

# JOINT SEGMENTATION OF PIECEWISE CONSTANT AUTOREGRESSIVE PROCESSES BY USING A HIERARCHICAL MODEL AND A BAYESIAN SAMPLING APPROACH

Nicolas Dobigeon<sup>†</sup>, Jean-Yves Tournet<sup>†</sup> and Manuel Davy<sup>\*</sup>

<sup>†</sup>IRIT/ENSEEIH/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France

<sup>\*</sup>LAGIS, BP 48, Cité Scientifique, 59651 Villeneuve d'Ascq cedex, France

Nicolas.Dobigeon@enseeiht.fr, Jean-Yves.Tournet@enseeiht.fr, Manuel.Davy@ec-lille.fr

## 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 provides interesting results compared to a signal-by-signal segmentation.

## 1. INTRODUCTION AND PROBLEM FORMULATION

This paper addresses the problem of segmenting correlated signals recorded from several sensors. Of course, signal and image segmentation has already received much attention in the signal and image processing literature (see [1, 2, 3] and references therein). The main contribution of this study is to introduce correlations between the change-points of the different observed signals. More precisely, when a change is detected in one or several signals at a given position, the proposed algorithm allows to favor the presence of a change at this position in the other signals. This change-point correlation is obtained within a Bayesian framework by defining appropriate change-point priors. A similar strategy was proposed in [4] for joint detection of variance changes.

The signals  $\mathbf{y}_j = [y_{j,1}, \dots, y_{j,n}]$ , where  $j = 1, \dots, J$  and  $n$  is the number of observed samples in each signal, are modeled by piecewise constant autoregressive (AR) processes:

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

In Eq. (1), each signal  $\mathbf{y}_j$  features  $K_j$  segments (i.e.,  $k = 1, \dots, K_j$ ). The AR parameter vector of the  $k^{\text{th}}$  segment in the  $j^{\text{th}}$  signal is  $\mathbf{a}_{j,k} = [a_{j,k,1}, \dots, a_{j,k,p}]^T$ . The time indexes immediately after which a changes occur are denoted  $l_{j,k}$  (with the convention that  $l_{j,0} = 0$  and  $l_{j,K_j} = n$ ). Finally,  $\mathbf{e}_j = [e_{j,1}, \dots, e_{j,n}]^T$  is a vector of i.i.d. zero mean Gaussian noise samples. The noise vectors  $\mathbf{e}_1, \dots, \mathbf{e}_J$  are assumed independent.

AR modeling of the observations 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  $S(f)$  [5, p. 130]. Note that the orders of the AR models in (1) are assumed equal in all segments of the observed time-series<sup>1</sup>. By using the notation

<sup>1</sup>This assumption could be relaxed: the orders of the AR processes in each segment could actually be estimated by introducing a convenient prior in the Bayesian model, as well as reversible jumps [6] in the algorithm presented in Section 3.

$\mathbf{x}_{j,i:i'} = [x_{j,i}, \dots, x_{j,i'}]^T$ , the set of equations (1) can be written in the following matrix form:

$$\mathbf{y}_{j,l_{k-1}+1:l_k} = \mathbf{Y}_{j,k} \mathbf{a}_{j,k} + \mathbf{e}_{j,l_{k-1}+1:l_k},$$

where  $\mathbf{Y}_{j,k}$  denotes a matrix of size  $(l_{j,k} - l_{j,k-1}) \times p$  (whose entries are delayed signal samples  $y_{j,i}$ ). This paper addresses the problem of estimating the change-point locations  $l_{j,k}$  from the  $J$  observed time series  $\mathbf{y}_j$ ,  $j = 1, \dots, J$ .

Section 2 studies the joint segmentation problem by using a hierarchical model. This model assumes that appropriate prior distributions for the unknown parameters (change-point locations, AR parameters and noise variances) are available. The parameters of these priors (referred to as hyperparameters) can be specified by using appropriate information regarding the observations. However, several authors have proposed to increase the robustness of the estimates by introducing a second level of hierarchy within the Bayesian paradigm [3] [7]. Such methodology consists of assigning vague priors to the hyperparameters and to estimate these parameters from the observed data. A hierarchical Bayesian model based on these ideas is presented in Section 2. The change-point Bayesian estimators such as the Maximum a posteriori (MAP) or minimum mean square (MMSE) estimators are difficult to derive from the hierarchical model. A joint segmentation procedure based on the Gibbs sampler is studied in Section 3. Theoretical properties are illustrated by simulation which are presented in Section 4. Conclusions are reported in Section 5.

## 2. HIERARCHICAL BAYESIAN MODEL

The unknown parameters for the joint segmentation problem introduced in the previous section are the numbers of segments  $K_j$ , the change-point locations  $l_{j,k}$ , the noise variances  $\sigma_{j,k}^2$  (with  $\boldsymbol{\sigma}_j^2 = [\sigma_{j,1}^2, \dots, \sigma_{j,K_j}^2]^T$ ) and the AR parameter vectors  $\mathbf{a}_{j,k}$  (which are collectively denoted  $\mathbf{A}_j = (\mathbf{a}_{j,1}, \dots, \mathbf{a}_{j,K_j})$  for each observation). A standard reparametrization consists of introducing indicator variables  $r_{j,i}$  ( $j \in \{1, \dots, J\}$ ,  $i \in \{1, \dots, n\}$ ) such that:

$$\begin{cases} r_{j,i} = 1 & \text{if there is a changepoint at time } i \text{ of the } j^{\text{th}} \text{ sequence,} \\ r_{j,i} = 0 & \text{otherwise,} \end{cases}$$

with  $r_{j,n} = 1$  (this condition ensures that the number of change-points and the number of steps in the  $j^{\text{th}}$  sequence are equal to  $K_j = \sum_{i=1}^n r_{j,i}$ ). When using indicator variables, the unknown parameter vector is  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J\}$  where  $\boldsymbol{\theta}_j = (\mathbf{r}_j, \boldsymbol{\sigma}_j^2, \mathbf{A}_j)$  and  $\mathbf{r}_j = [r_{j,1}, \dots, r_{j,n}]$ . 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

to estimate the unknown parameter vector  $\theta$  by using Bayesian estimation theory. Bayesian inference on  $\theta$  is based on the posterior distribution  $f(\theta|\mathbf{Y})$ , with  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_J)$ . This posterior distribution is related to the likelihood of the observations and the parameter priors via Bayes theorem  $f(\theta|\mathbf{Y}) \propto f(\mathbf{Y}|\theta)f(\theta)$ . The likelihood and priors for the joint segmentation problem are summarized below.

## 2.1. Approximate Likelihood

The dependence of the exact likelihood  $f(\mathbf{y}_j|\theta_j)$  on the  $p$  first samples  $\mathbf{y}_{j,1:p}$  is classically omitted, to simplify the analysis (the reader is invited to consult [8, p.186] for more details). In other words, by using the independence assumption between the noise vectors  $\mathbf{e}_j$ ,  $j = 1, \dots, J$ , the exact likelihood of  $\mathbf{Y}$  is approximated as follows:

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

where  $n_{j,k}(\mathbf{r}_j) = l_{j,k} - l_{j,k-1}$  is the length of segment  $k$  in the  $j^{\text{th}}$  observed sequence and

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

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

## 2.2. Parameter Priors

1) **Indicator variable prior:** Possible correlations between the change locations in the different observed signals are modeled by an appropriate prior distribution  $f(\mathbf{R}|\mathbf{P})$ , where  $\mathbf{R} = [r_1, \dots, r_J]^T$ . More precisely, we assume that the probability  $P_\epsilon$  of having  $[r_{1,i}, \dots, r_{J,i}] = \epsilon$  (where  $\epsilon \in \mathcal{E} = \{0, 1\}^J$ ) does not depend on the time index  $i$ . As a consequence, by assuming that  $[r_{1,i}, \dots, r_{J,i}]$  is independent of  $[r_{1,i'}, \dots, r_{J,i'}]$  for any  $i \neq i'$ , the indicator prior distribution is expressed as:

$$f(\mathbf{R}|\mathbf{P}) = \prod_{\epsilon \in \mathcal{E}} P_\epsilon^{S_\epsilon(\mathbf{R})},$$

where  $S_\epsilon(\mathbf{R})$  is the number of times  $i$  such that  $[r_{1,i}, \dots, r_{J,i}] = \epsilon$ . With this prior, a high value of  $P_\epsilon$  indicates very likely configurations  $[r_{1,i}, \dots, r_{J,i}] = \epsilon$  for all  $i = 1, \dots, n$ . For instance, by choosing a high value of  $P_{0\dots 0}$  (resp.  $P_{1\dots 1}$ ), we will favor a simultaneous absence (resp. presence) of changes in all observed signals. This choice introduces correlation between the change-point locations.

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

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

where  $\mathcal{IG}(a, b)$  denotes the Inverse-Gamma distribution with parameters  $a$  and  $b$ ,  $\nu = 2$  and  $\gamma$  is an adjustable hyperparameter. We assume in this paper that the hyperparameter  $\gamma$  is the same for all observed signals. Note, however, that a similar analysis could be conducted with a set of different hyperparameters

$\gamma_j$ ,  $j = 1, \dots, J$ . For instance, such analysis would be interesting when signal amplitudes differ significantly from one signal to another.

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

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

where  $\mathbf{I}_p$  is the  $p \times p$  identity matrix,  $\mathbf{0}_p$  is the vector made of  $p$  zeros and  $\delta_0^2$  is an adjustable hyperparameter. These conjugate priors have already been proposed in [3] and allow the marginalization of the posterior  $f(\theta|\mathbf{Y})$  with respect to noise variances and AR parameters.

The hyperparameter vector associated to the parameter priors defined above is  $\Phi = (\mathbf{P}, \delta_0^2, \gamma)$ . Of course, the ability of this Bayesian model to segment accurately depends on the values of the hyperparameters. In particular applications, these hyperparameters can be fixed from available information regarding the observed signals. However, in order to increase the model robustness, hyperparameters can be considered as unknown with vague priors, as in [3]. This strategy involves several levels of hierarchy within the Bayesian paradigm and results in a so-called hierarchical Bayesian model. This hierarchical model requires to define hyperparameter priors (sometimes referred to as hyper-priors) which are detailed in the next section.

## 2.3. Hyperparameter priors

The priors for hyperparameters  $\delta_0^2$  and  $\gamma$  are a noninformative Jeffreys' prior [3] and a vague conjugate Inverse-Gamma distribution (i.e, with large variance) which reflect the absence of precise knowledge regarding these hyperparameters:

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

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

The prior distribution for the hyperparameter  $\mathbf{P}$  is a Dirichlet distribution with parameter vector  $\alpha = [\alpha_{0\dots 0}, \dots, \alpha_{1\dots 1}]^T$  defined on the simplex  $\mathcal{P} = \{\mathbf{P} \text{ such that } \sum_{\epsilon \in \mathcal{E}} P_\epsilon = 1, P_\epsilon > 0\}$  denoted as  $\mathbf{P} \sim D_{2,J}(\alpha)$  (for more details see [7, p. 383]). This prior indicator variable distribution has been chosen since it allows marginalization of the posterior distribution  $f(\theta|\mathbf{Y})$  with respect to  $\mathbf{P}$ . Moreover, by choosing  $\alpha_\epsilon = 1, \forall \epsilon \in \mathcal{E}$ , the Dirichlet distribution reduces to the uniform distribution on  $\mathcal{P}$ .

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

$$f(\Phi | \alpha, \xi, \beta) \propto \left( \prod_{\epsilon \in \mathcal{E}} P_\epsilon^{\alpha_\epsilon - 1} \right) \frac{1}{\gamma} \frac{\beta^\xi}{\Gamma(\xi)(\delta_0^2)^\xi} \exp\left(-\frac{\beta}{\delta_0^2}\right), \quad (3)$$

where  $\Gamma(\cdot)$  is the gamma function.

## 2.4. Posterior distribution of $\theta$

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

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

where

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

and  $f(\mathbf{Y}|\theta)$  and  $f(\Phi)$  are defined in Eq.'s (2) and (3). This hierarchical structure allows to integrate out the nuisance parameters  $\mathbf{a}_{j,k}$ ,  $\sigma_{j,k}^2$  and  $\mathbf{P}$  from the joint distribution  $f(\theta, \Phi|\mathbf{Y})$ , yielding:

$$f(\mathbf{R}, \gamma, \delta_0^2 | \mathbf{Y}) \propto \prod_{j=1}^J \prod_{k=1}^{K_j(\mathbf{r}_j)} \left( \frac{\gamma^{\frac{\nu}{2}} |\mathbf{M}_{j,k}|^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)\right)}{(\gamma + T_{j,k}^2)^{\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)}} \right) \\ \times \frac{1}{\gamma} (\delta_0^2)^{-\frac{\nu}{2} \sum_{j=1}^J K_j(\mathbf{r}_j)} f(\delta_0^2 | \xi, \beta) C(\mathbf{R} | \mathbf{Y}, \alpha), \quad (4)$$

$$\text{with} \quad \begin{cases} T_{j,k}^2 = \mathbf{y}_{j,l_{k-1}+1:l_k}^\top \mathbf{Q}_{j,k} \mathbf{y}_{j,l_{k-1}+1:l_k}, \\ \mathbf{Q}_{j,k} = \mathbf{I}_p - \mathbf{Y}_{j,k} \mathbf{M}_{j,k} \mathbf{Y}_{j,k}^\top, \\ \mathbf{M}_{j,k} = \left( \mathbf{Y}_{j,k}^\top \mathbf{Y}_{j,k} + \frac{\mathbf{I}_p}{\delta_0^2} \right)^{-1}, \end{cases} \quad (5)$$

and

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

The posterior distribution in Eq. (4) 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 [3], with the two noticeable differences: 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.

### 3. A GIBBS SAMPLER FOR JOINT SIGNAL SEGMENTATION

Gibbs sampling is an iterative sampling strategy which consists of generating samples distributed according to the full conditional distributions of each parameter. This paper proposes to sample according to the distribution  $f(\mathbf{R}, \gamma, \delta_0^2 | \mathbf{Y})$  defined in (4) by the following two step procedure:

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

This generation is achieved by using the Gibbs Sampler, to generate Monte Carlo samples distributed according to  $f(r_{1,i}, \dots, r_{J,i} | \gamma, \delta_0^2, \mathbf{Y})$ . This vector is a random vector of Booleans in  $\mathcal{E}$ . Consequently, its distribution is fully characterized by the probabilities  $P(r_i = \epsilon | \gamma, \delta_0^2, \mathbf{Y})$ ,  $\epsilon \in \mathcal{F}$ . By using the notation  $\mathbf{R}_{-i}$  to denote the matrix  $\mathbf{R}$  where the column at time  $i$  is suppressed, the following result can be obtained:

$$P([r_{1,i}, \dots, r_{J,i}] = \epsilon | \mathbf{R}_{-i}, \gamma, \delta_0^2, \mathbf{Y}) \propto f(\mathbf{R}_i(\epsilon), \gamma, \delta_0^2 | \mathbf{Y}),$$

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

#### Generation of samples distributed 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  $f(\gamma, \delta_0^2, \boldsymbol{\sigma}^2, \mathbf{A} | \mathbf{R}, \mathbf{Y})$ , where  $\boldsymbol{\sigma}^2 = \{\sigma_1^2, \dots, \sigma_J^2\}$  and  $\mathbf{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_J\}$ , by using Gibbs moves. By looking

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

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

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

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

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

This is achieved as follows:

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

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

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

#### Posterior distribution of $P_\epsilon$

The hyperparameters  $P_\epsilon$ ,  $\epsilon \in \mathcal{E}$ , 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):

$$f(\mathbf{P} | \mathbf{R}, \mathbf{Y}) \propto \mathcal{D}_{2^J}(S_\epsilon(\mathbf{R}) + \alpha_\epsilon). \quad (6)$$

### 4. SIMULATION RESULTS

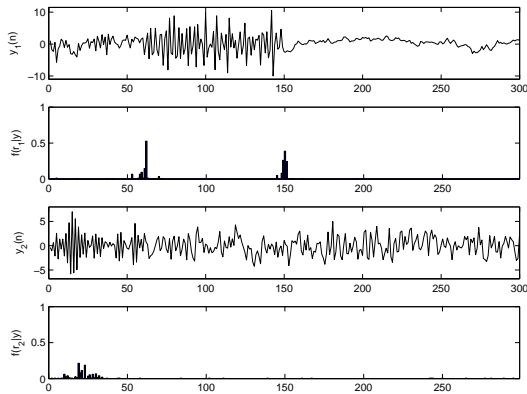
The first simulation is aimed at justifying the use of joint segmentation as opposed to parallel single signal segmentations. Here, one considers two independent AR processes. The change-point locations are  $l_1 = (60, 150, 300)$  for signal #1 and  $l_2 = (60, 300)$  for signal #2. The fixed parameters and hyperparameters have been chosen as follows:  $\nu = 2$  (as in [3]),  $\xi = 1$  and  $\beta = 100$  (in order to obtain vague hyperprior),  $\alpha_\epsilon = 1, \forall \epsilon \in \mathcal{E}$  so as to obtain a uniform distribution. It is interesting to note that the matrices  $T_{j,k}^2$ ,  $\mathbf{Q}_{j,k}$  and  $\mathbf{M}_{j,k}$  defined in (5) have been computed following the implementations described in [9]. Fig.'s 1 and 2 show the posterior distributions of the change-locations obtained for the two time-series. These posterior distributions have been obtained by averaging the  $N_r = 150$  last outputs of the Gibbs sampler (the first  $N_{bi} = 50$  burn-in iterations have been ignored). 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 first signal favors the detection of a change at the same time index in the other signal.

The estimation of the total number of change-points for the two time-series is an important problem. The proposed algorithm generates samples  $(\mathbf{R}^{(m)}, \gamma^{(m)}, \delta_0^{2(m)})$  distributed according to the posterior distribution  $f(\mathbf{R}, \gamma, \delta_0^2 | \mathbf{Y})$ , which allows for model selection. Indeed, for each sample  $\mathbf{R}^{(m)}$ , the number of change-points are  $\hat{K}_1^{(m)}(\mathbf{R}^{(m)}) = \sum_{k=1}^N r_{1,k}$  and  $\hat{K}_2^{(m)}(\mathbf{R}^{(m)}) = \sum_{k=1}^N r_{2,k}$ . Fig. 3 shows the histograms of  $\hat{K}_1$  and  $\hat{K}_2$  computed

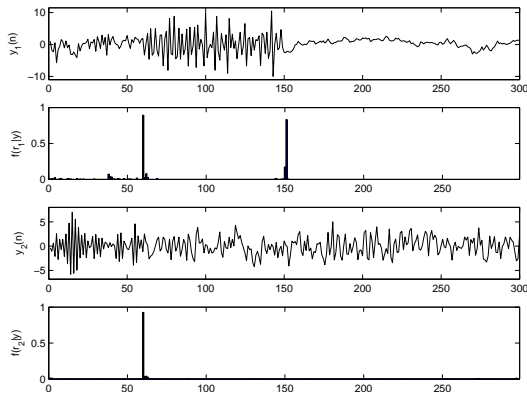
from the 150 last Markov chain samples with the joint approach. The histograms have maximum values for  $K_1 = 3$  and  $K_2 = 2$  which correspond to the actual numbers of changes. The estimation of the noise variances or AR parameters can be interesting in practical applications. Fig. 4 shows the posterior distributions of parameters  $\sigma_{21}^2$  and  $\sigma_{22}^2$  associated to the second time-series  $y_2$ . These histograms are in good agreement with the actual values of the parameters  $\sigma_{21}^2 = 1.49$  and  $\sigma_{22}^2 = 3.01$  (which have been generated uniformly in  $[0, 5]$ ). 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 shown in Fig. 5. The results clearly coincide with the theoretical distribution (6).

## 5. CONCLUSIONS

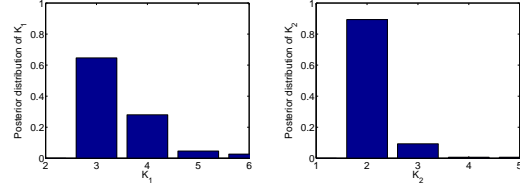
This paper studied a joint Bayesian segmentation procedure allowing to segment signals recorded from different sensors. The proposed approach assumed that the signals can be modeled by piecewise constant autoregressive processes. This mild assumption allows to handle a large class of real signals such as biomedical and speech signals. The application of this strategy to detect changes in arc tracking signals (see [4] for more details) is currently under investigation.



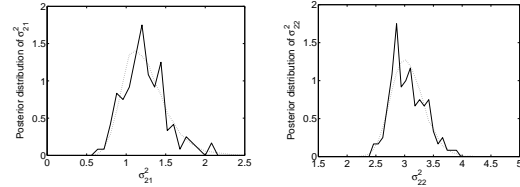
**Fig. 1.** Posterior distributions of the change-point locations (two single signal independent segmentations).



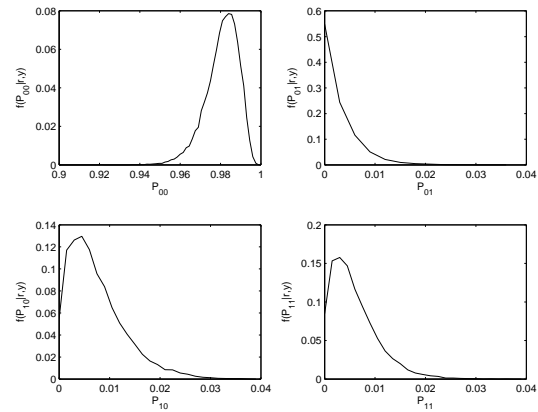
**Fig. 2.** Posterior distributions of the change-point locations (one joint signals segmentation).



**Fig. 3.** Posterior distributions of the change-point numbers.



**Fig. 4.** Posterior distributions of the noise variances  $\sigma_{2i}^2$  (for  $i = 1, 2$ ) conditioned to  $K_2 = 2$  (solid lines). Averaged posterior distributions from 100 Markov chains (dotted lines),



**Fig. 5.** Posterior distributions of the hyperparameters  $P_e$  (computed from 100 Markov chains).

## 6. REFERENCES

- [1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, 1993.
- [2] P. M. Djurić, "A MAP solution to off-line segmentation of signals," in *Proc. IEEE ICASSP-94*, vol. 4, pp. 505–508, 1994.
- [3] E. Punsakaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Processing*, vol. 50, pp. 747–758, March 2002.
- [4] M. Chabert, J.-Y. Tourneret, and M. Coulon, "Joint detection of variance changes using hierarchical Bayesian analysis," in *Proc. IEEE-SP Workshop Stat. and Signal Processing*, pp. 593–596, Sept. 2003.
- [5] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York: Springer Verlag, 1990.
- [6] P. J. Green, "Reversible jump MCMC computation and bayesian model determination," *Biometrika*, vol. 82, 1995.
- [7] C. P. Robert, *The Bayesian Choice - A Decision Theoretic Motivation*. New-York: Springer Verlag, 1994.
- [8] S. M. Kay, *Modern spectral estimation*. Prentice Hall, 1988.
- [9] M. Davy and J. Idier, "Fast MCMC computations for the estimation of sparse processes from noisy observations," in *Proc. ICASSP-04*, vol. 2, pp. 1041–1044, May 2004.