol{\sigma}_j^2 = \left[\sigma_{j,1}^2, \ldots, \sigma_{j,K_j}^2\right]^{\mathsf{T}}$) and $\mathbf{a}_{j,k}$ (AR parameter vectors which are

denoted jointly as $\mathbf{A}_j = \{\mathbf{a}_{j,1}, \ldots, \mathbf{a}_{j,K_j}\}$ for signal $\#j$). A standard re-parameterization consists of introducing indicator variables $\mathrm{r}_{j,i}$ ($j \in \{1, \ldots, J\}$, $i \in \{1, \ldots, n\}$) such that:

$$\begin{cases} \mathrm{r}_{j,i} = 1 \text{ if there is a change-point at time } i \text{ in signal } \#j, \\ \mathrm{r}_{j,i} = 0 \text{ otherwise}, \end{cases}$$

with $\mathrm{r}_{j,n} = 1$ (this condition ensures that the number of change-points equals the number of segments in signal $\#j$, that is $K_j = \sum_{i=1}^{n} \mathrm{r}_{j,i}$). Using these indicator variables, the unknown parameter vector is $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J\}$, where $\boldsymbol{\theta}_j = \left(\mathbf{r}_j, \boldsymbol{\sigma}_j^2, \mathbf{A}_j\right)$ and $\mathbf{r}_j = \left[\mathrm{r}_{j,1}, \ldots, \mathrm{r}_{j,n}\right]$. It is important to note that the parameter vector $\boldsymbol{\theta}$ belongs to a space whose dimension depends on K_j, i.e., $\boldsymbol{\theta} \in \Theta = \{0,1\}^{nJ} \times \prod_{j=1}^{J} (\mathbb{R}^+ \times \mathbb{R}^p)^{K_j}$. This paper proposes a Bayesian approach to the estimation of the unknown parameter vector $\boldsymbol{\theta}$. Bayesian inference on $\boldsymbol{\theta}$ is based on the posterior distribution $f(\boldsymbol{\theta}|\mathbf{Y})$, with $\mathbf{Y} = \left[\mathbf{y}_1, \ldots, \mathbf{y}_J\right]^{\mathsf{T}}$, which is related to the observations likelihood and to the parameter priors via Bayes rule $f(\boldsymbol{\theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})$. The likelihood and priors used for the joint abrupt change detection are presented below.

A. Approximate Likelihood

Though the likelihood of a single AR model is easy to write exactly, the likelihood of a *piecewise stationary* AR model is much more complicated, as each stationary segment needs to be initialized using the samples from the previous segment. In many works, the dependence of the exact likelihood $f(\mathbf{y}_j|\boldsymbol{\theta}_j)$ on the p first samples $\mathbf{y}_{j,1:p}$ is omitted, (see [15, p. 186] for more details), and we adopt this approximation. In other words, by using the independence assumption between the noise vectors $\mathbf{e}_j$, $j \in \{1, \ldots, J\}$, the exact likelihood of $\mathbf{Y}$ is approximated as follows:

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}) &\approx \prod_{j=1}^{J} f(\mathbf{y}_{j,p+1:n}|\mathbf{y}_{j,1:p}, \boldsymbol{\theta}_j) \\ &\approx \prod_{j=1}^{J} \prod_{k=1}^{K_j} \frac{1}{\left(2\pi\sigma_{j,k}^2\right)^{n_{j,k}(\mathbf{r}_j)/2}} \exp\left(-\frac{E_{j,k}(\mathbf{r}_j)}{2\sigma_{j,k}^2}\right), \end{aligned} \tag{4}$$

where $n_{j,k}(\mathbf{r}_j) = l_{j,k} - l_{j,k-1}$ is the length of segment $\#k$ in signal $\#j$ and

$$E_{j,k}(\mathbf{r}_j) \triangleq \left\| \mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}^{\mathsf{T}} - \mathbf{Y}_{j,k}\mathbf{a}_{j,k} \right\|^2, \tag{5}$$

where $\|\mathbf{x}\|^2 = \mathbf{x}^{\mathsf{T}}\mathbf{x}$.

B. Parameter Priors

In our approach, the abrupt changes are detected via the indicator variables $\mathbf{r}_j$, $j = 1, \ldots, J$ (we recall that there is one variable for each signal j, and one variable for each time index $i = 1, ..., n$). This section defines the indicator, variance and AR parameter priors.

1) Indicators: Possible correlations between change locations in the J observed signals are modeled by an appropriate prior distribution $f(\mathbf{R}|\mathbf{P})$, where $\mathbf{R} = \left[\mathbf{r}_1, \ldots, \mathbf{r}_J\right]^\mathsf{T}$ and $\mathbf{P}$ is defined below. Before being more precise, we define a *global abrupt change configuration* as follows: the matrix $\mathbf{R}$ is composed of 0's and 1's, and a global configuration is a specific instance of this matrix. In our formulation, this corresponds to a specific solution to the joint abrupt change detection problem. A *local abrupt change configuration*, denoted $\boldsymbol{\epsilon}$ (where $\boldsymbol{\epsilon} \in \mathcal{E} = \{0, 1\}^J$), is a specific instance of a column of $\mathbf{R}$: this corresponds to a the presence/asbsence of abrupt changes at a given time, across the J signals.

Denote as $\mathrm{P}_{\boldsymbol{\epsilon}}$ the probability of having a local abrupt change configuration $\boldsymbol{\epsilon}$ at time i, that is, of having $\left[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}\right]^\mathsf{T} = \boldsymbol{\epsilon}$. We first assume that $\mathrm{P}_{\boldsymbol{\epsilon}}$ does not depend on the time index i. As a consequence, by assuming that $\left[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}\right]$ is independent of $\left[\mathrm{r}_{1,i'}, \ldots, \mathrm{r}_{J,i'}\right]$ for any $i \neq i'$, the indicator prior distribution is expressed as:

$$f(\mathbf{R}|\mathbf{P}) = \prod_{\boldsymbol{\epsilon} \in \mathcal{E}} \mathrm{P}_{\boldsymbol{\epsilon}}^{S_{\boldsymbol{\epsilon}}(\mathbf{R})}, \tag{6}$$

where $\mathbf{P} = \{\mathrm{P}_{\boldsymbol{\epsilon}}\}_{\boldsymbol{\epsilon} \in \mathcal{E}}$ and $S_\epsilon(\mathbf{R})$ is the number of times i such that $\left[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}\right]^\mathsf{T} = \boldsymbol{\epsilon}$. For example, in the case of two observed signals $\mathbf{y}_1$ and $\mathbf{y}_2$ (i.e., $J = 2$), the prior distribution of $\mathbf{R}$ can be written as:

$$f(\mathbf{R}|\mathbf{P}) = \mathrm{P}_{00}^{S_{00}} \mathrm{P}_{10}^{S_{10}} \mathrm{P}_{01}^{S_{01}} \mathrm{P}_{11}^{S_{11}}, \tag{7}$$

where $S_{00} = \sum_{i=1}^{n-1} (1 - \mathrm{r}_{1,i})(1 - \mathrm{r}_{2,i})$, $S_{11} = \sum_{i=1}^{n-1} \mathrm{r}_{1,i}\mathrm{r}_{2,i}$, $S_{10} = \sum_{i=1}^{n-1} \mathrm{r}_{1,i}(1 - \mathrm{r}_{2,i})$ and $S_{01} = \sum_{i=1}^{n-1} (1 - \mathrm{r}_{1,i})\mathrm{r}_{2,i}$. With this prior, a high value of $\mathrm{P}_{\boldsymbol{\epsilon}}$ indicates a very likely configuration $\left[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}\right]^\mathsf{T} = \boldsymbol{\epsilon}$ for all $i = 1, \ldots, n$. For instance, by choosing a high value of $\mathrm{P}_{0\ldots0}$ (resp. $\mathrm{P}_{1\ldots1}$), we will favor a simultaneous absence (resp. presence) of changes in all observed signals. This choice introduces correlation between the change-point locations.

2) Variances and AR parameters: : Inverse-Gamma distributions are selected for the noise variances:

$$\sigma_{j,k}^2 \mid \left(\frac{\nu}{2}, \frac{\gamma}{2}\right) \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right), \tag{8}$$

where $\mathcal{IG}(a,b)$ denotes the Inverse-Gamma distribution with parameters a and b, $\nu = 2$ (as in [9]) and γ is an adjustable hyperparameter. This is a so-called *conjugate prior* which has been used successfully in [9] for Bayesian curve fitting. We assume here that the hyperparameter γ is the same for all the observed signals. Note, however, that a similar analysis could be conducted with a set of non-equal hyperparameters $\gamma_j, j = 1,\ldots,J$. Such analysis is interesting when signal amplitudes differ significantly from one signal to another, and it will be implemented in Subsection V-A.

Conjugate zero-mean Gaussian priors are chosen for the AR parameters:

$$\mathbf{a}_{j,k}|\sigma^2_{j,k},\delta^2_0 \sim \mathcal{N}\left(\mathbf{0}_p, \sigma^2_{j,k}\delta^2_0\mathbf{I}_p\right), \tag{9}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix, $\mathbf{0}_p$ is the vector made of p zeros and δ^2_0 is an adjustable hyperparameter. One motivation for selecting conjugate priors is that they allow to integrate out (marginalize) the noise variances and AR parameters in the posterior $f(\boldsymbol{\theta}|\mathbf{Y})$, making the computations easier.

C. Hyperparameter priors

The hyperparameter vector associated with the parameter priors defined above is $\boldsymbol{\Phi} = (\mathbf{P}, \delta^2_0, \gamma)$. Of course, the ability of this Bayesian model to detect abrupt changes accurately in the J signals depends on the values of the hyperparameters. As outlined in Section I, these hyperparameters should be considered as unknown, and estimated as this makes the overall model more robust, see [9] for example. The resulting hierarchical model requires to define hyperparameter priors (sometimes referred to as hyper-priors) which are detailed below.

1) Hyperparameters δ^2_0 and γ: The priors for hyperparameters δ^2_0 and γ are selected as a noninformative Jeffreys' prior and a vague conjugate Inverse-Gamma distribution (i.e, with large variance) which reflect the lack of precise knowledge regarding these hyperparameters:

$$\delta^2_0|\xi,\beta \sim \mathcal{IG}\left(\xi,\beta\right), \qquad f(\gamma) = \frac{1}{\gamma}\mathbb{I}_{\mathbb{R}^+}(\gamma), \tag{10}$$

where $\mathbb{I}_{\mathbb{R}^+}(x)$ is the indicator function defined on $\mathbb{R}^+$.

2) Hyperparameter $\mathbf{P}$: The prior distribution for the hyperparameter $\mathbf{P}$ is a Dirichlet distribution with parameter vector $\boldsymbol{\alpha} = \left[\alpha_{0\ldots0},\ldots,\alpha_{1\ldots1}\right]$ defined on the simplex $\mathcal{P} = \{\mathbf{P} \text{ such that } \sum_{\boldsymbol{\epsilon}\in\mathcal{E}} \mathrm{P}_{\boldsymbol{\epsilon}} =$

$1, \mathrm{P}_{\boldsymbol{\epsilon}} > 0\}$ denoted as $\mathbf{P} \sim D_{2^J}(\boldsymbol{\alpha})$. This distribution has been chosen since it allows marginalization of the posterior distribution $f(\boldsymbol{\theta}|\mathbf{Y})$ with respect to $\mathbf{P}$. Moreover, by choosing $\alpha_{\boldsymbol{\epsilon}} = 1, \forall \boldsymbol{\epsilon} \in \mathcal{E}$, the Dirichlet distribution reduces to the uniform distribution on $\mathcal{P}$.

Assuming that the individual hyperparameters are independent, the full hyperparameter prior distribution $\boldsymbol{\Phi}$ can be written (up to a normalizing constant):

$$f(\boldsymbol{\Phi}|\boldsymbol{\alpha}, \xi, \beta) \propto \left(\prod_{\boldsymbol{\epsilon} \in \mathcal{E}} \mathrm{P}_{\boldsymbol{\epsilon}}^{\alpha_{\boldsymbol{\epsilon}} - 1} \right) \frac{1}{\gamma} \frac{\beta^{\xi}}{\Gamma(\xi)(\delta_0^2)^{\xi+1}} \exp\left(-\frac{\beta}{\delta_0^2} \right) \mathbb{I}_{\mathbb{R}^+}(\gamma) \mathbb{I}_{\mathbb{R}^+}(\delta_0^2) \mathbb{I}_{\mathcal{P}}(\mathbf{P}), \tag{11}$$

where $\propto$ means "proportional to" and $\Gamma(\cdot)$ is the gamma function.

D. Posterior distribution of $\boldsymbol{\theta}$

The posterior distribution of the unknown parameter vector $\boldsymbol{\theta}$ can be computed from the following hierarchical structure:

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \int f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y}) d\boldsymbol{\Phi} \propto \int f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}) d\boldsymbol{\Phi}, \tag{12}$$

where

$$f(\boldsymbol{\theta}|\boldsymbol{\Phi}) = f(\mathbf{R}|\mathbf{P}) \prod_{j=1}^{J} \prod_{k=1}^{K_j} f\left(\mathbf{a}_{j,k}|\sigma_{j,k}^2, \delta_0^2\right) f\left(\sigma_{j,k}^2|\frac{\nu}{2}, \frac{\gamma}{2}\right), \tag{13}$$

and $f(\mathbf{Y}|\boldsymbol{\theta})$ and $f(\boldsymbol{\Phi})$ are defined in Eq.'s (4) and (11). This hierarchical structure allows to integrate out the nuisance parameters $\boldsymbol{\sigma}^2 = \{\boldsymbol{\sigma}_1^2, \ldots, \boldsymbol{\sigma}_J^2\}$, $\mathbf{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_J\}$ and $\mathbf{P}$ from the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y})$, yielding:

$$\begin{aligned} f(\mathbf{R}, \gamma, \delta_0^2|\mathbf{Y}) \propto \frac{1}{\gamma} \mathbb{I}_{\mathbb{R}^+}(\gamma) \left(\delta_0^2\right)^{-\frac{p}{2}\sum_{j=1}^{J} K_j(\mathbf{r}_j)} f(\delta_0^2|\xi, \beta) C(\mathbf{R}|\mathbf{Y}, \boldsymbol{\alpha}) \\ \times \prod_{j=1}^{J} \prod_{k=1}^{K_j(\mathbf{r}_j)} \left(\frac{\gamma^{\frac{\nu}{2}} \left|\mathbf{M}_{j,k}\right|^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)\right)}{\left(\gamma + T_{j,k}^2\right)^{\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)}} \right), \end{aligned} \tag{14}$$

with

$$\begin{cases} T_{j,k}^2 = \mathbf{y}_{j, l_{j,k-1}+1:l_{j,k}}^{\mathsf{T}} \, \mathbf{Q}_{j,k} \mathbf{y}_{j, l_{j,k-1}+1:l_{j,k}}, \\ \mathbf{Q}_{j,k} = \mathbf{I}_p - \mathbf{Y}_{j,k} \mathbf{M}_{j,k} \mathbf{Y}_{j,k}^{\mathsf{T}}, \\ \mathbf{M}_{j,k} = \left(\mathbf{Y}_{j,k}^{\mathsf{T}} \mathbf{Y}_{j,k} + \dfrac{\mathbf{I}_p}{\delta_0^2} \right)^{-1}, \end{cases} \tag{15}$$

and

$$C(\mathbf{R}|\mathbf{Y}) = \frac{\prod_{\boldsymbol{\epsilon} \in \{0,1\}^J} \Gamma\left(S_{\boldsymbol{\epsilon}}(\mathbf{R}) + \alpha_{\boldsymbol{\epsilon}}\right)}{\Gamma\left(\sum_{\boldsymbol{\epsilon} \in \{0,1\}^J} \left(S_{\boldsymbol{\epsilon}}(\mathbf{R}) + \alpha_{\boldsymbol{\epsilon}}\right)\right)}. \tag{16}$$

The posterior distribution in Eq. (14) is too complex to enable the closed-form calculation of Bayesian estimators (e.g., MMSE or MAP) for the unknown parameters. In this case, it is very usual to apply MCMC methods to generate samples which are asymptotically distributed according to the posteriors of interest. The samples can then be used to estimate the unknown parameters by replacing integrals by empirical averages over the MCMC samples.

Here, we propose a Gibbs sampler strategy that is similar to that in [9], with two noticeable differences, however: 1) our approach enables to perform joint signal segmentation and 2) the use of indicator variables sets our model into a fixed dimensional space, which avoids the costly implementation of reversible jumps. Section III presents the MCMC algorithm designed to perform the joint abrupt change detection in the case where the orders or the AR models, as wall as the hyperparameter γ, are the same for all the signals. These assumptions will be removed in Section V.

III. A Gibbs Sampler for joint signal segmentation

Gibbs sampling is an iterative sampling strategy which consists of generating random samples (denoted by $\cdot^{(t)}$ where t is the iteration index) distributed according to the conditional posterior distributions of each parameter. This paper proposes to sample according to the distribution $f(\mathbf{R}, \gamma, \delta_0^2|\mathbf{Y})$ defined in Eq. (14) by the three step procedure outlined below. The main steps of Algorithm 1, as well as the key equations, are detailed in Subsections III-A to III-C below.

A. Generation of samples according to $f(\mathbf{R}|\gamma, \delta_0^2, \mathbf{Y})$

This step is achieved by using the Gibbs Sampler, to generate Monte Carlo samples distributed according to $f\left(\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}|\gamma, \delta_0^2, \mathbf{Y}\right)$. This vector is a random vector of Booleans in $\mathcal{E}$. Consequently, its distribution is fully characterized by the probabilities $P\left(\left[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}\right]^{\mathsf{T}} = \boldsymbol{\epsilon}|\gamma, \delta_0^2, \mathbf{Y}\right), \boldsymbol{\epsilon} \in \mathcal{E}$. By using the notation $\mathbf{R}_{-i}$ to denote the matrix $\mathbf{R}$ where the column at time i is removed, the following result can be obtained:

$$P\left(\left[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}\right]^{\mathsf{T}} = \boldsymbol{\epsilon}|\mathbf{R}_{-i}, \gamma, \delta_0^2, \mathbf{Y}\right) \propto f(\mathbf{R}_i(\boldsymbol{\epsilon}), \gamma, \delta_0^2|\mathbf{Y}), \tag{17}$$

where $\mathbf{R}_i(\boldsymbol{\epsilon})$ is the matrix $\mathbf{R}$ where the column at time i is replaced by the vector $\boldsymbol{\epsilon}$. This yields a closed-form expression of the probabilities $P\left(\left[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}\right]^{\mathsf{T}} = \boldsymbol{\epsilon}|\mathbf{R}_{-i}, \gamma, \delta_0^2, \mathbf{Y}\right)$ after appropriate normalization.

Algorithm 1: Gibbs sampling algorithm for abrupt change detection

- Initialization:
  - Sample hyperparameter vector $\widetilde{\mathbf{\Phi}}^{(0)} = \left(\widetilde{\delta}_0^{2(0)}, \widetilde{\gamma}^{(0)}, \widetilde{\mathbf{P}}^{(0)}\right)$ from the pdf in Eq. (11),
  - For $i = 1, \ldots, n-1$ sample, $\left[\widetilde{\mathrm{r}}_{1,i}^{(0)}, \ldots, \widetilde{\mathrm{r}}_{J,i}^{(0)}\right]$ from the pdf in Eq. (6),
  - For $j = 1, \ldots, J$, $k = 1, \ldots, K$, sample $\widetilde{\sigma}_{j,k}^{2(0)}$ and $\widetilde{\mathbf{a}}_{j,k}^{(0)}$ from the pdf's in Eq.'s (8) and (9) ,
  - Set $t \leftarrow 1$,
- Iterations: for $t = 1, 2, 3, \ldots,$ do
  - For each time instant $i = 1, \ldots, n-1$, sample the local abrupt change configuration at time i $\left[\widetilde{\mathrm{r}}_{1,i}^{(t)}, \ldots, \widetilde{\mathrm{r}}_{J,i}^{(t)}\right]$ from its conditional distribution given in Eq. (17),
  - For signals $j = 1, \ldots, J$, and segments $k = 1, \ldots, K$, sample the noise variance $\widetilde{\sigma}_{j,k}^{2(t)}$ from its conditional posterior given in Eq. (18),
  - Sample the hyperparameter $\widetilde{\gamma}^{(t)}$ from its posterior given in Eq. (19),
  - For signals $j = 1, \ldots, J$ and segments $k = 1, \ldots, K$, sample the AR coefficients $\widetilde{\mathbf{a}}_{j,k}^{(t)}$ from their conditional posterior given in Eq. (20),
  - Sample the hyperparameter $\widetilde{\delta}_0^{2(t)}$ from its conditional posterior given in Eq. (21),
  - *Optional step:* sample the hyperparameter $\widetilde{\mathbf{P}}^{(t)}$ from the pdf in Eq (22),
  - Set $t \leftarrow t + 1$.

B. Generation of samples according to $f(\gamma, \delta_0^2|\mathbf{R}, \mathbf{Y})$

To obtain samples distributed according to $f(\gamma, \delta_0^2|\mathbf{R}, \mathbf{Y})$, it is very convenient to generate vectors distributed according to the joint distribution $f(\gamma, \delta_0^2, \boldsymbol{\sigma}^2, \mathbf{A}|\mathbf{R}, \mathbf{Y})$ by using Gibbs moves. By looking carefully at the joint distribution of $f(\boldsymbol{\theta}, \mathbf{\Phi}|\mathbf{Y})$, this step can be decomposed as follows:

- **Generate samples according to** $f(\gamma, \boldsymbol{\sigma}^2|\mathbf{R}, \delta_0^2, \mathbf{Y})$

By integrating the joint distribution $f(\boldsymbol{\theta}, \mathbf{\Phi}|\mathbf{Y})$ with respect to the AR parameters, the following results can be obtained:

$$\sigma_{j,k}^2|\mathbf{R}, \gamma, \delta_0^2, \mathbf{Y} \sim \mathcal{IG}\left(\frac{\nu + n_{j,k}(\mathbf{r}_j)}{2}, \frac{\gamma + T_{j,k}^2}{2}\right), \tag{18}$$

$$\gamma|\mathbf{R}, \boldsymbol{\sigma}^2 \sim \mathcal{G}\left(\frac{\nu}{2}\sum_{j=1}^{J} K_j(\mathbf{r}_j), \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{K_j(\mathbf{r}_j)} \frac{1}{\sigma_{j,k}^2}\right), \tag{19}$$

where $\mathcal{G}(a,b)$ is the Gamma distribution with parameters (a,b).

- **Generate samples according to** $f(\delta_0^2, \mathbf{A}|\mathbf{R}, \boldsymbol{\sigma}^2, \mathbf{Y})$

This is achieved as follows:

$$\mathbf{a}_{j,k}|\mathbf{R}, \boldsymbol{\sigma}^2, \delta_0^2, \mathbf{Y} \sim \mathcal{N}\left(\boldsymbol{\mu}_{j,k}, \sigma_{j,k}^2 \mathbf{M}_{j,k}\right), \tag{20}$$

$$\delta_0^2|\mathbf{R}, \mathbf{A}, \boldsymbol{\sigma}^2 \sim \mathcal{IG}\left(\xi + \frac{p}{2}\sum_{j=1}^{J} K_j(\mathbf{r}_j), \beta + \sum_{j=1}^{J}\sum_{k=1}^{K_j(\mathbf{r}_j)} \frac{\|\mathbf{a}_{j,k}\|^2}{2\sigma_{j,k}^2}\right), \tag{21}$$

with $\boldsymbol{\mu}_{j,k} = \mathbf{M}_{j,k}\mathbf{Y}_{j,k}^{\mathsf{T}}\mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}$.

C. Posterior distribution of $\mathrm{P}_{\boldsymbol{\epsilon}}$

The hyperparameters $\mathrm{P}_{\boldsymbol{\epsilon}}$, $\boldsymbol{\epsilon} \in \mathcal{E}$, carry information regarding the correlations between the change locations in the different time series. As a consequence it is interesting for practical applications to estimate them from their posterior distribution (which is Dirichlet):

$$\mathbf{P}|\mathbf{R}, \mathbf{Y} \sim \mathcal{D}_{2^J}(S_\epsilon(\mathbf{R}) + \alpha_{\boldsymbol{\epsilon}}). \tag{22}$$

IV. Segmentation of synthetic data

The simulation presented in this section have been obtained for $J = 2$ with sample size $n = 300$. The change-point locations are $\boldsymbol{l}_1 = (60, 150)$ for signal $\#1$ and $\boldsymbol{l}_2 = (60)$ for signal $\#2$. The parameters of the two AR processes are summarized in Table I. The fixed parameters and hyperparameters have been chosen as follows: $\nu = 2$ (as in [9]), $\xi = 1$ and $\beta = 100$ (vague hyperprior), $\alpha_{\boldsymbol{\epsilon}} = \alpha = 1, \forall \boldsymbol{\epsilon} \in \mathcal{E}$. The hyperparameters α_ϵ are equal to insure the Dirichlet distribution reduces to a uniform distribution. Moreover, the common value to the hyperparameters α_ϵ has been set to $\alpha = 1 \ll n$ in order to reduce the influence of this parameter in the posterior (22). In order to speed up the computations, the quantities $T_{j,k}^2$, $\mathbf{Q}_{j,k}$ and $\mathbf{M}_{j,k}$ defined in Eq. (15) have been computed following the implementations described in [16] and reported in the Appendix. All figures have been obtained after averaging the results of 64 Markov chains. The total number of runs for each Markov chain is $N_{MC} = 700$, including $N_{bi} = 200$ burn-in iterations. Thus only the last 500 Markov chain output samples are used for the estimations (the choice of parameters N_{MC} and N_{bi} will be discussed later). Note that running 100 iterations of the proposed algorithm for joint segmentation takes approximately 2 minutes and 30 seconds for a MATLAB implementation on a 2.8 Ghz Pentium IV. However, the computational cost may be prohibitive for very long time series.

TABLE I

PARAMETERS OF THE AR MODEL AND NOISE VARIANCES FOR EACH SEGMENT OF EACH SEQUENCE.

Sequence	Segment	$\sigma^2_{j,k}$	$a_{j,k,l}$					
$j=1$	$k=1$	0.50	0.0746	0.1664	-0.0693	-0.1571	-0.3543	-0.4277
	$k=2$	0.52	0.0135	0.1525	0.8170	2.3037	3.5316	2.8567
	$k=3$	3.80	0.0189	-0.0571	0.1502	-0.3173	0.4824	0.1607
$j=2$	$k=1$	0.81	0.0011	-0.0104	0.0538	-0.0646	0.3713	-0.0717
	$k=2$	4.63	0.0074	0.0138	0.1244	0.2660	0.7677	0.8705

A. Posterior distributions of the change-point locations

The first simulation shows the interest of joint segmentation compared to signal-by-signal segmentation for two independent autoregressive processes. Fig. 1 shows the posterior distributions of the change-locations obtained for the two time-series. As can be seen, the change-point of the second time-series can be detected when using our joint segmentation technique whereas it is not detected when applying two single signal independent segmentations. When joint segmentation is performed, the change point located at time $i = 60$ in the second signal favors the detection of a change at the same time index in the other signal. Note that the results presented in Fig. 1 (left) obtained with $J = 1$ (univariate segmentation procedure) correspond exactly to the Bayesian curve fitting strategy of Punskaya *et al.* [9].

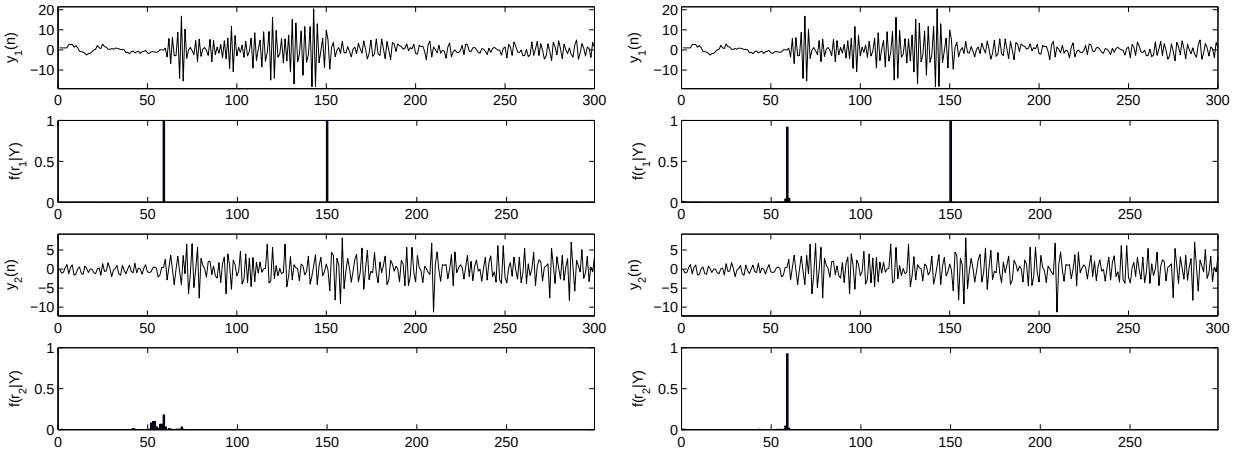

Fig. 1. Posterior distributions of the change-point locations for 1D (left) and joint segmentations (right) obtained after $N_{bi} = 200$ burn-in iterations and $N_r = 500$ iterations of interest.

B. Posterior distribution of the change-point numbers.

The estimation of the total number of change-points for the two time-series is an important problem. The proposed algorithm generates samples $(\mathbf{R}^{(t)}, \gamma^{(t)}, \delta_0^{2\,(t)})$ distributed according to the posterior distribution $f\left(\mathbf{R}, \gamma, \delta_0^2 | \, \mathbf{Y}\right)$, which allows for model selection. Indeed, for each sample $\mathbf{R}^{(t)}$, the number of change-points are $\widehat{K}_1^{(t)}(\mathbf{r}_1^{(t)}) = \sum_{i=1}^{N} \mathrm{r}_{1,i}^{(t)}$ and $\widehat{K}_2^{(t)}(\mathbf{r}_2^{(t)}) = \sum_{i=1}^{N} \mathrm{r}_{2,i}^{(t)}$. Fig. 2 shows the means of $\widehat{K}_1$ and $\widehat{K}_2$ as well as means $\pm$ standard deviations computed from the 500 last Markov chain samples with the joint approach. The histograms have maximum values for $K_1 = 3$ and $K_2 = 2$ which correspond to the actual numbers of changes.

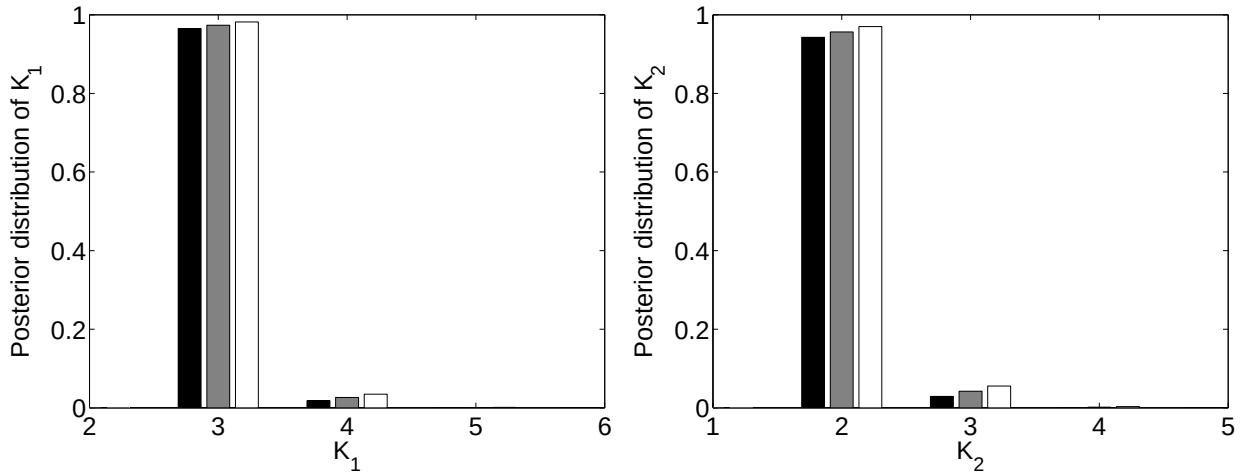

Fig. 2. Posterior distributions of the change-point numbers computed from $N_r = 500$ iterations of interest (mean in gray, mean $\pm$ standard deviations in white and black).

C. Noise variances and AR parameters

The estimation of the noise variances or AR parameters can be interesting in practical applications. Fig. 3 shows the posterior distributions of parameters $\{\sigma_{1,k}^2\}_{k=1,\ldots,3}$ associated with the time-series $\mathbf{y}_1$. These histograms are in good agreement with the actual values of the parameters $\sigma_{1,1}^2 = 0.50$, $\sigma_{1,2}^2 = 0.52$, $\sigma_{1,3}^2 = 3.80$ and $\sigma_{21}^2 = 0.81$, $\sigma_{22}^2 = 4.63$. Similar results could be obtained for AR parameters. They are omitted here for brevity.

D. Hyperparameter estimation

The last simulation results illustrate the performance of the hyperparameter estimation procedure. The estimated posteriors of hyperparameters P_{00}, P_{01}, P_{10} and P_{11} are depicted in Fig. 5.

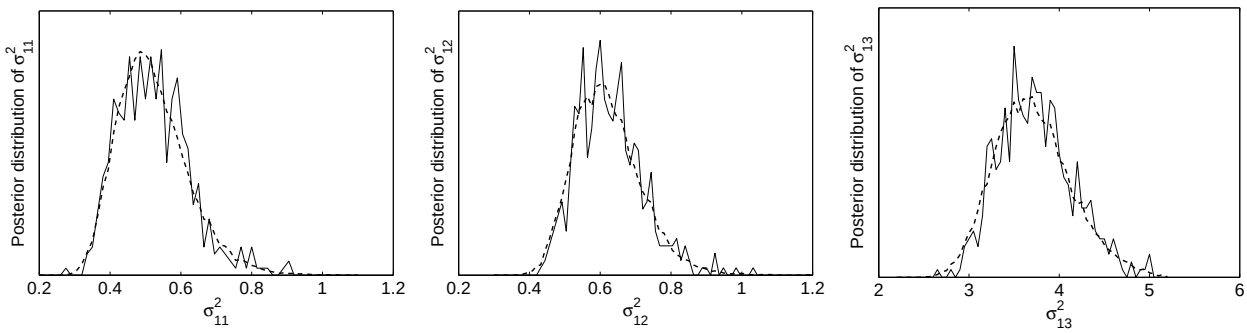

Fig. 3. Posterior distributions of the noise variances σ_{1i}^2 (for $i = 1, \ldots, 3$) conditioned to $K_1 = 3$ computed from $N_r = 500$ iterations of interest (solid lines). Averaged posterior distributions from 64 Markov chains (dotted lines).

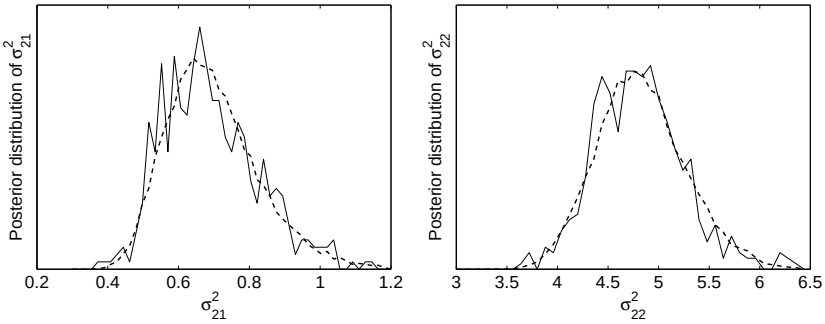

Fig. 4. Posterior distributions of the noise variances σ_{2i}^2 (for $i = 1, 2$) conditioned to $K_2 = 2$ computed from $N_r = 500$ iterations of interest (solid lines). Averaged posterior distributions from 64 Markov chains (dotted lines),

This shows that our Gibbs sampler actually generates samples distributed according to the true distribution in Eq. (22).

E. Sampler convergence

The Gibbs sampler allows to draw samples $\left(\mathbf{R}^{(t)}, \gamma^{(t)}, \delta_0^{2(t)}\right)$ asymptotically distributed according to $f(\mathbf{R}, \gamma, \delta_0^2|, \mathbf{Y})$. The change-point posterior probabilities can then be estimated by the empirical average (according to the minimum mean square error (MMSE) principle):

$$\hat{\mathbf{R}}_{\text{MMSE}} = \frac{1}{N_r} \sum_{t=1}^{N_r} \mathbf{R}^{(N_{bi}+t)}, \tag{23}$$

where N_{bi} is the number of burn-in iterations. However, two important questions are: 1) When can we decide that the samples $\left\{\mathbf{R}^{(t)}\right\}$ are actually distributed according to the target distribution? 2) How many samples are necessary to obtain an accurate estimate of $\mathbf{R}$ when using Eq. (23)?

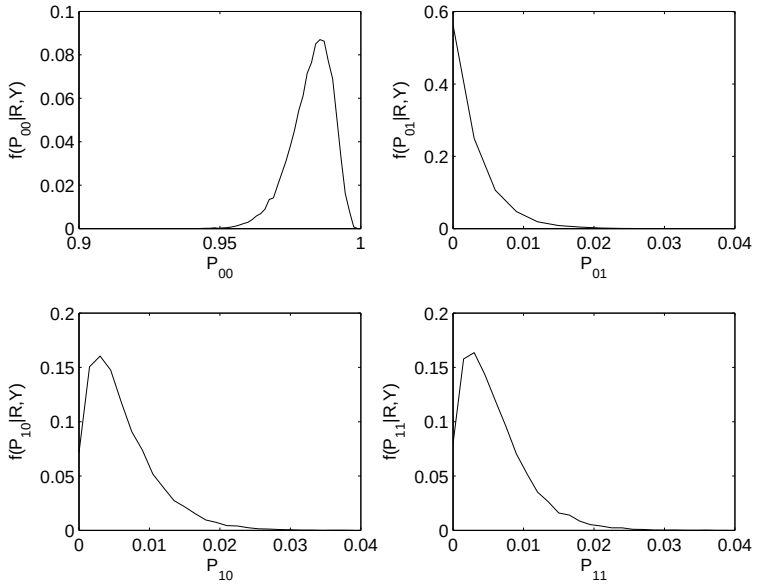

Fig. 5. Posterior distributions of the hyperparameters P_{ϵ} (computed from $N_r = 500$ iterations of interests of 64 Markov chains).

This section surveys some works allowing to determine appropriate values for parameters N_r and N_{bi}.

Running multiple chains with different initializations allows to define various convergence measures for MCMC methods [17]. This section proposes to use the popular between-within variance criterion to ensure the convergence of the algorithm. This method was initially studied by Gelman and Rubin in [18] and has been often used to monitor convergence (see for example [19], [20] or [17, p. 33]). This criterion requires to run M parallel chains of length N_r with different starting values. The between-sequence variance B and within-sequence variance W for the M Markov chains are defined by

$$B = \frac{N_r}{M-1} \sum_{m=1}^{M} \left(\overline{\kappa}_m - \overline{\kappa}\right)^2, \tag{24}$$

and

$$W = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{N_r} \sum_{t=1}^{N_r} \left(\kappa_m^{(t)} - \overline{\kappa}_m\right)^2, \tag{25}$$

with

$$\begin{cases} \overline{\kappa}_m = \frac{1}{N_r} \sum_{t=1}^{N_r} \kappa_m^{(t)}, \\ \overline{\kappa} = \frac{1}{M} \sum_{m=1}^{M} \overline{\kappa}_m, \end{cases} \tag{26}$$

where κ is the parameter of interest and $\kappa_m^{(t)}$ is the t^{th} run of the m^{th} chain. The convergence of the chain is monitored by a so-called *potential scale reduction factor* $\hat{\rho}$ defined as [21, p. 332]:

$$\sqrt{\hat{\rho}} = \sqrt{\frac{1}{W}\left(\frac{N_r - 1}{N_r} W + \frac{1}{N_r} B\right)}. \tag{27}$$

A value of $\sqrt{\hat{\rho}}$ close to 1 indicates a good convergence of the sampler.

Different choices for parameter κ could be considered for the proposed joint segmentation procedure. This paper proposes to monitor the convergence of the Gibbs sampler with the parameters $\mathrm{P}_{\boldsymbol{\epsilon}}$, $\boldsymbol{\epsilon} \in \mathcal{E}$. As an example, the outputs of $M = 5$ chains for parameter P_{00} are depicted in Fig. 6. The chains clearly converge to similar values. The potential scale reduction factors for all parameters $\mathrm{P}_{\boldsymbol{\epsilon}}$ are given in Table II. These values of $\sqrt{\hat{\rho}}$ confirm the good convergence of the sampler (a recommendation for convergence assessment is a value of $\sqrt{\hat{\rho}}$ below 1.2 [21, p. 332]). It is important to make the following comments:

- For segmentation purposes, the important information is contained in the change locations, which has motivated the choice of the parameters $\mathrm{P}_{\boldsymbol{\epsilon}}$ for monitoring convergence. However, for applications requiring signal reconstruction, the AR parameters $a_{j,k,l}$ and noise variances $\sigma_{j,k}^2$ are important parameters. Therefore, the potential scale reduction factors $\sqrt{\hat{\rho}}$ computed for the estimated variances are also indicated in Table II. The obtained values confirm that a burn-in of 200 iterations is sufficient for this example.
- Other simulation examples with smaller changes or closer changepoints can yield MCMC convergence problems. In such cases, an alternative based on perfect simulation might be implemented (see [22] for more details).

In order to determine the number of runs which are required to obtain an accurate estimate of $\mathbf{R}$ when using Eq. (23), an *ad hoc* approach consists of assessing convergence via appropriate graphical evaluations [17, p. 28]. Here, a reference estimate denoted as $\widetilde{\mathbf{R}}$ has been computed for a large number of iterations $N_r = 10000$ and $N_{bi} = 200$ (to ensure convergence of the sampler and good accuracy of the approximation in Eq. (23)). Figure 7 shows the mean square

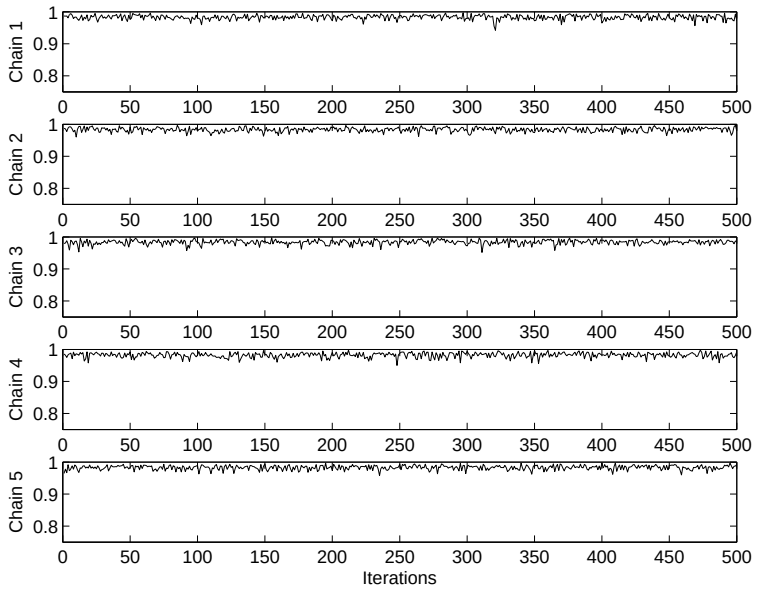

Fig. 6. Convergence assessment: the outputs of $M = 5$ chains for the parameter P_{00} converge to the same value.

TABLE II

POTENTIAL SCALE REDUCTION FACTORS OF $\mathrm{P}_{\boldsymbol{\epsilon}}$ (COMPUTED FROM $M = 64$ MARKOV CHAINS).

	$\mathrm{P}_{\boldsymbol{\epsilon}}$				$\sigma^2_{1,k}$			$\sigma^2_{2,k}$	
	P_{00}	P_{01}	P_{10}	P_{11}	$\sigma^2_{1,1}$	$\sigma^2_{1,2}$	$\sigma^2_{1,3}$	$\sigma^2_{2,1}$	$\sigma^2_{2,2}$
$\sqrt{\hat{\rho}}$	1.0005	1.0002	1.0006	0.9997	1.0005	1.0002	1.0006	1.0005	1.0002

error (MSE) between this reference estimate $\widetilde{\mathbf{R}}$ and the estimate obtained after $N_r = p$ runs (and $N_{bi} = 200$ burn-in iterations):

$$e_r^2(p) = \left\| \widetilde{\mathbf{R}} - \frac{1}{p} \sum_{t=1}^{p} \mathbf{R}^{(N_{bi}+t)} \right\|^2 .$$

This figure indicates that that a number of runs equal to $N_r = 500$ is sufficient to ensure an accurate estimation of the empirical average in Eq. (23) for this example. Of course, for more difficult problems, a larger number of runs will be necessary to obtain an accurate estimation of the posterior distribution.

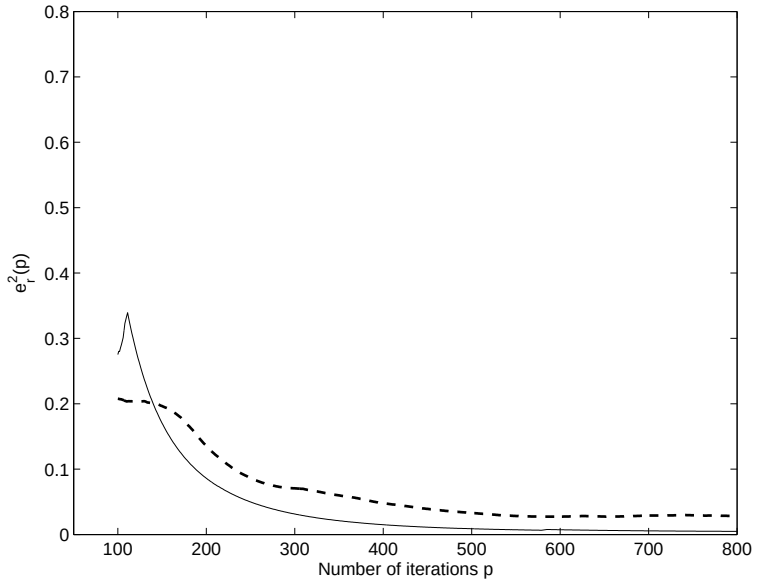

Fig. 7. MSE between the reference and estimated *a posteriori* change-point probabilities versus p (solid line). Averaged MSE computed from 64 chains (dotted line) ($N_{bi} = 200$).

V. Alternative models

This section presents possible modifications to the model and Gibbs sampler presented above.

A. Signals with different dynamics

We have assumed in the previous model that the hyperparameter γ is the same for all observed signals. This section shows that a similar analysis can be conducted with a set of different hyperparameters $\gamma_j, j = 1, \ldots, J$. Such analysis is interesting when signal amplitudes differ significantly from one signal to another. In this case, conjugate Inverse-Gamma distributions are selected for the noise variances as follows:

$$\sigma_{j,k}^2 \mid \left(\frac{\nu}{2}, \frac{\gamma_j}{2}\right) \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma_j}{2}\right). \tag{28}$$

The priors for $\gamma_j, j = 1, \ldots, J$ are still vague conjugate Inverse-Gamma distributions which reflect the lack of precise knowledge regarding these hyperparameters:

$$f(\gamma_j) = \frac{1}{\gamma_j} \mathbb{I}_{\mathbb{R}^+}(\gamma_j). \tag{29}$$

By assuming independence between γ_j and $\gamma_{j'}$ for all $j \neq j'$, the joint prior distribution of $\boldsymbol{\gamma} = \left[\gamma_1, \ldots, \gamma_J\right]^\mathsf{T}$ is:

$$f(\boldsymbol{\gamma}) = \prod_{j=1}^{J} \frac{1}{\gamma_j} \mathbb{I}_{\mathbb{R}^+}(\gamma_j). \tag{30}$$

Consequently, the posterior of interest is given by:

$$f(\mathbf{R}, \boldsymbol{\gamma}, \delta_0^2 | \mathbf{Y}) \propto \left(\delta_0^2\right)^{-\frac{p}{2}\sum_{j=1}^{J} K_j(\mathbf{r}_j)} f(\delta_0^2|\xi,\beta) C(\mathbf{R}|\mathbf{Y}, \boldsymbol{\alpha}) \times \prod_{j=1}^{J} \left[\frac{1}{\gamma_j} \mathbb{I}_{\mathbb{R}^+}(\gamma_j) \prod_{k=1}^{K_j} \left(\frac{\gamma_j^{\frac{\nu}{2}} \left|\mathbf{M}_{j,k}\right|^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)\right)}{\left(\gamma_j + T_{j,k}^2\right)^{\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)}} \right) \right],$$

where $T_{j,k}^2, \mathbf{Q}_{j,k}, \mathbf{M}_{j,k}$ and $C(\mathbf{R}|\mathbf{Y})$ have been defined in Eq.'s (15) and (16) respectively.

The generation of $f(\boldsymbol{\gamma}, \boldsymbol{\sigma}^2 | \mathbf{R}, \delta_0^2, \mathbf{Y})$ in Algorithm 1 is achieved by replacing the corresponding steps by the following two:

$$\begin{aligned} \sigma_{j,k}^2 | \boldsymbol{\gamma}, \delta_0^2, \mathbf{R}, \mathbf{Y} &\sim \mathcal{IG}\left(\frac{\nu + n_{j,k}(\mathbf{r}_j)}{2}, \frac{\gamma_j + T_{j,k}^2}{2} \right), \\ \gamma_j | \boldsymbol{\sigma}^2, \mathbf{R}, \mathbf{Y} &\sim \mathcal{G}\left(\frac{\nu}{2} K_j(\mathbf{r}_j), \frac{1}{2} \sum_{k=1}^{K_j(\mathbf{r}_j)} \frac{1}{\sigma_{j,k}^2} \right). \end{aligned} \tag{31}$$

We illustrate the performance of the method by processing the synthetic data presented in section IV. However the amplitude of the second signal have been multiplicated by 0.005 to produce a significant amplitude change from one signal to another. Fig. 8 shows the posterior distributions of the change-locations obtained for the two time-series with the two approaches. It clearly appears that the initial algorithm with one hyperparameter γ (left figure) is unable to detect the first change-point that occurs at time $i = 60$ in the first sequence. The algorithm with different hyperparameters γ_j, $j = 1, \ldots, J$ (right figure) allows to detect all changes accurately.

B. Markov model for $\mathbf{R}$

The vectors $\mathbf{R}_i = \left[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}\right]^\mathsf{T}$ ($i = 1, \ldots, n$) have been previously assumed to be independent. As a consequence, the algorithm can hesitate between close change-point values. Introducing constraints on the length of the segment could be an efficient means to limit this phenomenon. This section presents a standard discrete-time finite state Markov model on $\mathbf{R}$ that rejects all segmentation models involving segments shorter than a minimal length L. For the

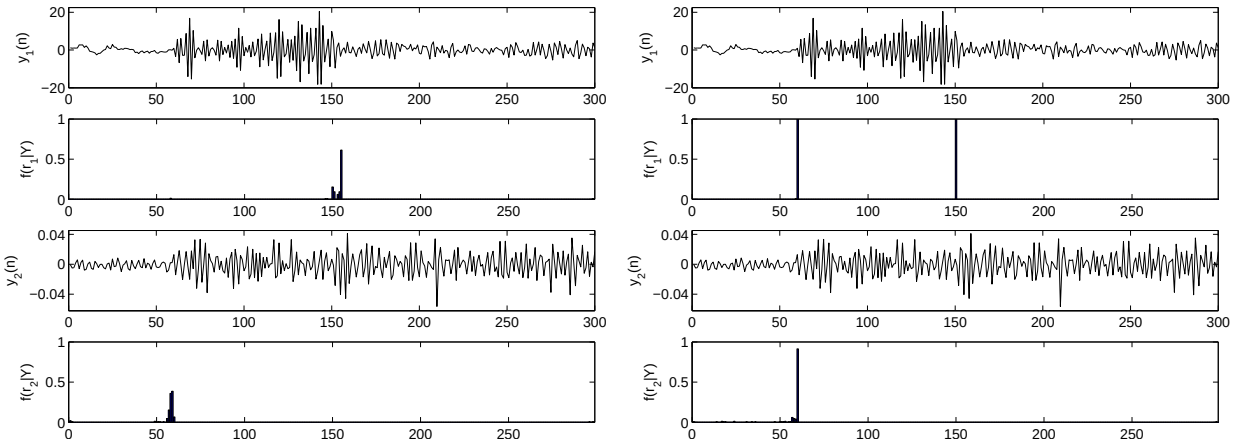

Fig. 8. Posterior distributions of the change-point locations for one hyperparameter γ (left) and several hyperparameters γ_j, $j = 1, \ldots, J$ (right).

sake of clarity, we have considered the simple case $L = 1$. However this procedure could be generalized to any value of L.

We propose a 2^J-state Markovian model on $\mathbf{R}$ with the following transition matrix:

$$\mathbf{P} = \begin{pmatrix} \mathrm{P}_{0\ldots0} & \mathrm{P}_{0\ldots1} & \cdots & \mathrm{P}_{1\ldots1} \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}. \tag{32}$$

By denoting $\Pi(\mathbf{R}_i) = \prod_{\boldsymbol{\epsilon}\in\mathcal{E}} \mathrm{P}_{\boldsymbol{\epsilon}}^{\delta(\mathbf{R}_i-\boldsymbol{\epsilon})}$ and $\Xi(\mathbf{R}_i) = \delta(\mathbf{R}_i - 0)$ (with $\delta(\mathbf{R}_i - \boldsymbol{\epsilon}) = 1$ if $\mathbf{R}_i = \boldsymbol{\epsilon}$, $\delta(\mathbf{R}_i - \boldsymbol{\epsilon}) = 0$ otherwise), it can be shown that:

$$f(\mathbf{R}_i|\mathbf{R}_{i-1}, \mathbf{P}) = \begin{cases} \Pi(\mathbf{R}_i) & \text{if } \Xi(\mathbf{R}_{i-1}) = 1, \\ \Xi(\mathbf{R}_i) & \text{otherwise.} \end{cases} \tag{33}$$

As a consequence, the prior for $\mathbf{R}$ can be written:

$$f(\mathbf{R}_i|\mathbf{R}_{i-1}, \mathbf{P}) = \prod_{i=2}^{n-1} \left[(\Pi(\mathbf{R}_i) - \Xi(\mathbf{R}_i))\, \Xi(\mathbf{R}_{i-1}) + \Xi(\mathbf{R}_i) \right] f(\mathbf{R}_i). \tag{34}$$

The resulting factor $C(\mathbf{R}|\mathbf{Y})$ in the posterior (14) is now defined as:

$$C(\mathbf{R}|\mathbf{Y}) = \frac{\prod_{\boldsymbol{\epsilon}\in\mathcal{E}} \Gamma\left(\widetilde{S}_{\boldsymbol{\epsilon}}(\mathbf{R}) + \alpha_{\boldsymbol{\epsilon}}\right)}{\Gamma\left(\sum_{\boldsymbol{\epsilon}\in\mathcal{E}} (\widetilde{S}_{\boldsymbol{\epsilon}}(\mathbf{R}) + \alpha_{\boldsymbol{\epsilon}})\right)}, \tag{35}$$

where $\widetilde{S}_{\boldsymbol{\epsilon}}(\mathbf{R})$ is the number of lags such as $\mathbf{R}_i = \boldsymbol{\epsilon}$ and $\mathbf{R}_{i-1} = \mathbf{0}$. The posterior distribution of $\mathrm{P}_{\boldsymbol{\epsilon}}$ is still a Dirichlet distribution:

$$\mathbf{P}|\mathbf{R}, \mathbf{Y} \sim \mathcal{D}_{2^J}\left(\widetilde{S}_{\boldsymbol{\epsilon}}(\mathbf{R}) + \alpha_{\boldsymbol{\epsilon}}\right). \tag{36}$$

We illustrate the performance of the method by processing the synthetic data presented in the section IV. However, we have modified the second sequence by inserting a new change-point at lag $i = 69$. We perform the proposed algorithm with $L = 2p$, i.e. the models whose segments are smaller than twice the order of the AR processes are prohibited. Fig. 9 shows the posterior distributions of the change-locations obtained for the two time-series with the two approaches. It clearly appears that the initial algorithm detects successive model changes between $i = 60$ and $i = 69$ (left figure), contrary to the modified algorithm that includes a Markov model for $\mathbf{R}$ (right figure).

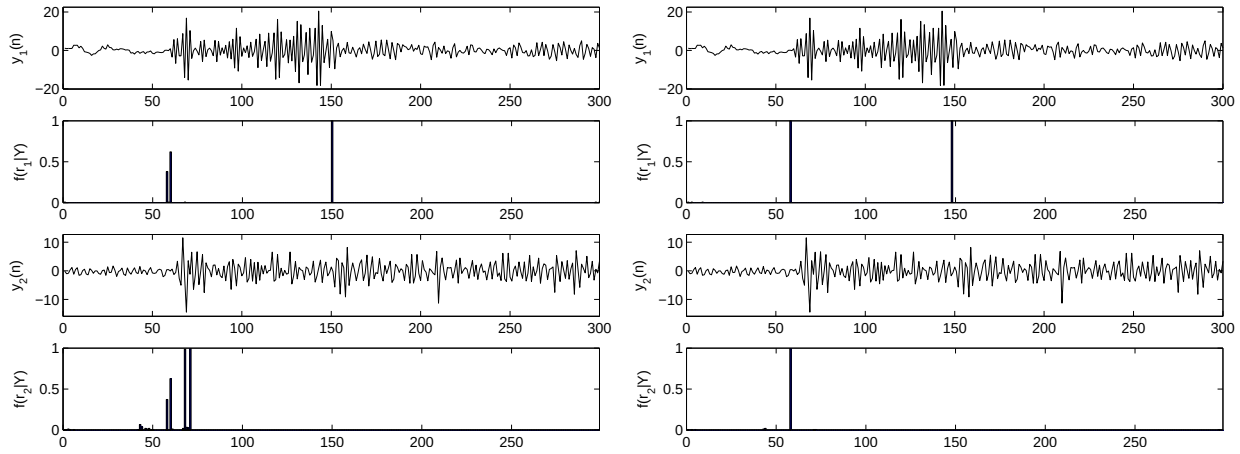

Fig. 9. Posterior distributions of the change-point locations for independent priors for $\mathbf{R}_i$ (left) and a markovian model for $\mathbf{R}$ (right).

C. Unknown AR model orders

This section generalizes the previous algorithm to AR processes whose orders differ from one segment to another.

1) Extented Bayesian model: We define appropriate priors for the new parameters to be estimated. A truncated Poisson distribution is chosen for the model order priors:

$$f(p_{j,k}|\psi) = \frac{1}{\Psi_{p_{\max}}(\psi)} \frac{\psi^{p_{j,k}}}{p_{j,k}!} \mathbb{I}_{\{0,\ldots,p_{\max}\}}(p_{j,k}), \qquad \Psi_{p_{\max}}(\psi) = \sum_{p=0}^{p_{\max}} \frac{\psi^p}{p!}. \tag{37}$$

Classically, a vague conjugate Gamma distribution is assigned to the hyperparameter ψ with fixed parameters μ and ρ:

$$\psi|\mu,\rho \sim \mathcal{G}\left(\rho,\mu\right), \tag{38}$$

where $\mathcal{G}\left(a,b\right)$ denotes the Gamma distribution with parameters a and b. Therefore, by assuming the independence between p_{j_1,k_1} and p_{j_2,k_2} for all $j_1 \neq j_2$ and $k_1 \neq k_2$, and by denoting $\mathbf{p} = \{\mathbf{p}_1,\ldots,\mathbf{p}_J\}$ with $\mathbf{p}_j = \left[p_{j,1},\ldots,p_{j,K_j}\right]^{\mathsf{T}}$, the posterior of interest can be written:

$$f(\mathbf{R},\mathbf{p},\gamma,\boldsymbol{\delta}^2,\psi|\mathbf{Y}) \propto \frac{1}{\gamma}\mathbb{I}_{\mathbb{R}^+}(\gamma)C(\mathbf{R}|\mathbf{Y},\boldsymbol{\alpha})f(\delta_0^2|\xi,\beta)\Psi_{p_{\max}}(\psi)^{-\sum_{j=1}^{J}K_j} \times \prod_{j=1}^{J}\prod_{k=1}^{K_j}\left(\frac{\gamma^{\frac{\nu}{2}}\left|\mathbf{M}_{j,k}\right|^{\frac{1}{2}}\Gamma\left(\frac{\nu}{2}+\frac{1}{2}n_{j,k}\right)}{\left(\gamma+T_{j,k}^2\right)^{\frac{\nu}{2}+\frac{1}{2}n_{j,k}}}\frac{\psi^{p_{j,k}}}{p_{j,k}!}\left(\delta_0^2\right)^{-\frac{p_{j,k}}{2}}\right),$$

where $C(\mathbf{R}|\mathbf{Y})$ has been defined in Eq. (16). We point out that the dimensions of the matrix $\mathbf{M}_{j,k}$ and therefore the quantity $T_{j,k}^2$ defined in Eq. (15) depend on the model order $p_{j,k}$.

2) Reversible jump MCMC algorithm: The previous distribution requires to develop an efficient strategy to sample according to $f(\mathbf{R},\mathbf{p},\gamma,\delta_0^2,\psi|\mathbf{Y})$. In this case, the vectors to be sampled $(\mathbf{R},\mathbf{p},\gamma,\delta_0^2,\psi)$ belong to the space $\{0,1\}^{nJ}\times\prod_{j=1}^{J}\{0,\ldots,p_{\max}\}^{K_j}\times\mathbb{R}^+\times\mathbb{R}$ whose dimension depends on K_j. In order to sample directly on this space, we propose an hybrid Gibbs algorithm whose main steps are detailed below:

- **a- Generation of samples according to $f(\mathbf{R}|\mathbf{p},\gamma,\delta_0^2,\psi,\mathbf{Y})$:**
  As in the initial model, this generation is achieved by using $n-1$ Gibbs moves to generate Monte Carlo samples distributed according to $f(\mathrm{r}_{1,i},\ldots,\mathrm{r}_{J,i}|\mathbf{p},\gamma,\delta_0^2,\psi,\mathbf{Y})$. The 2^J probabilities $P\left(\left[\mathrm{r}_{1,i},\ldots,\mathrm{r}_{J,i}\right]^{\mathsf{T}} = \boldsymbol{\epsilon}|\mathbf{R}_{-i},\mathbf{p},\gamma,\delta_0^2,\psi,\mathbf{Y}\right)$ could be evaluated in an exact way with the two following updating rules for $\mathbf{p}$:
  - if two segments with orders p_{j,k_1} and p_{j,k_2} have to be merged, the model order $p_{j,k}^*$ of the resulting segment is $p_{j,k}^* = p_{j,k_1} + p_{j,k_2}$.
  - if one segment with order $p_{j,k}$ has to be split, the model orders p_{j,k_1}^* and p_{j,k_2}^* of the two resulting segments are chosen as follows: $p_{j,k_1}^* \sim \mathcal{U}_{\{0,\ldots,p_{j,k}\}}$ and $p_{j,k_2}^* = p_{j,k} - p_{j,k_1}^*$.

  These choices ensure the reversibility of the different moves.
- **b- Generation of samples according to $f(\mathbf{p}|\mathbf{R},\gamma,\delta_0^2,\psi,\mathbf{Y})$:**
  As in [9], the update of the model orders is performed by using a reversible jump MCMC

Algorithm 2: Hybrid Gibbs algorithm for abrupt change detection with unknown model orders

- Initialization:
  - Sample hyperparameter vector $\widetilde{\mathbf{\Phi}}^{(0)} = \left(\widetilde{\delta}_0^{2(0)}, \widetilde{\gamma}^{(0)}, \widetilde{\mathbf{P}}^{(0)}\right)$ from the pdf in Eq. (11),
  - Sample hyperparameter $\widetilde{\psi}^{(0)}$ from the pdf in Eq. (38),
  - For $i = 1, \ldots, n-1$ sample, $\left[\widetilde{\mathrm{r}}_{1,i}^{(0)}, \ldots, \widetilde{\mathrm{r}}_{J,i}^{(0)}\right]$ from the pdf in Eq. (6),
  - For $j = 1, \ldots, J$, $k = 1, \ldots, K$, sample $\widetilde{\sigma}_{j,k}^{2(0)}$, $\widetilde{\mathbf{a}}_{j,k}^{(0)}$ and $\widetilde{p}_{j,k}^{(0)}$ from the pdf's in Eq.'s (8), (9) and (37),
  - Set $t \leftarrow 1$,
- Iterations: for $t = 1, 2, 3, \ldots,$ do
  - For $i = 1, \ldots, n-1$, sample $\left[\widetilde{\mathrm{r}}_{1,i}^{(t)}, \ldots, \widetilde{\mathrm{r}}_{J,i}^{(t)}\right]$ according to the 2^J probabilities defined in step a- below,
  - For $j = 1, \ldots, J$, $k = 1, \ldots, K$, update the model order $\widetilde{p}_{j,k}^{(t)}$ (see step b-):
    - if $\left(u \sim \mathcal{U}_{[0,1]}\right) \leq b_{\widetilde{p}_{j,k}^{(t-1)}}$, then propose $p_{j,k}^* = \widetilde{p}_{j,k}^{(t-1)} + 1$,
      else if $\left(u \sim \mathcal{U}_{[0,1]}\right) \leq b_{\widetilde{p}_{j,k}^{(t-1)}} + d_{\widetilde{p}_{j,k}^{(t-1)}}$, then propose $p_{j,k}^* = \widetilde{p}_{j,k}^{(t-1)} - 1$,
    - if $\left(v_p \sim \mathcal{U}_{[0,1]}\right) \leq \lambda_{p_{j,k} \to p_{j,k}^*}$ (see Eq. (39)), $\widetilde{p}_{j,k}^{(t)} = p_{j,k}^*$,
      else $\widetilde{p}_{j,k}^{(t)} = \widetilde{p}_{j,k}^{(t-1)}$,
  - Update $\widetilde{\psi}^{(t)}$ (see step c-):
    - Propose ψ^* according to the Gamma proposal distribution defined in step d-,
    - if $\left(v_\psi \sim \mathcal{U}_{[0,1]}\right) \leq \lambda_{\psi \to \psi^*}$ (see Eq. (40)), $\widetilde{\psi}^{(t)} = \psi^*$,
      else $\widetilde{\psi}^{(t)} = \widetilde{\psi}^{(t-1)}$,
  - For $j = 1, \ldots, J$, $k = 1, \ldots, K$, sample $\widetilde{\sigma}_{j,k}^{2(t)}$ from the pdf in Eq. (41),
  - Sample $\widetilde{\gamma}^{(t)}$ from the pdf in Eq. (42),
  - For $j = 1, \ldots, J$, $k = 1, \ldots, K$, sample $\widetilde{\mathbf{a}}_{j,k}^{(t)}$ from the pdf in Eq. (43),
  - Sample $\widetilde{\delta}_0^{2(t)}$ from the pdf in Eq. (44),
  - *Optional step:* sample $\widetilde{\mathbf{P}}^{(t)}$ from the pdf in Eq. (45),
  - Set $t \leftarrow t + 1$.

procedure:

- a birth move $p_{j,k}^* = p_{j,k} + 1$ is proposed with the probability $b_{p_{j,k}}$,
- a death move $p_{j,k}^* = p_{j,k} - 1$ is proposed with the probability $d_{p_{j,k}}$.

The acceptance probability for the new Monte Carlo state is:

$$\lambda_{p_{j,k}\to p^*_{j,k}} = \frac{\lambda^{p^*_{j,k}}}{\lambda^{p_{j,k}}}\frac{p^*_{j,k}!}{p_{j,k}!}\left(\frac{1}{\delta_0^2}\right)^{\pm\frac{1}{2}}\frac{\left|\mathbf{M}_{j,k}(p^*_{j,k})\right|^{\frac{1}{2}}}{\left|\mathbf{M}_{j,k}(p_{j,k})\right|^{\frac{1}{2}}}\frac{\left[T^2_{j,k}(p^*_{j,k})+\gamma\right]^{\frac{\nu}{2}+\frac{1}{2}n_{j,k}}}{\left[T^2_{j,k}(p_{j,k})+\gamma\right]^{\frac{\nu}{2}+\frac{1}{2}n_{j,k}}}\left(\frac{d_{p_{j,k}}}{b_{p_{j,k}}}\right)^{\pm 1}, \quad (39)$$

with $p^*_{j,k} = p_{j,k} \pm 1$,

- **c- Generation of samples according to $f(\psi|\mathbf{R},\mathbf{p},\gamma,\delta_0^2,\mathbf{Y})$:**

  Looking carefully at its posterior distribution, we can sample ψ by a simple Metropolis-Hastings step with a Gamma proposal distribution $\psi^* \sim \mathcal{G}\left(\mu+\sum_{j,k}p_{j,k}, \rho+\sum_j K_j\right)$ and the following acceptance probability:

$$\lambda_{\psi\to\psi^*} = \left[\frac{\Psi_{p_{\max}}(\psi)}{\Psi_{p_{\max}}(\psi^*)}\exp\left(\psi^*-\psi\right)\right]^{\sum_{j=1}^J K_j}, \quad (40)$$

- **d- Generation of samples according to $f(\gamma,\boldsymbol{\sigma}^2|\mathbf{R},\mathbf{p},\delta_0^2,\mathbf{Y})$:**

  As in the initial model, after appropriate integration, the following posteriors are obtained:

$$\sigma^2_{j,k}|\mathbf{R},\mathbf{p},\gamma,\delta_0^2,\mathbf{Y} \sim \mathcal{IG}\left(\frac{\nu+n_{j,k}}{2}, \frac{\gamma+T^2_{j,k}}{2}\right), \quad (41)$$

$$\gamma|\mathbf{R},\boldsymbol{\sigma}^2 \sim \mathcal{G}\left(\frac{\nu}{2}\sum_j K_j, \frac{1}{2}\sum_{j,k}\frac{1}{\sigma^2_{j,k}}\right), \quad (42)$$

- **e- Generation of samples according to $f(\delta_0^2,\mathbf{A}|\mathbf{R},\mathbf{p},\boldsymbol{\sigma}^2,\mathbf{Y})$:**

  This is achieved as follows:

$$\mathbf{a}_{j,k}|\mathbf{R},\mathbf{p},\boldsymbol{\sigma}^2,\delta_0^2,\mathbf{Y} \sim \mathcal{N}\left(\boldsymbol{\mu}_{j,k}, \sigma^2_{j,k}\mathbf{M}_{j,k}\right), \quad (43)$$

$$\delta_0^2|\mathbf{R},\mathbf{p},\mathbf{A},\boldsymbol{\sigma}^2 \sim \mathcal{IG}\left(\xi+\sum_{j,k}\frac{p_{j,k}}{2}, \beta+\sum_{j,k}\frac{\|\mathbf{a}_{j,k}\|^2}{2\sigma^2_{j,k}}\right). \quad (44)$$

- **e- Generation of samples according to $f(\mathbf{P}|\mathbf{R},\mathbf{Y})$:**

  As in the initial model, the following posterior is obtained:

$$\mathbf{P}|\mathbf{R},\mathbf{Y} \sim \mathcal{D}_{2^J}(S_\epsilon(\mathbf{R})+\alpha_{\boldsymbol{\epsilon}}). \quad (45)$$

It is important to note that the preceding scheme requires only one model order selection (i.e. one reversible jump MCMC procedure) contrary to the approach presented in [9].

3) Simulations: In order to assess the accuracy of the proposed method, we consider $J = 2$ synthetic signals of $n = 300$ samples. The change-point locations are $\boldsymbol{l}_1 = (60, 150)$ for signal $\#1$ and $\boldsymbol{l}_2 = (60)$ for signal $\#2$. The parameters of the two AR processes (which have been extracted from [9]) are summarized in Table III. The fixed parameters and hyperparameters have been chosen as follows: $\nu = 2$, $\xi = 1$, $\mu = 1$, $\beta = 10^2$, $\rho = 10^{-2}$ (vague hyperpriors) and $\alpha_{\boldsymbol{\epsilon}} = 1, \forall \boldsymbol{\epsilon} \in \mathcal{E}$ so as to obtain a uniform prior distribution for $\mathbf{P}$. The estimated values for AR model orders associated to the two signals are depicted on Fig.'s 10 and 11. The corresponding change-point posterior distributions are shown on 12. The proposed algorithm achieves accurate estimation of changes in the two sequences. The orders of the AR processes in each segment are also estimated with good accuracy.

TABLE III

PARAMETERS OF THE AR MODEL AND NOISE VARIANCES FOR EACH SEGMENT OF EACH SEQUENCE.

Sequence	Segment	$\sigma^2_{j,k}$	$p_{j,k}$	$a_{j,k,l}$			
$j = 1$	$k = 1$	1.7	2	-0.8000	0.5200	–	–
	$k = 2$	1.6	4	2.3000	2.6675	1.8437	0.5936
	$k = 3$	1.8	3	0.5000	-0.6100	-0.5850	–
$j = 2$	$k = 1$	0.5	3	-2.0000	1.6350	-0.5075	–
	$k = 2$	0.6	2	1.7000	0.7450	–	–

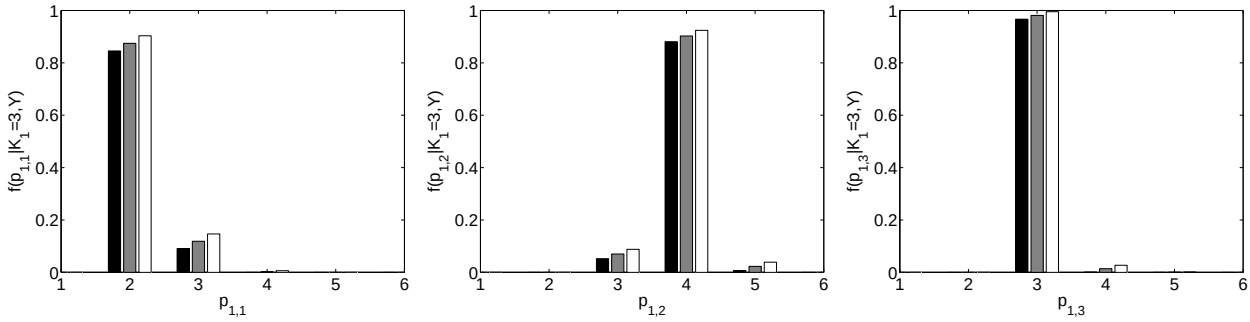

Fig. 10. Posterior distributions of the model orders p_{1i} (for $i = 1, \ldots, 3$) conditioned to $K_1 = 3$.

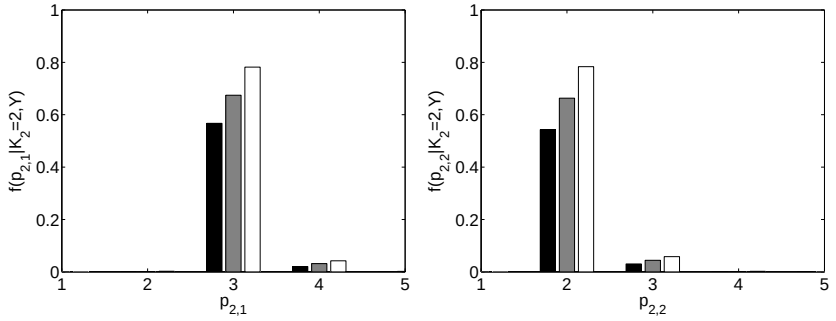

Fig. 11. Posterior distributions of the model orders p_{2i} (for $i = 1, 2$) conditioned to $K_2 = 2$.

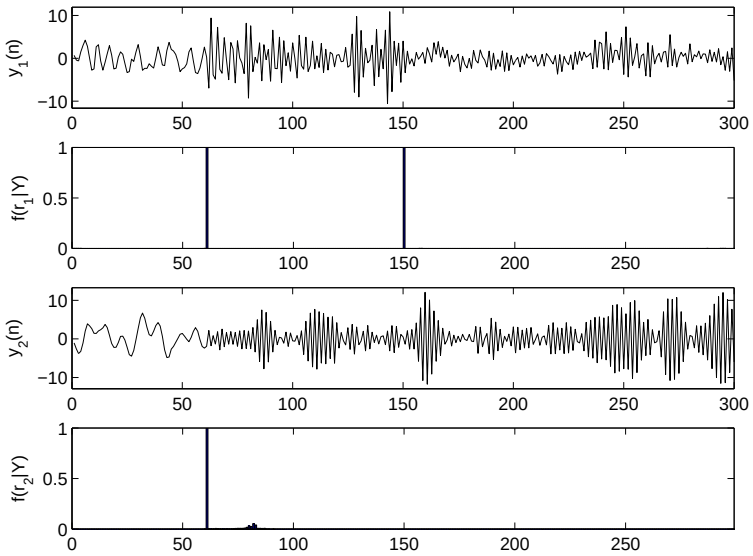

Fig. 12. Posterior distribution of the change-point locations estimated by the reversible jump algorithm.

VI. Applications

A. *"Arc-tracking" detection*

We illustrate the performance of the proposed segmentation procedure by processing real aeronautical data, where the issue is to prevent the phenomenon referred to as "arc-tracking". This phenomenon is responsible of many fatal aircraft crashes in the last years. The few hundreds of kilometers of wires embedded on military and commercial aircrafts are subject to various constraints (chemical, mechanical, thermic...) resulting in insulation damages. These breakdowns expose the cable to intermittent fault-arc currents that could ignite the neighboring wires [23]. Several methods for detection of wiring failures have been studied in the literature: they are

mainly based on dielectric properties [24] or, more recently, on electromagnetic properties [25]. We propose here an "arc-tracking" detection procedure that searches for transients in the predamaged wires, which is a early phenomenon announcing "arc-tracking" problems.

The analyzed data have been recorded from a common 3-phase (A, B and C) supply voltage whose electric network frequency is f_0. The phenomenon we are looking for affects the signals at frequencies higher than f_c[1]. Therefore, the $J = 3$ sequences whose sample size is $n = 551$ are filtered by an highpass filter in order to highlight the transients which are much less energetic. The filtered voltages can be accurately modeled as autoregressive processes. The presence of transients in the observed time series results in changes in the AR parameters.

We propose to detect the transients that appear on phases A, B and C between $t_1 = 0.04$s and $t_2 = 0.17$s. The observed data corresponding to the three phases have been processed by the proposed joint segmentation algorithm. The estimated number of change-points and their positions are obtained after $N_{MC} = 450$ iterations including a burn-in period of $N_{bi} = 100$ iterations. The parameters N_{bi} and N_{MC} have been chosen to provide appropriate potential reduction factors $\sqrt{\hat{\rho}}$ for the hyperparameters $\mathrm{P}_{\boldsymbol{\epsilon}}$. Running 30 iterations of the proposed algorithm takes approximately 4 minutes for a MATLAB implementation on a 2.8 Ghz Pentium IV.

Fig. 13 shows the MAP estimates of parameters K_j: $\widehat{K_1} = 4$, $\widehat{K_2} = 8$ and $\widehat{K_3} = 8$. The estimated posterior distribution of $\mathbf{R}$ depicted in Fig. 14 can then be used to estimate the beginning and the end of transients in each phase. Indeed, by keeping the $\widehat{K}_j$ largest peaks of the posterior distribution $f(\mathbf{R}|\mathbf{Y})$, the segments corresponding to the different transients (outlined by vertical lines on Fig.14) can be reconstructed.

B. Speech segmentation

This section illustrates the performance of the proposed algorithm by processing a real speech signal which has received much attention in the literature (see [1], [3], [9], [14] and more recently [22]). As explained in [1, p. 401], it belongs to a database designed by the French National Agency for Telecommunications. It consists of a noisy speech recorded in a car with the sampling frequency 8kHz and quantized with 16bits. It is prefiltered by a highpass filter with cutt-off frequency equal to 150Hz. The raw 1D data $\mathbf{y} = \left[y_1, \ldots, y_n\right]$ have been processed by

[1]For confidentiality reasons, the actual values of f_0 and f_c corresponding to these real aeronautical data cannot be provided.

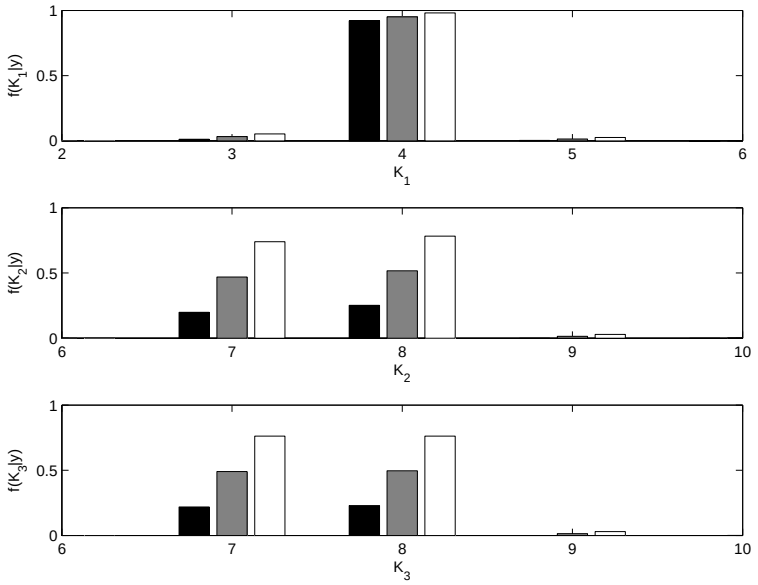

Fig. 13. Posterior distribution of the change-point number (3D aeronautical data).

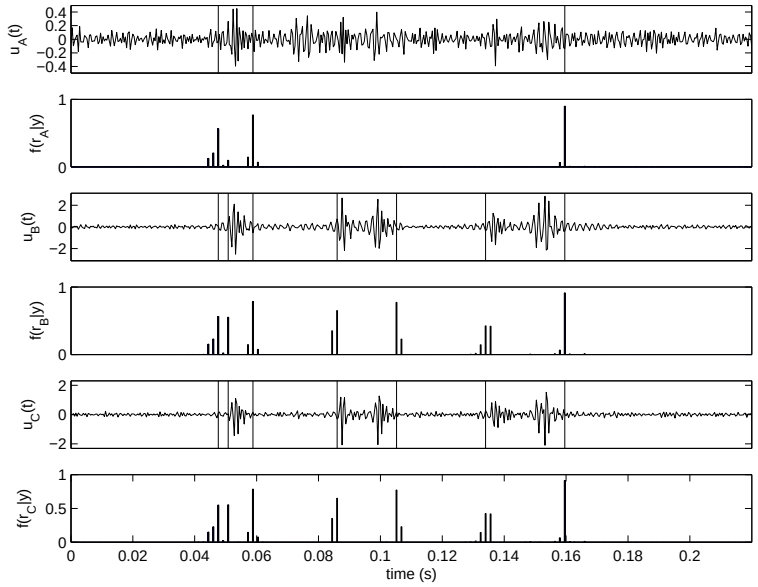

Fig. 14. Posterior distribution of the change-point locations and segmentation of 3D aeronautical data.

the proposed algorithm with $J = 1$. The estimated number of change-points and their positions are obtained after $N_{MC} = 600$ iterations including a burn-in period of $N_{bi} = 200$ iterations (N_{MC} and N_{bi} have been chosen in order to obtain appropriate potential reduction factors $\sqrt{\hat{\rho}}$

for the hyperparameters $\mathrm{P}_{\boldsymbol{\epsilon}}$). Running 1 iteration of the proposed algorithm for joint segmentation takes approximately 30 seconds for a MATLAB implementation on a 2.8Ghz Pentium IV. The estimated changes are depicted in Fig. 15 (top figure). Table IV also compares our results with those obtained with several other methods previously studied in the literature. It clearly appears that our method gives similar segmentation models. However, it has the advantage to be able to handle signals coming from different sensors. To illustrate this point, the data have been converted into stereo measurements $\mathbf{Y} = \left[\mathbf{y}_1, \mathbf{y}_2\right]^{\mathsf{T}}$ with $\mathbf{y}_1 = [y_{1,1}, \ldots, y_{1,n}]$ and $\mathbf{y}_2 = [y_{2,1}, \ldots, y_{2,n}]$ by using a standard mono-stereo converter. The change-point posterior distributions for the two signals $\mathbf{y}_1$ and $\mathbf{y}_2$ have been computed with the proposed algorithm with $J = 2$. The segments for the two time-series can be obtained by keeping the largest values of the change-point posteriors (corresponding to the estimated change-point numbers $\widehat{K}_j$, $j = 1, 2$). The results are presented in Table IV and in Fig. 15 (middle and bottom plots). They are in good agreement with the 1D segmentation. Note however that the segmentation of stereo signals does not estimate the first change $l_{1,1} = l_{2,1} = 448$ since it is not significant in both sequences.

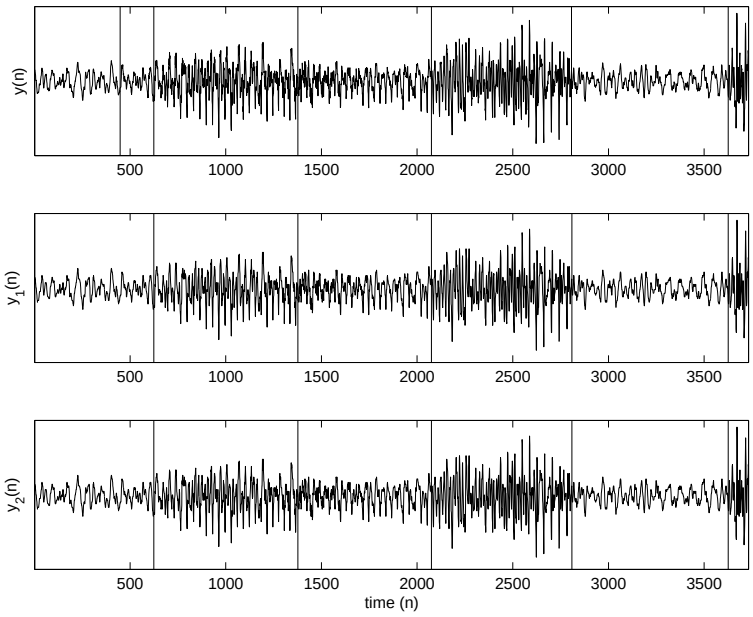

Fig. 15. Segmentations of 1D (top) and 2D (middle and bottom) real speech data.

TABLE IV

CHANGE-POINT POSITIONS ESTIMATED BY DIFFERENT METHODS.

Method	AR order	Estimated change-points									
Divergence [26]	16	445	–	645	1550	1800	2151	2797	–	3626	–
GLR [27]	16	445	–	645	1550	1800	2151	2797	–	3626	–
GLR [27]	2	445	–	645	1550	1750	2151	2797	3400	3626	–
Approx. ML [3]	16	445	–	626	1609	–	2151	2797	–	3627	–
1D MCMC [9]	estimated	448	–	624	1377	–	2075	2807	–	3626	–
Conditional MAP [22]	estimated	449	585	620	1365	1795	2144	2811	–	3624	3657
Proposed 1D approach	estimated	448	–	624	1377	–	2075	2807	–	3626	–
Proposed 2D approach Ch. 1	estimated	–	–	624	1377	–	2075	2809	–	3626	–
Proposed 2D approach Ch. 2	estimated	–	–	624	1377	–	2075	2809	–	3626	–

VII. CONCLUSIONS

This paper studied a joint Bayesian segmentation procedure allowing to segment signals recorded from different sensors. The proposed approach assumed that the signals can be modeled by piecewise constant autoregressive processes. A hierarchical Bayesian model was defined allowing to estimate jointly the change-point locations, the AR parameters and the noise variances for the multiple observed signals. To circumvent the complexity of the unknown parameters distributions, an appropriate Gibbs sampler was proposed to simulate samples distributed according to the posteriors of interest. Different extensions of the algorithm were discussed allowing to handle signals with different unknown AR orders or different dynamics. Two applications were finally investigated: arc-tracking detection and stereo speech signal segmentation. Note that our assumption regarding the observed signals are sufficiently mild to handle a large class of other real signals such as seismic [1] or biomedical [13] signals. Extending this work to more general models including models for long range dependent data [4] or generalized autoregressive conditional heteroskedastic (GARCH) models [28] would also be an interesting issue.

ACKNOWLEDGEMENTS

The authors would like to thank F. Gustafsson for providing the speech data and the reviewers for their thoughtful and incisive comments about this paper.

APPENDIX
FAST COMPUTATIONS

It is interesting to notice that the matrices $\mathbf{Q}_{j,k}$ and $\mathbf{M}_{j,k}$, and the variable $T^2_{j,k}$ defined in Eq. (15) could be computed following the implementations described in [16]. We note $\mathbf{y}_{j,[k]} = \mathbf{y}^{\mathsf{T}}_{j,l_{j,k-1}+1:l_{j,k}}$.

Algorithm 3: Fast computations of $T^2_{j,k}$

- Compute $\mathbf{M}^{-1}_{j,k} = \mathbf{Y}^{\mathsf{T}}_{j,k}\mathbf{Y}_{j,k} + \frac{\mathbf{I}_p}{\delta^2_0}$,
- Compute Cholesky's factors $\mathbf{C}_{j,k}$ such as $\mathbf{C}_{j,k}\mathbf{C}^{\mathsf{T}}_{j,k} = \mathbf{M}_{j,k}$,
- Compute $\mathbf{u}_{j,k} = \mathbf{Y}^{\mathsf{T}}_{j,k}\mathbf{y}_{j,[k]}$,
- Solve the system $\mathbf{C}_{j,k}\mathbf{v}_{j,k} = \mathbf{u}_{j,k}$ for $\mathbf{v}_{j,k}$,
- Compute $T^2_{j,k} = \mathbf{y}^{\mathsf{T}}_{j,[k]}\mathbf{y}_{j,[k]} - \mathbf{v}^{\mathsf{T}}_{j,k}\mathbf{v}_{j,k}$.

Such implementations allow us to develop a strategy to sample $\mathbf{a}_{j,k}|\delta^2_0, \sigma^2_{j,k}, \mathbf{R}, \mathbf{Y}$ according to $\mathcal{N}(\boldsymbol{\mu}_{j,k}, \sigma^2_{j,k}\mathbf{M}_{j,k})$ in the effective following scheme.

Algorithm 4: Fast multivariate Gaussian sampling of $\mathbf{a}_{j,k}$

- Sample a i.i.d. vector $\mathbf{w}_{j,k}$ according to $\mathcal{N}(\mathbf{0}, \sigma^2_{j,k}\mathbf{I}_p)$
- Solve the system $\mathbf{C}_{j,k}\widetilde{\boldsymbol{\mu}}_{j,k} = \mathbf{w}_{j,k}$ for $\widetilde{\boldsymbol{\mu}}_{j,k}$,
- Solve the system $\mathbf{C}^{\mathsf{T}}_{j,k}\boldsymbol{\mu}_{j,k} = \boldsymbol{\mu}'_{j,k}$ for $\boldsymbol{\mu}_{j,k}$,
- Compute $\mathbf{a}_{j,k} = \widetilde{\boldsymbol{\mu}}_{j,k} + \boldsymbol{\mu}_{j,k}$.

Another advantage of this scheme is that it is not necessary to compute directly the determinant of the matrices $\mathbf{M}_{j,k}$ that appear in Eq.(14). We can compute $\left|\mathbf{M}_{j,k}\right| = \left|\mathbf{C}_{j,k}\right|^{-2}$ where $\mathbf{C}_{j,k}$ are upper triangular matrices and, consequently, whose determinants are easy to perform.

REFERENCES

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs NJ: Prentice-Hall, 1993.

[2] B. Brodsky and B. Darkhovsky, *Nonparametric methods in change-point problems*. Boston (MA): Kluwer Academic Publishers, 1993.

[3] F. Gustafsson, *Adaptive filtering and change detection*. New-York: Wiley, 2000.

[4] M. Lavielle and E. Moulines, "Least squares estimation of an unknown number of shifts in a time series," *Jour. of Time Series Anal.*, vol. 21, no. 1, pp. 33–59, Jan. 2000.

[5] L. Birgé and P. Massart, "Gaussian model selection," *Jour. Eur. Math. Soc.*, vol. 3, pp. 203–268, 2001.

[6] E. Lebarbier, "Detecting multiple change-points in the mean of gaussian process by model selection," *Signal Processing*, vol. 85, no. 4, pp. 717–736, April 2005.

[7] R. E. McCulloch and R. S. Tsay, "Bayesian inference and prediction for mean and variance shifts in autoregressive time series," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 968–978, 1993.

[8] P. M. Djurić, "A MAP solution to off-line segmentation of signals," in *Proc. IEEE ICASSP-94*, vol. 4, 1994, pp. 505–508.

[9] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 747–758, March 2002.

[10] E. Kuhn and M. Lavielle, "Coupling a stochastic approximation version of EM with an MCMC procedure," *ESAIM P & S*, vol. 8, pp. 115–131, 2004.

[11] M. Lavielle and E. Lebarbier, "An application of MCMC methods for the multiple change-points problem," *Signal Processing*, vol. 81, no. 1, pp. 39–53, Jan. 2001.

[12] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York: Springer Verlag, 1990.

[13] M. Lavielle, "Optimal segmentation of random processes," *IEEE Trans. Signal Processing*, vol. 46, no. 5, pp. 1365–1373, May 1998.

[14] R. André-Obrecht, "A new statistical approach for automatic segmentation of continuous speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 29–40, Jan. 1988.

[15] S. M. Kay, *Modern spectral estimation*. Prentice Hall, 1988.

[16] M. Davy and J. Idier, "Fast MCMC computations for the estimation of sparse processes from noisy observations," in *Proc. ICASSP-04*, vol. 2, Montréal, Quebec, Canada, May 2004, pp. 1041–1044.

[17] C. P. Robert and S. Richardson, "Markov Chain Monte Carlo methods," in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed. New York: Springer Verlag, 1998, pp. 1–25.

[18] A. Gelman and D. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, pp. 457–511, 1992.

[19] S. Godsill and P. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech, Audio Proc.*, vol. 6, no. 4, pp. 352–372, 1998.

[20] P. M. Djurić and J.-H. Chun, "An MCMC sampling approach to estimation of nonstationary hidden markov models," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1113–1123, 2002.

[21] A. Gelman, J. B. Carlin, H. P. Robert, and D. B. Rubin, *Bayesian Data Analysis*. London: Chapman & Hall, 1995.

[22] P. Fearnhead, "Exact Bayesian curve fitting and signal segmentation," *IEEE Trans. Signal Processing*, vol. 53, no. 6, pp. 2160–2166, June 2005.

[23] C. Furse and R. Haupt, "Down to the wire," *IEEE Spectrum*, pp. 35–39, Feb. 2001.

[24] J. Hanson and D. Koenig, "Fault arc effects in cable bundles for space applications in vacuum," *IEEE Trans. Dielectrics and Electrical Insulation*, vol. 4, no. 4, pp. 394–399, Aug. 1997.

[25] C. Lo and C. Furse, "Noise-domain reflectometry for locating wiring faults," *IEEE Trans. Electromagn. Compat.*, vol. 47, no. 1, pp. 97–104, Feb. 2005.

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

[27] U. Appel and A. V. Brandt, "Adaptive sequential segmentation of piecewise stationary time series," *Inform. Sci.*, vol. 29, no. 1, pp. 27–56, 1983.

[28] D. Guégan, *Séries chronologiques non linéaires à temps discret*. Paris: Economica, 1994.