Joint Segmentation of Piecewise Constant Autoregressive Processes by Using a Hierarchical Model and a Bayesian Sampling Approach

Nicolas Dobigeon, *Student Member, IEEE*, Jean-Yves Tourneret, *Member, IEEE*, and Manuel Davy

***Abstract*—We propose a joint segmentation algorithm for piecewise constant autoregressive (AR) processes recorded by several independent sensors. The algorithm is based on a hierarchical Bayesian model. Appropriate priors allow us 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 with parallel and independent individual signal segmentations. The initial algorithm is derived for piecewise constant AR processes whose orders are fixed on each segment. However, an extension to models with unknown model orders is also discussed. Theoretical results are illustrated by many simulations conducted with synthetic signals and real arc-tracking and speech signals.**

***Index Terms*—Gibbs sampling, hierarchical Bayesian analysis, Markov chain Monte Carlo (MCMC), reversible jumps, segmentation.**

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 oversegmentation. 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 that 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 noninformative 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 that 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 occurrence 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 autoregressive (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 $\mathbf{y}_j = [y_{j,1}, \ldots, y_{j,n}]$ for $j = 1, \ldots, J$, where $y_{j,i}$ is the

Manuscript received December 21, 2005; revised May 29, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Erchin Serpedin.

N. Dobigeon and J.-Y. Tourneret are with IRIT/ENSEEIHT/TéSA, 31071 Toulouse, Cedex 7, France (e-mail: Nicolas.Dobigeon@enseeiht.fr; Jean-Yves.Tourneret@enseeiht.fr).

M. Davy is with the INRIA-Futurs SequeL Team, 59651 Villeneuve d'Ascq Cedex, France (e-mail: Manuel.Davy@ec-lille.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2006.889090

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 that 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 $\mathbf{a}_{j,k} = [a_{j,k,1}, \ldots, a_{j,k,p}]^\top$. 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, $\mathbf{e}_j = [e_{j,1}, \ldots, e_{j,n}]$ is a vector of independent and identically distributed (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] and [14]. We assume in a first step that the orders of the AR models in (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 $\mathbf{x}_{j,i:i'} = [x_{j,i}, \ldots, x_{j,i'}]$, the equations in (1) can be written in the following matrix form:

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

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. The initial algorithm is generalized for AR models whose orders on each signal segment are unknown in Section V. 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 $\boldsymbol{\sigma}_j^2 = [\sigma_{j,1}^2, \ldots, \sigma_{j,K_j}^2]^\top$), 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 reparameterization 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 $\boldsymbol{\theta}_j = (\mathbf{r}_j, \boldsymbol{\sigma}_j^2, \mathbf{A}_j)$ and $\mathbf{r}_j = [\mathrm{r}_{j,1}, \ldots, \mathrm{r}_{j,n}]$. 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 $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_J]^\top$, 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

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

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

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 $i = 1, \ldots, n$). 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 $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_J]^\top$ 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 $[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}]^\top = \boldsymbol{\epsilon}$. We first assume that $\mathrm{P}_{\boldsymbol{\epsilon}}$ does not depend on the time index i. As a consequence, by assuming that $[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}]$ is independent of $[\mathrm{r}_{1,i'}, \ldots, \mathrm{r}_{J,i'}]$ 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 $[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}]^\top = \boldsymbol{\epsilon}$. 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 $[\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}]^\top = \boldsymbol{\epsilon}$ for all $i = 1, \ldots, n$. For instance, by choosing a high value of $\mathrm{P}_{0\ldots0}$ (respectively, $\mathrm{P}_{1\ldots1}$), we will favor a simultaneous absence (respectively, 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

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

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 nonequal hyperparameters γ_j, $j = 1, \ldots, J$. Such analysis is interesting when signal amplitudes differ significantly from one signal to another, and it is developed in [16].

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

$$\mathbf{a}_{j,k} | \sigma_{j,k}^2, \delta_0^2 \sim \mathcal{N}\left(\mathbf{0}_p, \sigma_{j,k}^2 \delta_0^2 \mathbf{I}_p\right) \tag{9}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix, $\mathbf{0}_p$ is the vector made of p zeros, and δ_0^2 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 $\boldsymbol{\Phi} = (\mathbf{P}, \delta_0^2, \gamma)$. 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_0^2 and γ: The priors for hyperparameters δ_0^2 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:

$$\delta_0^2 | \xi, \beta \sim \mathcal{IG}(\xi, \beta), \qquad f(\gamma) = \frac{1}{\gamma} \mathbb{I}_{\mathbb{R}^+}(\gamma) \tag{10}$$

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 $\boldsymbol{\alpha} = [\alpha_{0\ldots0}, \ldots, \alpha_{1\ldots1}]$ 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)

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

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 (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

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

with

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

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 (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 well as the hyperparameter γ, are the same for all the signals. These assumptions will be removed in Section V.

III. Gibbs Sampler for Joint Signal Segmentation

Gibbs sampling is an iterative sampling strategy which consists of generating random samples (denoted by $\widetilde{\cdot}^{(t)}$, 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 (14) by the three-step procedure outlined below. The main steps of Algorithm 1, as well as the key equations, are detailed in Sections III-A to III-C below.

Algorithm 1: Gibbs Sampling Algorithm for Abrupt Change Detection

- Initialization:
  - sample hyperparameter vector $\widetilde{\boldsymbol{\Phi}}^{(0)} = \left(\widetilde{\delta}_0^{2(0)}, \widetilde{\gamma}^{(0)}, \widetilde{\mathbf{P}}^0\right)$ from the probability density function (pdf) in (11);
  - for $i = 1, \ldots, n-1$ sample, $[\widetilde{\mathrm{r}}_{1,i}^{(0)}, \ldots, \widetilde{\mathrm{r}}_{J,i}^{(0)}]$ from the pdf in (6);
  - for $j = 1, \ldots, J$, $k = 1, \ldots, K_j$, sample $\widetilde{\sigma}_{j,k}^{2(0)}$ and $\widetilde{\mathbf{a}}_{j,k}^{(0)}$ from the pdf's in (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 $[\widetilde{\mathrm{r}}_{1,i}^{(t)}, \ldots, \widetilde{\mathrm{r}}_{J,i}^{(t)}]$ from its conditional distribution given in (17);
  - for signals $j = 1, \ldots, J$, and segments $k = 1, \ldots, K_j$, sample the noise variance $\widetilde{\sigma}_{j,k}^{2(t)}$ from its conditional posterior given in (18);
  - sample the hyperparameter $\widetilde{\gamma}^{(t)}$ from its posterior given in (19);
  - for signals $j = 1, \ldots, J$ and segments $k = 1, \ldots, K_j$, sample the AR coefficients $\mathbf{a}_{j,k}^{(t)}$ from their conditional posterior given in (20);
  - sample the hyperparameter $\widetilde{\delta}_0^{2(t)}$ from its conditional posterior given in (21);
  - (*optional step*) sample the hyperparameter $\widetilde{\mathbf{P}}^{(t)}$ from the pdf in (22);
  - set $t \leftarrow t + 1$.

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 $f(\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}|\gamma, \delta_0^2, \mathbf{Y})$. This vector is a random vector of Booleans in $\mathcal{E}$. Consequently, its distribution is fully characterized by the probabilities $P([\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}]^{\top} = \boldsymbol{\epsilon}|\gamma, \delta_0^2, \mathbf{Y})$, $\boldsymbol{\epsilon} \in \mathcal{E}$. 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:

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

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 $P([\mathrm{r}_{1,i}, \ldots, \mathrm{r}_{J,i}]^{\top} = \boldsymbol{\epsilon}|\mathbf{R}_{-i}, \gamma, \delta_0^2, \mathbf{Y})$ after appropriate normalization.

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 $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y})$, this step can be decomposed as follows.

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

- **Generate samples according to** $f(\gamma, \boldsymbol{\sigma}^2|\mathbf{R}, \delta_0^2, \mathbf{Y})$
  By integrating the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y})$ with respect to the AR parameters, the following results can be obtained:

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

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

  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:

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

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

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

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:

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

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 $\boldsymbol{l}_1 = (60, 150)$ and $\boldsymbol{l}_2 = (60)$. 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 $T^2_{j,k}$, $\mathbf{Q}_{j,k}$, and $\mathbf{M}_{j,k}$ defined in (15) have been computed following the implementations described in [17] 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 $N_{\mathrm{MC}} = 700$, including $N_{\mathrm{bi}} = 200$ 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 of signals with sample size $n = 300$ takes approximately 2 min and 30 s for a MATLAB implementation on a 2.8-GHz Pentium IV. Of course, the computational cost will increase for longer time series and may become prohibitive.

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 AR 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 the joint segmentation technique (right figures) whereas it is not detected when applying two single signal independent segmentations (left figures). 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 figures) are obtained with univariate segmentations ($J = 1$), which correspond to the Bayesian curve fitting strategy of Punskaya *et al.* [9].

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 $f(\mathbf{R}, \gamma, \delta_0^2|\mathbf{Y})$, 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 ± 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.

C. Noise Variances and AR Parameters

The estimation of the noise variances or AR parameters can be interesting in practical applications. Figs. 3 and 4 show the posterior distributions of parameters $\{\sigma^2_{1,k}\}_{k=1,\ldots,3}$ and $\{\sigma^2_{2,k}\}_{k=1,2}$ associated with the two time-series $\mathbf{y}_1$ and $\mathbf{y}_2$. These histograms are in good agreement with the actual values of the parameters $\sigma^2_{1,1} = 0.50$, $\sigma^2_{1,2} = 0.52$, $\sigma^2_{1,3} = 3.80$ and $\sigma^2_{21} = 0.81$, $\sigma^2_{22} = 4.63$. Similar results could be obtained for AR parameters. They are omitted here for brevity.

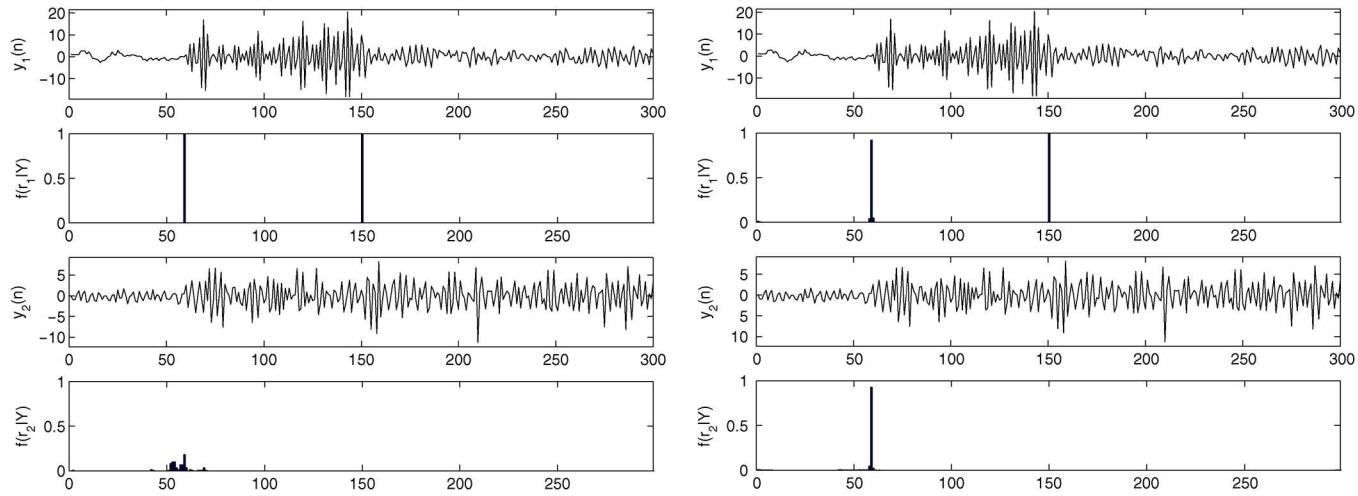


Fig. 1. Posterior distributions of the change-point locations for one-dimensional (left) and joint segmentations (right) obtained after $N_{\mathrm{bi}} = 200$ burn-in iterations and $N_r = 500$ iterations of interest.

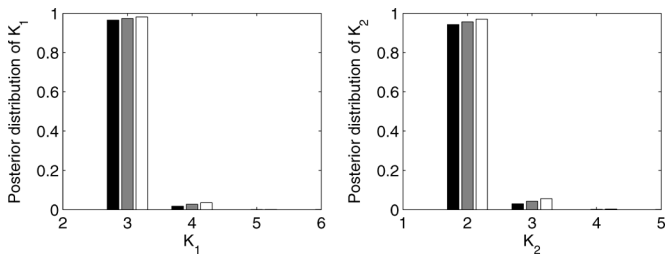


Fig. 2. Posterior distributions of the change-point numbers computed from $N_r = 500$ iterations of interest (mean in gray, mean $\pm$ standard deviations in white and black).

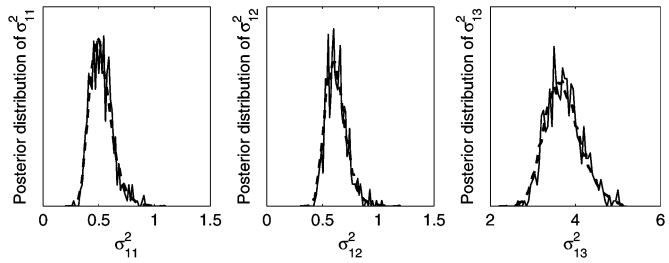


Fig. 3. Posterior distributions of the noise variances σ^2_{1i} (for $i = 1, \ldots, 3$) conditioned to $K_1 = 3$ computed from $N_r = 500$ iterations of interest (solid lines). Averaged posterior distributions from 64 Markov chains (dashed lines).

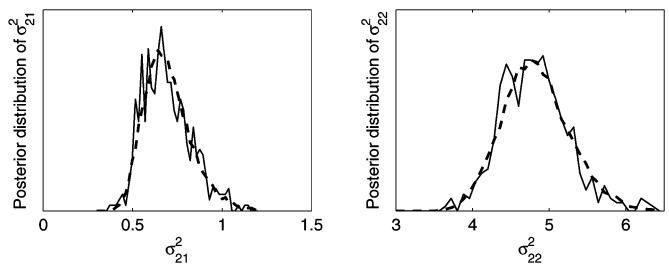


Fig. 4. Posterior distributions of the noise variances σ^2_{2i} (for i = 1,2) conditioned to $K_2 = 2$ computed from $N_r = 500$ iterations of interest (solid lines). Averaged posterior distributions from 64 Markov chains (dashed lines).

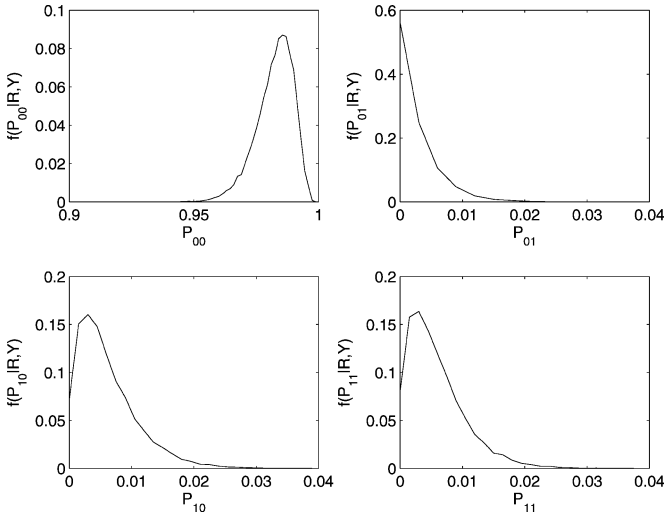


Fig. 5. Posterior distributions of the hyperparameters $\mathrm{P}_{\boldsymbol{\epsilon}}$ (computed from $N_r = 500$ iterations of interests of 64 Markov chains).

D. Hyperparameter Estimation

The performance of the hyperparameter estimation procedure needs to be investigated. The estimated posteriors of hyperparameters P_{00}, P_{01}, P_{10}, and P_{11} are depicted in Fig. 5. This shows that the proposed Gibbs sampler actually generates samples distributed according to the true distribution in (22).

E. Robustness to Correlated Noise Vectors

This section shows that the proposed joint segmentation procedure is robust to correlated noise vectors. For this, assume that the two noise vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are correlated with the following covariance matrices:

$$\Sigma_i = \begin{cases} \begin{pmatrix} \sigma^2_{1,1} & \rho \\ \rho & \sigma^2_{2,1} \end{pmatrix}, & \text{for } i = 1, \ldots, 60 \\ \begin{pmatrix} \sigma^2_{1,2} & \rho \\ \rho & \sigma^2_{2,2} \end{pmatrix}, & \text{for } i = 61, \ldots, 150 \\ \begin{pmatrix} \sigma^2_{1,3} & \rho \\ \rho & \sigma^2_{2,2} \end{pmatrix}, & \text{for } i = 151, \ldots, 300 \end{cases} .$$

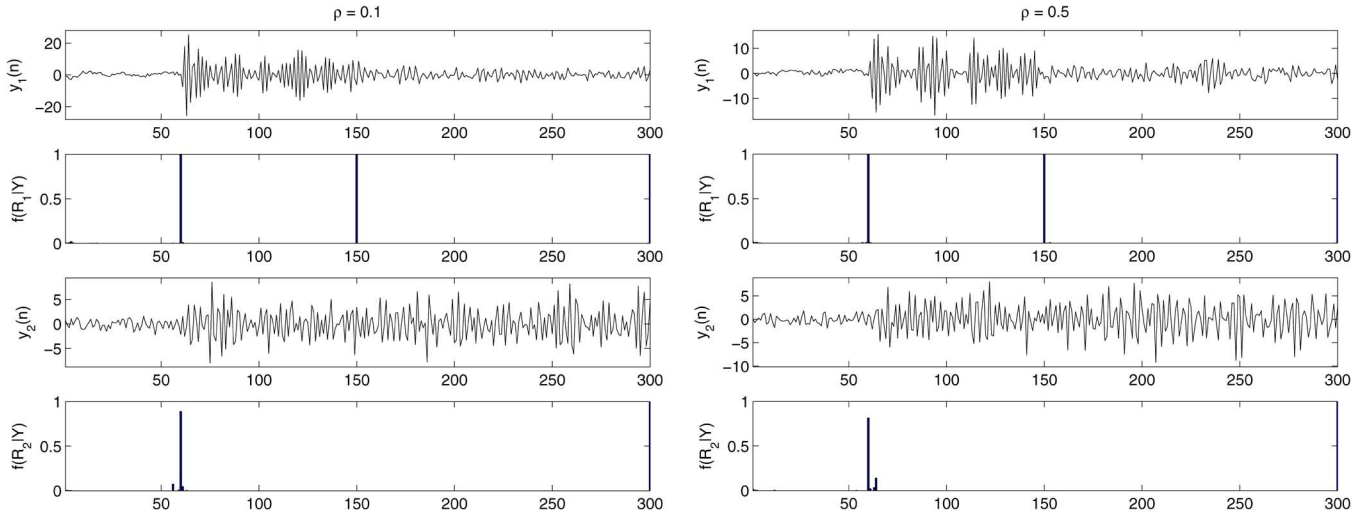


Fig. 6. Posterior distributions of the change-point locations for correlated noise vectors with $\rho = 0.1$ (left) and $\rho = 0.5$ (right) ($N_{\mathrm{bi}} = 200$ and $N_r = 500$).

Note that for $\rho = 0$, this example reduces to the previous synthetic data. The results of the joint hierarchical Bayesian segmentation procedure are depicted in Fig. 6 for two different values of ρ, i.e., two different correlations. Figs. 1 and 6 show that similar results are obtained for uncorrelated and correlated noise vectors. Consequently, the proposed model appears to be robust to noise correlations.

F. Sampler Convergence

The Gibbs sampler allows to draw samples $(\mathbf{R}^{(t)}, \gamma^{(t)}, \delta_0^{2(t)})$ 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 MMSE principle), as follows:

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

where N_{bi} is the number of burn-in iterations. However, two important questions are 1) When can we decide that the samples $\{\mathbf{R}^{(t)}\}$ are actually distributed according to the target distribution? and 2) How many samples are necessary to obtain an accurate estimate of $\mathbf{R}$ when using (23)? Running multiple chains with different initializations allows us to define various convergence measures for MCMC methods [18]. 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 [19] and has been often used to monitor convergence (see, for example, [20], [21], or [18, p. 33]). This criterion requires us 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

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

and

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

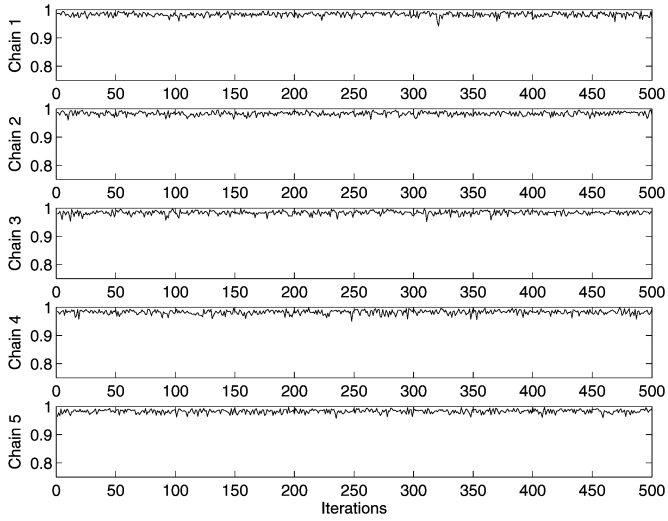


Fig. 7. Convergence assessment: the outputs of $M = 5$ chains for the parameter P_{00} converge to the same value.

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 [22, 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. 7. The chains clearly converge to similar values. The potential scale reduction factors for

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

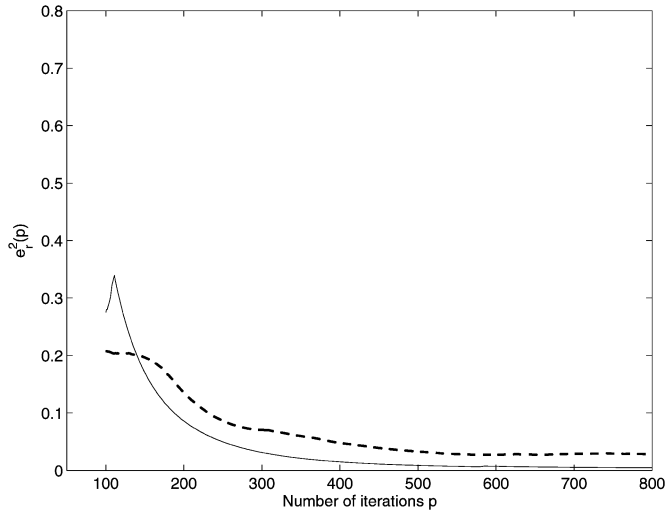


Fig. 8. MSE between the reference and estimated *a posteriori* change-point probabilities versus p (solid line). Averaged MSE computed from 64 chains (dashed line) ($N_{\mathrm{bi}} = 200$).

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 [22, 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 $\sigma^2_{j,k}$ 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 [23] for more details).

In order to determine the number of runs that are required to obtain an accurate estimate of $\mathbf{R}$ when using (23), an *ad hoc* approach consists of assessing convergence via appropriate graphical evaluations [18, p. 28]. Here, a reference estimate denoted as $\widetilde{\mathbf{R}}$ has been computed for a large number of iterations $N_r = 10000$ and $N_{\mathrm{bi}} = 200$ (to ensure convergence of the sampler and good accuracy of the approximation in (23)). Fig. 8 shows the mean-square error (MSE) between this reference estimate $\widetilde{\mathbf{R}}$ and the estimate obtained after $N_r = p$ iterations (and $N_{\mathrm{bi}} = 200$):

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

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

V. UNKNOWN AR MODEL ORDERS

This section generalizes the previous hierarchical Bayesian model to AR processes whose orders are unknown and differ from one segment to another.

A. 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

$$f(p_{j,k}|\psi) = \frac{1}{\Psi_{p_{\max}}(\psi)} \frac{\psi^{p_{j,k}}}{p_{j,k}!} \mathbb{I}_{\{0,\ldots,p_{\max}\}}(p_{j,k})$$
$$\Psi_{p_{\max}}(\psi) = \sum_{p=0}^{p_{\max}} \frac{\psi^p}{p!}. \tag{28}$$

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

$$\psi|\mu,\rho \sim \mathcal{G}(\rho,\mu).Al \tag{29}$$

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 $\mathbf{p}_j = [p_{j,1},\ldots,p_{j,K_j}]^\top$, the posterior of interest can be written

$$\begin{aligned} &f(\mathbf{R},\mathbf{p},\gamma,\boldsymbol{\delta}^2,\psi|\mathbf{Y}) \\ &\propto \frac{C(\mathbf{R}|\mathbf{Y},\boldsymbol{\alpha})}{\gamma} f\left(\delta_0^2|\xi,\beta\right) \Psi_{p_{\max}}(\psi)^{-\sum_j K_j} \\ &\times \prod_{j=1}^{J}\prod_{k=1}^{K_j} \left(\frac{\gamma^{\frac{\nu}{2}} |\mathbf{M}_{j,k}|^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2}+\frac{1}{2}n_{j,k}\right)}{\left(\gamma+T_{j,k}^2\right)^{\frac{\nu}{2}+\frac{1}{2}n_{j,k}}} \frac{\psi^{p_{j,k}}}{p_{j,k}!} \left(\delta_0^2\right)^{-\frac{p_{j,k}}{2}} \right) \mathbb{I}_{\mathbb{R}^+}(\gamma) \end{aligned} \tag{30}$$

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

B. 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 referred to as "algorithm 2" 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 $P([\mathrm{r}_{1,i},\ldots,\mathrm{r}_{J,i}]^\top = \boldsymbol{\epsilon}|\mathbf{R}_{-i},\mathbf{p},\gamma,\delta_0^2,\psi,\mathbf{Y})$ 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 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

$$\lambda_{p_{j,k}\to p^*_{j,k}}=\frac{\psi^{p^*_{j,k}}}{\psi^{p_{j,k}}}\frac{p^*_{j,k}!}{p_{j,k}!}\left(\frac{1}{\delta_0^2}\right)^{\pm\frac{1}{2}}\frac{\left|\mathbf{M}_{j,k}\left(p^*_{j,k}\right)\right|^{\frac{1}{2}}}{\left|\mathbf{M}_{j,k}(p_{j,k})\right|^{\frac{1}{2}}}\times\frac{\left[T^2_{j,k}(p_{j,k})+\gamma\right]^{\frac{\nu}{2}+\frac{1}{2}n_{j,k}}}{\left[T^2_{j,k}\left(p^*_{j,k}\right)+\gamma\right]^{\frac{\nu}{2}+\frac{1}{2}n_{j,k}}}\left(\frac{d_{p_{j,k}}}{b_{p_{j,k}}}\right)^{\pm1} \quad (31)$$

with $p^*_{j,k}=p_{j,k}\pm1$.

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 $\psi^*\sim\mathcal{G}(\mu+\sum_{j,k}p_{j,k},\rho+\sum_j K_j)$ and the following acceptance probability:

$$\lambda_{\psi\to\psi^*}=\left[\frac{\Psi_{p_{\max}}(\psi)}{\Psi_{p_{\max}}(\psi^*)}\exp(\psi^*-\psi)\right]^{\sum_{j=1}^J K_j}. \quad (32)$$

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:

$$\sigma^2_{j,k}|\mathbf{R},\mathbf{p},\gamma,\delta_0^2,\mathbf{Y}\sim\mathcal{IG}\left(\frac{\nu+n_{j,k}}{2},\frac{\gamma+T^2_{j,k}}{2}\right) \quad (33)$$

$$\gamma|\mathbf{R},\boldsymbol{\sigma}^2\sim\mathcal{G}\left(\frac{\nu}{2}\sum_j K_j,\frac{1}{2}\sum_{j,k}\frac{1}{\sigma^2_{j,k}}\right). \quad (34)$$

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:

$$\mathbf{a}_{j,k}|\mathbf{R},\mathbf{p},\boldsymbol{\sigma}^2,\delta_0^2,\mathbf{Y}\sim\mathcal{N}\left(\boldsymbol{\mu}_{j,k},\sigma^2_{j,k}\mathbf{M}_{j,k}\right) \quad (35)$$

$$\delta_0^2|\mathbf{R},\mathbf{p},\mathbf{A},\boldsymbol{\sigma}^2\sim\mathcal{IG}\left(\xi+\sum_{j,k}\frac{p_{j,k}}{2},\beta+\sum_{j,k}\frac{||\mathbf{a}_{j,k}||^2}{2\sigma^2_{j,k}}\right). \quad (36)$$

f) **Generation of samples according to** $f(\mathbf{P}|\mathbf{R},\mathbf{Y})$: As in the initial model, the following posterior is obtained:

$$\mathbf{P}|\mathbf{R},\mathbf{Y}\sim\mathcal{D}_{2^J}\left(S_\epsilon(\mathbf{R})+\alpha_{\boldsymbol{\epsilon}}\right). \quad (37)$$

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

Algorithm 2: Hybrid Gibbs Algorithm for Abrupt Change Detection

- Initialization:
  — sample hyperparameter vector $\widetilde{\boldsymbol{\Phi}}^{(0)}=\left(\widetilde{\delta}_0^{2(0)},\widetilde{\gamma}^{(0)},\widetilde{\mathbf{P}}^{(0)}\right)$ from the pdf in (11);
  — sample hyperparameter $\widetilde{\psi}^{(0)}$ from the pdf in (29);
  — for $i=1,\ldots,n-1$ sample, $[\widetilde{\mathrm{r}}^{(0)}_{1,i},\ldots,\widetilde{\mathrm{r}}^{(0)}_{J,i}]$ from the pdf in (6);
  — for $j=1,\ldots,J$, $k=1,\ldots,K_j$, sample $\widetilde{\sigma}^{2(0)}_{j,k}$, $\widetilde{\mathbf{a}}^{(0)}_{j,k}$ and $\widetilde{p}^{(0)}_{j,k}$ from the pdf's in (8), (9) and (28);
  — set $t\leftarrow1$.
- Iterations: for $t=1,2,3,\ldots$, do:
  — for $i=1,\ldots,n-1$, sample $[\widetilde{\mathrm{r}}^{(t)}_{1,i},\ldots,\widetilde{\mathrm{r}}^{(t)}_{J,i}]$ according to the 2^J probabilities defined in step a) below;
  — for $j=1,\ldots,J$, $k=1,\ldots,K_j$, update the model order $\widetilde{p}^{(t)}_{j,k}$ [see step b)]:
    - if $u\sim\mathcal{U}_{[0,1]}\leq b_{\widetilde{p}^{(t-1)}_{j,k}}$, then propose $p^*_{j,k}=\widetilde{p}^{(t-1)}_{j,k}+1$,
      else if $u\sim\mathcal{U}_{[0,1]}\leq b_{\widetilde{p}^{(t-1)}_{j,k}}+d_{\widetilde{p}^{(t-1)}_{j,k}}$, then propose $p^*_{j,k}=\widetilde{p}^{(t-1)}_{j,k}-1$,
    - if $v_p\sim\mathcal{U}_{[0,1]}\leq\lambda_{p_{j,k}\to p^*_{j,k}}$ (see (31)), $\widetilde{p}^{(t)}_{j,k}=p^*_{j,k}$, else $\widetilde{p}^{(t)}_{j,k}=\widetilde{p}^{(t-1)}_{j,k}$;
  — update $\widetilde{\psi}^{(t)}$ [see step c)]:
    - propose ψ^* according to the Gamma proposal distribution defined in step d);
    - if $v_\psi\sim\mathcal{U}_{[0,1]}\leq\lambda_{\psi\to\psi^*}$ (see (32)), $\widetilde{\psi}^{(t)}=\psi^*$, else $\widetilde{\psi}^{(t)}=\widetilde{\psi}^{(t-1)}$;
  — for $j=1,\ldots,J$, $k=1,\ldots,K_j$, sample $\widetilde{\sigma}^{2(t)}_{j,k}$ from the pdf in (33);
  — sample $\widetilde{\gamma}^{(t)}$ from the pdf in (34);
  — for $j=1,\ldots,J$, $k=1,\ldots,K_j$, sample $\widetilde{\mathbf{a}}^{(t)}_{j,k}$ from the pdf in (35);
  — sample $\widetilde{\delta}_0^{2(t)}$ from the pdf in (36);
  — (*optional step*) sample $\widetilde{\mathbf{P}}^{(t)}$ from the pdf in (37);
  — set $t\leftarrow t+1$.

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	–	–

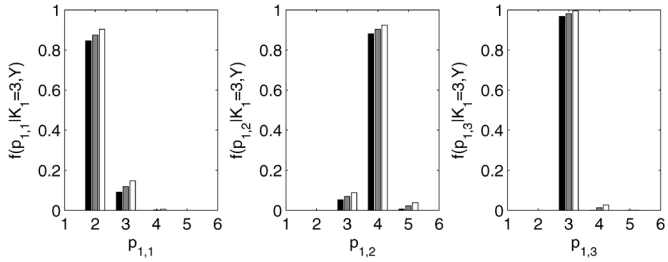


Fig. 9. Posterior distributions of the model orders p_{1i} (for $i = 1, \ldots, 3$) conditioned to $K_1 = 3$.

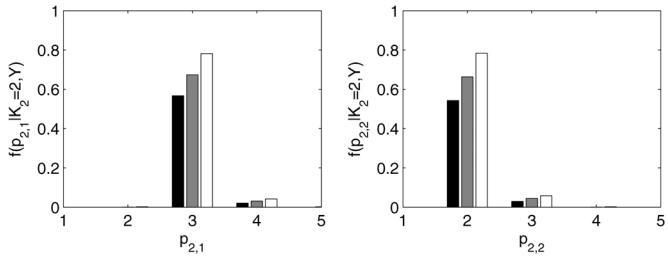


Fig. 10. Posterior distributions of the model orders p_{2i} (for $i = 1,2$) conditioned to $K_2 = 2$.

C. 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 $\boldsymbol{l}_1 = (60, 150)$ and $\boldsymbol{l}_2 = (60)$. 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 Figs. 9 and 10. The corresponding change-point posterior distributions are shown on Fig. 11. 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.

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

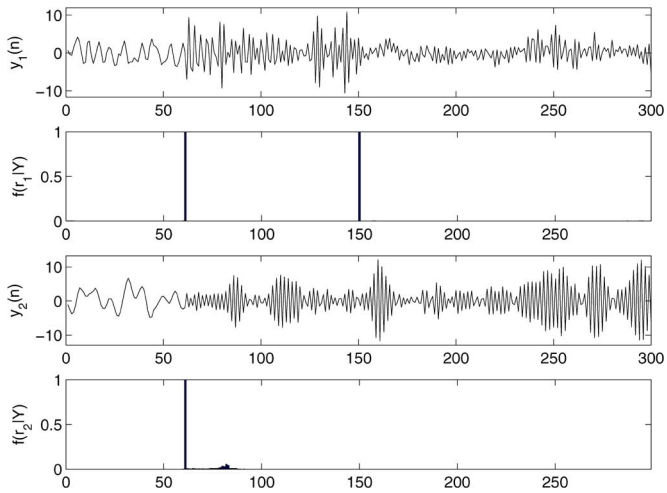


Fig. 11. Posterior distribution of the change-point locations estimated by the reversible-jump algorithm.

on military and commercial aircrafts are subject to various constraints (chemical, mechanical, thermic, etc.) resulting in insulation damages. These breakdowns expose the cable to intermittent fault-arc currents that could ignite the neighboring wires [24]. Several methods for detection of wiring failures have been studied in the literature: they are mainly based on dielectric properties [25] or, more recently, on electromagnetic properties [26]. 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 three-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 high-pass filter in order to highlight the transients which are much less energetic. The filtered voltages can be accurately modeled as AR 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 $N_{\mathrm{MC}} = 450$ iterations including a burn-in period of $N_{\mathrm{bi}} = 100$ iterations. The parameters N_{bi} and N_{MC} have been

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

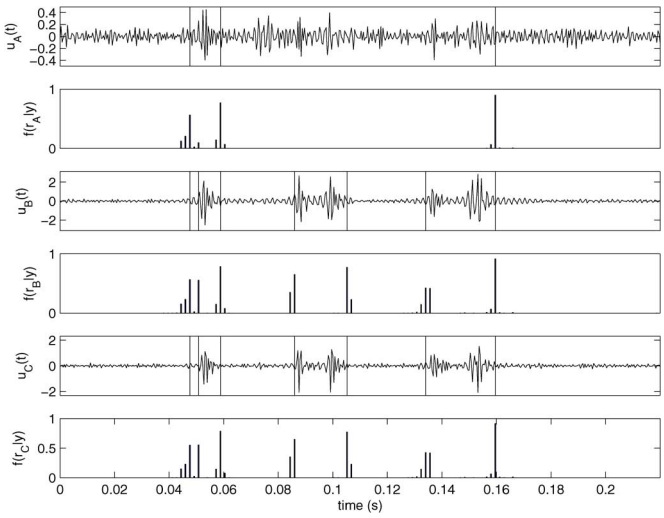


Fig. 12. Posterior distribution of the change-point locations and segmentation of three-dimensional aeronautical data ($N_{\rm bi} = 100$ and $N_{\rm MC} = 450$).

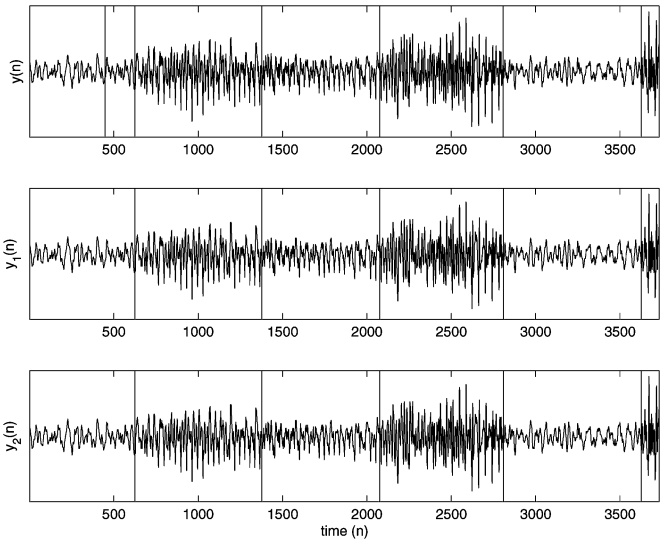


Fig. 13. Segmentations of 1-D (top) and 2-D (middle and bottom) real speech data ($N_{\rm bi} = 200$ and $N_{\rm MC} = 600$).

chosen to provide appropriate potential reduction factors $\sqrt{\hat{\rho}}$ for the hyperparameters $\mathrm{P}_{\boldsymbol{\epsilon}}$. Note that running 30 iterations of the proposed algorithm takes approximately 4 min for a MATLAB implementation on a 2.8-GHz Pentium IV.

In the first step of the analysis, we estimate the number of change-points for the observed sequences. The posteriors of the number of changes in each signal are not depicted here for brevity. The corresponding MAP estimators are $\widehat{K}_1 = 4$, $\widehat{K}_2 = 8$, and $\widehat{K}_3 = 8$. The estimated posterior distribution of $\mathbf{R}$ depicted in Fig. 12 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 in Fig. 12) 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 [23]). As explained in [1, p. 401], this signal 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 8 kHz and quantized with 16 bits. It is prefiltered by a highpass filter with cutoff frequency equal to 150 Hz. The raw one-dimensional (1-D) data $\mathbf{y} = [y_1, \ldots, y_n]$ have been processed by the proposed algorithm with $J = 1$. The estimated number of change points and their positions are obtained after $N_{\rm MC} = 600$ iterations, including a burn-in period of $N_{\rm bi} = 200$ iterations ($N_{\rm MC}$ and $N_{\rm bi}$ have been chosen in order to obtain appropriate potential reduction factors $\sqrt{\hat{\rho}}$ for the hyperparameters $\mathrm{P}_{\boldsymbol{\epsilon}}$). The estimated changes are depicted in Fig. 13 (top figure). Table IV compares the estimates with those obtained with several other methods previously studied in the literature. It clearly appears that the proposed 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 $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2]^\top$ 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. 13 (middle and bottom plots). They are in good agreement with the 1-D 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. Finally, it is interesting to point out that running 1 iteration of the joint segmentation algorithm takes approximately 30 s for a MATLAB implementation on a 2.8-GHz Pentium IV.

VII. Conclusion

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 AR 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. The proposed algorithm was initially developed for AR signals with known orders. However, an extension to models with unknown orders was also presented. Two applications were finally investigated: arc-tracking detection and stereo speech signal segmentation. The results obtained in these applications are very encouraging.

Note that the assumptions 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

TABLE IV
CHANGE-POINT POSITIONS ESTIMATED BY DIFFERENT METHODS

Method	AR order	Estimated change-points									
Divergence [28]	16	445	–	645	1550	1800	2151	2797	–	3626	–
GLR [29]	16	445	–	645	1550	1800	2151	2797	–	3626	–
GLR [29]	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 [23]	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	–

to more general models appropriate for signals with heterogeneous dynamics [16], long-range-dependent data [4] or generalized autoregressive conditional heteroskedastic (GARCH) signals [27] would also be an interesting issue.

APPENDIX
FAST COMPUTATIONS

It is interesting to notice that the matrices $\mathbf{Q}_{j,k}$ and $\mathbf{M}_{j,k}$, and the variable $T_{j,k}$ defined in (15) could be computed following the implementations described in [17]. We note $\mathbf{y}_{j,[k]} = \mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}^{\top}$.

Algorithm 3: Fast Computations of $T_{j,k}^2$

- Compute $\mathbf{M}_{j,k}^{-1} = \mathbf{Y}_{j,k}^{\top}\mathbf{Y}_{j,k} + \mathbf{I}_p/\delta_0^2$.
- Compute Cholesky's factors $\mathbf{C}_{j,k}$ such as $\mathbf{C}_{j,k}\mathbf{C}_{j,k}^{\top} = \mathbf{M}_{j,k}$.
- Compute $\mathbf{u}_{j,k} = \mathbf{Y}_{j,k}^{\top}\mathbf{y}_{j,[k]}$.
- Solve the system $\mathbf{C}_{j,k}\mathbf{v}_{j,k} = \mathbf{u}_{j,k}$ for $\mathbf{v}_{j,k}$.
- Compute $T_{j,k}^2 = \mathbf{y}_{j,[k]}^{\top}\mathbf{y}_{j,[k]} - \mathbf{v}_{j,k}^{\top}\mathbf{v}_{j,k}$.

Such implementations allow us to develop a strategy to sample $\mathbf{a}_{j,k}|\delta_0^2, \sigma_{j,k}^2, \mathbf{R}, \mathbf{Y}$ according to $\mathcal{N}(\boldsymbol{\mu}_{j,k}, \sigma_{j,k}^2\mathbf{M}_{j,k})$ in the effective following scheme.

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

- Sample an i.i.d. vector $\mathbf{w}_{j,k}$ according to $\mathcal{N}(\mathbf{0}, \sigma_{j,k}^2\mathbf{I}_p)$.
- 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 $\mathbf{C}_{j,k}^{\top}\boldsymbol{\mu}_{j,k} = \boldsymbol{\mu}'_{j,k}$ 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 (14). Indeed, $|\mathbf{M}_{j,k}| = |\mathbf{C}_{j,k}|^{-2}$, where $\mathbf{C}_{j,k}$ are upper triangular matrices whose determinants can be computed very easily.

ACKNOWLEDGMENT

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.

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, 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," *J. Time Series Anal.*, vol. 21, no. 1, pp. 33–59, Jan. 2000.
[5] L. Birgé and P. Massart, "Gaussian model selection," *J. 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 Process.*, vol. 85, no. 4, pp. 717–736, Apr. 2005.
[7] R. E. McCulloch and R. S. Tsay, "Bayesian inference and prediction for mean and variance shifts in autoregressive time series," *J. Amer. Statist. Assoc.*, vol. 88, no. 423, pp. 968–978, 1993.
[8] P. M. Djurić, "A MAP solution to off-line segmentation of signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1994, vol. 4, 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 Process.*, vol. 50, no. 3, pp. 747–758, Mar. 2002.
[10] E. Kuhn and M. Lavielle, "Coupling a stochastic approximation version of EM with an MCMC procedure," *ESAIM Probab. Statist.*, vol. 8, pp. 115–131, 2004.
[11] M. Lavielle and E. Lebarbier, "An application of MCMC methods for the multiple change-points problem," *Signal Process.*, 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 Process.*, 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 Process.*, vol. 36, pp. 29–40, Jan. 1988.
[15] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[16] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," IRIT/ENSEEIHT/TéSA and LAGIS, Tech. Rep., Mar. 2006 [Online]. Available: http://www.enseeiht.fr/~dobigeon
[17] M. Davy and J. Idier, "Fast MCMC computations for the estimation of sparse processes from noisy observations," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montréal, QC, Canada, May 2004, vol. 2, pp. 1041–1044.
[18] 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.
[19] A. Gelman and D. Rubin, "Inference from iterative simulation using multiple sequences," *Statist. Sci.*, vol. 7, no. 4, pp. 457–511, 1992.
[20] S. Godsill and P. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech, Audio Process.*, vol. 6, no. 4, pp. 352–372, 1998.
[21] P. M. Djurić and J.-H. Chun, "An MCMC sampling approach to estimation of nonstationary hidden Markov models," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1113–1123, 2002.

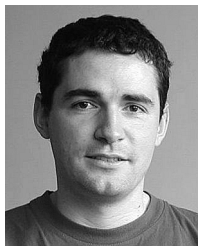
[22] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 1995.
[23] P. Fearnhead, "Exact Bayesian curve fitting and signal segmentation," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2160–2166, Jun. 2005.
[24] C. Furse and R. Haupt, "Down to the wire," *IEEE Spectrum*, pp. 35–39, Feb. 2001.
[25] J. Hanson and D. Koenig, "Fault arc effects in cable bundles for space applications in vacuum," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 4, no. 4, pp. 394–399, Aug. 1997.
[26] 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.
[27] D. Guégan, *Séries chronologiques non linéaires à temps discret*. Paris, France: Economica, 1994.
[28] M. Basseville and A. Benveniste, "Design and comparative study of some sequential jump detection algorithms for digital signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, pp. 521–535, 1983.
[29] U. Appel and A. V. Brandt, "Adaptive sequential segmentation of piecewise stationary time series," *Inf. Sci.*, vol. 29, no. 1, pp. 27–56, 1983.

Nicolas Dobigeon (S'05) was born in Angoulême, France, in 1981. He received the Eng. degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, France, both in June 2004. He is currently working towards the Ph.D. degree with the Signal and Communication Group at the IRIT Laboratory, Toulouse, France. His research interests are centered around Bayesian inference and Markov chain Monte Carlo (MCMC) methods for the joint segmentation of multiple signals or images.

Jean-Yves Tourneret (M'94) received the ingénieur degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique in Toulouse (ENSEEIHT), France, and the Ph.D. degree from the National Polytechnic Institute, Toulouse, France, in 1992.

He is currently a Professor in the University of Toulouse (ENSEEIHT), France. He is a member of the IRIT laboratory (UMR 5505 of the CNRS), where his research activity is centered around estimation, detection, and classification of non-Gaussian and nonstationary processes.

Dr. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO), which was held in Toulouse, France, in 2002. He was also a member of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2006 organizing committee. He has been a member of different technical committees, including the Signal Processing Theory and Methods (SPTM) Committee of the IEEE Signal Processing Society.



Manuel Davy was born in Caen, France, in 1972. He received the Ingénieur degree in electrical engineering from the Ecole Centrale de Nantes, Nantes, France, in 1996 and the Ph.D. degree from the University of Nantes in 2000.

From 2000 to 2002, he was a Research Associate with the Signal Processing Group, University of Cambridge, Cambridge, U.K. He is currently a Chargé de Recherches CNRS, LAGIS, Lille, France, and member of the INRIA-FUTURS "SequeL" team, where his research activity is centered around kernel algorithms and Bayesian/Monte Carlo methods for signal processing and reinforcement learning.