Joint Segmentation of Multivariate Astronomical Time Series: Bayesian Sampling With a Hierarchical Model

Nicolas Dobigeon, Jean-Yves Tourneret, Jeffrey D. Scargle

Joint Segmentation of Multivariate Astronomical Time Series: Bayesian Sampling with a Hierarchical Model

Nicolas Dobigeon, Jean-Yves Tourneret, and Jeffrey D. Scargle

***Abstract*— Astronomy and other sciences often face the problem of detecting and characterizing structure in two or more related time series. This paper approaches such problems using Bayesian priors to represent relationships between signals with various degrees of certainty, and not just rigid constraints. The segmentation is conducted by using a hierarchical Bayesian approach to a piecewise constant Poisson rate model. A Gibbs sampling strategy allows joint estimation of the unknown parameters and hyperparameters. Results obtained with synthetic and real photon counting data illustrate the performance of the proposed algorithm.**

***Index Terms*— Segmentation, photon counting data, Gibbs sampling, Markov chain Monte Carlo, hierarchical Bayesian analysis.**

I. Introduction

The problem of signal segmentation has received increasing attention in the signal processing literature. A complete bibliography of references published before 1993 can be found in the textbooks of Basseville and Nikiforov [1] and Brodsky and Darkhovsky [2]. However, intensive research has been conducted since 1993 on developing new segmentation algorithms. A first class of algorithms adopts a model selection approach via penalization. A parametric model is defined for the signal of interest including changepoints between an unknown number of segments. The changepoints are then estimated by minimizing an appropriate penalized criterion. Note that the penalization is necessary to avoid over-segmentation. Penalized contrast criteria that have been proposed for signal segmentation include the penalized least squares criterion [3] and the generalized C_p criterion for Gaussian model selection [4], [5]. A second class of algorithms is based on Bayesian inference. These algorithms consist of defining appropriate prior distributions for the unknown signal parameters (including the changepoints between the different segments) and estimating these unknown parameters from their posterior distributions. Bayesian estimators recently used for signal segmentation include the maximum *a posteriori* (MAP) estimator [6], [7], the minimum mean square estimator (MMSE) [8] and the hierarchical Bayesian curve fitting estimator [9]. Note that the complexity of the posterior distributions for the unknown parameters generally requires appropriate simulation methods such as Markov Chain Monte Carlo (MCMC) methods [7], [8], [9] or perfect simulation techniques [10].

The prior distributions appropriate for Bayesian signal segmentation involve hyper-parameters, which may be difficult to estimate. There are two main approaches to estimating these hyperparameters. The first approach couples MCMCs with an expectation-maximization (EM) algorithm, which allows one to estimate the unknown hyperparameters [11]. The second approach defines non-informative prior distributions for the hyperparameters, introducing a second level of hierarchy within the Bayesian paradigm. The hyperparameters are then integrated out of the joint posterior distribution or estimated from the observed data [9].

Surprisingly, the segmentation of astronomical time series has received less attention in the signal and image processing community. An iterative Bayesian algorithm based on a constant Poisson rate model was recently studied to solve this problem [12]. The main idea of the proposed algorithm is to decompose the observed signal into two subintervals (by optimizing an appropriate criterion), to apply the same procedure on the two subintervals and to continue this operation several times. The main advantage of this procedure is to handle only one changepoint at each step. However, the accuracy of the algorithm is limited by its greediness and the fact that an appropriate stopping rule is required. The multiple changepoint algorithm presented in [13] removes these limitations, but requires the specification of a prior distribution for the number of changepoints, and has the disadvantage that it does not automatically provide information on the significance of the optimally determined parameters.

This paper studies a new Bayesian time series segmentation algorithm that does not use a stopping rule and allows one to segment jointly multiple signals coming from different sensors. The proposed strategy is based on a hierarchical model for the segmentation problem. This model assumes that appropriate prior distributions for the unknown parameters (change-point locations, Poisson parameters) are available. Vague priors are then assigned to the corresponding hyperparameters, which are integrated out from the joint posterior distribution (when possible) or estimated from the observed data. MCMC methods are used to draw samples according to the posteriors of interest. The Bayesian estimators are finally computed from these simulated samples. The proposed methodology is similar to the hierarchical Bayesian curve fitting technique studied in [9]. However, the method studied in [9] was designed for linear regression models with additive

Nicolas Dobigeon and Jean-Yves Tourneret are with IRIT/ENSEEIHT/TéSA, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France (e-mail: Nicolas.Dobigeon@enseeiht.fr; Jean-Yves.Tourneret@enseeiht.fr).

Jeffrey D. Scargle is with Space Science Division, NASA, Ames Research Center, Moffett Field, CA, USA (e-mail: Jeffrey.D.Scargle@nasa.gov).

Gaussian noise and cannot be applied directly to Poisson data. Our change-point detection strategy can be viewed as an adaptation of the Bayesian curve fitting estimator to Poisson data. Also the segmentation procedure studied in this paper allows joint segmentation of signals recorded by different sensors, contrary to the algorithm proposed in [9]. To our knowledge, this is the first fully Bayesian algorithm developed for joint segmentation of Poisson data.

This paper is organized as follows. The segmentation problem is formulated in Section II. Section III describes different elements of the hierarchical model that with be used to solve this segmentation problem. Section IV studies a Gibbs sampler for the posteriors of the unknown parameters to be estimated. Some simulation results on synthetic and real data are presented in Section V. Conclusions are reported in Section VI.

II. Problem Formulation

As explained in [12], the arrival times of photons can be modeled accurately by a discrete-time Poisson counting process. The numbers of photons counted in n successive equally spaced intervals (bins), from j different signals, are denoted $y_{j,i}$, where $y_{j,i}$ is the count in bin i of signal j ($i = 1, \ldots, n$ and $j = 1, \ldots, J$). The bins are grouped into blocks (intervals containing one or more bins). The summed counts of which are assumed to obey Poisson distributions whose parameters may or may not vary from one interval to another. Consequently, the statistical properties of such *binned data* can be defined as follows:

$$y_{j,i} \sim \mathcal{P}\left(\lambda_{j,k}\right),$$

where $j = 1, \ldots, J$, $k = 1, \ldots, K_j$, $i \in I_{j,k} = \{l_{j,k-1} + 1, \ldots, l_{j,k}\}$, and the following notations have been used:

- $\mathcal{P}(\lambda)$ denotes a Poisson distribution with parameter λ,
- J is the number of signals to be segmented,
- K_j is the number of segments in the j^{th} observed signal,
- $l_{j,k}$ is the sample point after which the k^{th} change occurs in the j^{th} signal (by convention $l_{j,0} = 0$ and $l_{j,K_j} = n$, where n is the number of observed samples). In other words, the actual change locations are $t_{j,k} = l_{j,k}T + \tau$ with $0 \leq \tau < T$, where T is the sampling period.

Moreover, the sequences $\mathbf{y}_l = \left[y_{l,1}, \ldots, y_{l,n}\right]$ and $\mathbf{y}_m = \left[y_{m,1}, \ldots, y_{m,n}\right]$ are assumed to be independent for $l \neq m$. Segmenting the astronomical time series $\mathbf{y}_j = \left[y_{j,1}, \ldots, y_{j,n}\right]$ jointly consists of estimating the change-points numbers K_j and their positions $l_{j,k}$ (for $j = 1, \ldots, J$ and $k = 1, \ldots, K_j$) from the observations $\mathbf{Y} = \left[\mathbf{y}_1, \ldots, \mathbf{y}_J\right]^\mathsf{T}$.

III. Hierarchical Bayesian Model

The unknown parameters for the segmentation problem (introduced in the previous section) are the numbers of segments K_j, the change-point locations $l_{j,k}$ and the Poisson parameters $\lambda_{j,k}$ (with $\boldsymbol{\lambda}_j = \left[\lambda_{j,1}, \ldots, \lambda_{j,K_j}\right]^\mathsf{T}$ and $\boldsymbol{\Lambda} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_J\}$). A standard reparameterization consists of introducing indicators $r_{j,i}$, $j \in \{1, \ldots, J\}$, $i \in \{1, \ldots, n\}$ such that:

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

with $r_{j,n} = 1$ (this condition ensures that the number of change-points and the number of steps of the j^{th} signal are equal to $K_j = \sum_{i=1}^{n} r_{j,i}$). The unknown parameter vector resulting from this reparameterization is $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J\}$ where $\boldsymbol{\theta}_j = \{\mathbf{r}_j, \boldsymbol{\lambda}_j\}$ and $\mathbf{r}_j = \left[r_{j,1}, \ldots, r_{j,n}\right]$. Note that the unknown parameter vector $\boldsymbol{\theta}$ belongs to a space $\Theta = \{0,1\}^{J \times n} \times \prod_{j=1}^{J} \mathbb{R}_+^{K_j}$, whose dimension depends on the parameters K_j, $j = 1, \ldots, J$. This paper proposes to estimate the unknown parameter vector $\boldsymbol{\theta}$ by using Bayesian estimation theory. Bayesian inference on $\boldsymbol{\theta}$ is based on the posterior distribution $f(\boldsymbol{\theta}|\mathbf{Y})$. This posterior distribution is related to the likelihood of the observations and the parameter priors via Bayes' theorem $f(\boldsymbol{\theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})$. The likelihood and priors for the segmentation problem are summarized below.

A. Likelihood

The likelihood of the observed vector $\mathbf{Y}$ can be expressed as follows:

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}) &= \prod_{j=1}^{J} \prod_{k=1}^{K_j} \prod_{i \in I_{j,k}} \frac{\lambda_{j,k}^{y_{j,i}} \exp\left(-\lambda_{j,k}\right)}{y_{j,i}!} \\ &= \frac{1}{\prod_{j=1}^{J} \prod_{i=1}^{n} y_{j,i}!} \prod_{j=1}^{J} \prod_{k=1}^{K_j} \lambda_{j,k}^{s_{j,k}(\mathbf{r}_j)} \exp\left(-\lambda_{j,k} n_{j,k}\left(\mathbf{r}_j\right)\right) \\ &\propto \prod_{j=1}^{J} \prod_{k=1}^{K_j} \lambda_{j,k}^{s_{j,k}(\mathbf{r}_j)} \exp\left(-\lambda_{j,k} n_{j,k}\left(\mathbf{r}_j\right)\right), \end{aligned} \tag{1}$$

where $\propto$ means "proportional to", $s_{j,k}\left(\mathbf{r}_j\right) = \sum_{i \in I_{j,k}} y_{j,i}$ and $n_{j,k}\left(\mathbf{r}_j\right) = l_{j,k} - l_{j,k-1}$ (number of samples in the k^{th} interval $I_{j,k}$ of the j^{th} signal).

B. Parameter Priors

1) Indicator Vector: The indicator vectors $\mathbf{R}_i = \left[r_{1,i}, \ldots, r_{J,i}\right]^\mathsf{T}$ and $\mathbf{R}_{i'} = \left[r_{1,i'}, \ldots, r_{J,i'}\right]^\mathsf{T}$ are assumed to be independent for any $i \neq i'$. Therefore, the prior distribution of $\mathbf{R} = \left[\mathbf{R}_1, \ldots, \mathbf{R}_n\right]$ can be decomposed as follows:

$$f(\mathbf{R}) = \prod_{i=1}^{n-1} f(\mathbf{R}_i). \tag{2}$$

The possible correlations between the change locations in the different observed signals are adjusted by choosing an appropriate prior distribution $f(\mathbf{R}|\mathbf{P})$. We assume that the probability of having $\left[r_{1,i}, \ldots, r_{J,i}\right]^\mathsf{T} = \boldsymbol{\epsilon}$ does not depend on i (with $\boldsymbol{\epsilon} \in \mathcal{E} = \{0,1\}^J$) and is denoted $\mathrm{P}_{\boldsymbol{\epsilon}}$. As a consequence, the indicator prior distribution is:

$$f(\mathbf{R}|\mathbf{P}) = \prod_{\boldsymbol{\epsilon} \in \mathcal{E}} \mathrm{P}_{\boldsymbol{\epsilon}}^{S_{\boldsymbol{\epsilon}}(\mathbf{R})}, \tag{3}$$

where $\mathbf{P} = \{\mathrm{P}_{\boldsymbol{\epsilon}}\}_{\boldsymbol{\epsilon} \in \mathcal{E}}$, $\mathrm{P}_{\boldsymbol{\epsilon}} \in \{\mathrm{P}_{0\ldots0}, \ldots, \mathrm{P}_{1\ldots1}\}$ and $S_{\boldsymbol{\epsilon}}(\mathbf{R})$ is the number of lags such that $\left[r_{1,i}, \ldots, r_{J,i}\right]^\mathsf{T} = \boldsymbol{\epsilon}$. The most likely configuration $\left[r_{1,i}, \ldots, r_{J,i}\right]^\mathsf{T} = \boldsymbol{\epsilon}$ will correspond to the highest value of $\mathrm{P}_{\boldsymbol{\epsilon}}$. For instance, by choosing high values of $\mathrm{P}_{0\ldots0}$ (resp. $\mathrm{P}_{1\ldots1}$), the absence (resp. presence) of simultaneous changes in all observed signals will be favored. This choice induces correlation between change-point locations in the different time series.

2) Poisson Parameters: By assuming the parameters $\lambda_{j,k}$ are *a priori* independent, the prior distribution for $\mathbf{\Lambda} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_J\}$ is:

$$f(\mathbf{\Lambda}|\gamma) = \prod_{j=1}^{J}\prod_{k=1}^{K_j} f(\lambda_{j,k}|\nu,\gamma).$$

Gamma distributions are assigned to these Poisson parameters:

$$\lambda_{j,k}|\nu,\gamma \sim \mathcal{G}(\nu,\gamma), \tag{4}$$

where $\nu = 1$ (as in [9]), γ is an adjustable hyperparameter and $\mathcal{G}(a,b)$ denotes the Gamma distribution with parameters a and b. The variety of distributions available when γ varies indicates it is possible to incorporate either vague or more specific prior information about the parameters $\lambda_{j,k}$. Moreover, the Gamma distribution is the conjugate prior for the parameters $\lambda_{j,k}$, which allows us to integrate out these parameters from the joint posterior (see section III-D). It would be possible to define a set of hyperparameters $\gamma_j, j = 1, \ldots, J$ for signals whose amplitudes differ significantly; however, such situations are not considered in this paper. The previous assumptions yield the following prior distribution for $\mathbf{\Lambda}$:

$$\begin{aligned} f(\mathbf{\Lambda}|\gamma) &= \prod_{j=1}^{J}\prod_{k=1}^{K_j} f(\lambda_{j,k}|\nu,\gamma) \\ &= \prod_{j=1}^{J}\prod_{k=1}^{K_j} \frac{\gamma^{\nu}}{\Gamma(\nu)}\lambda_{j,k}^{\nu-1}e^{-\gamma\lambda_{j,k}}\mathbb{I}_{\mathbb{R}^+}(\lambda_{j,k}) \\ &= \prod_{j=1}^{J}\left[\frac{\gamma^{\nu K_j}e^{-\gamma\sum_{k=1}^{K_j}\lambda_{j,k}}}{\Gamma(\nu)^{K_j}}\prod_{k=1}^{K_j}\left(\lambda_{j,k}^{\nu-1}\mathbb{I}_{\mathbb{R}^+}(\lambda_{j,k})\right)\right], \end{aligned}$$

where $\mathbb{I}_{\mathbb{R}^+}(x)$ is the indicator function defined on $\mathbb{R}^+$ (i.e. $\mathbb{I}_{\mathbb{R}^+}(x) = 1$ if $x \geq 0$ and $\mathbb{I}_{\mathbb{R}^+}(x) = 0$ otherwise).
The hyperparameter vector associated with the priors defined above is $\mathbf{\Phi} = (\mathbf{P}, \gamma)$. Of course, the quality of the Bayesian segmentation depends on the values of the hyperparameters. In particular applications, these hyperparameters can be fixed from available information regarding the observed signals as in [14]. However, to increase the robustness of the algorithm, hyperparameters can be considered as random variables with noninformative priors as in [9]. This strategy, involving different levels in a Bayesian prior hierarchy, results in so-called *hierarchical Bayesian models*. Such models require that one define hyperparameter priors (sometimes referred to as *hyper-priors*), as detailed in the next section.

C. Hyperparameter Priors

1) Hyperparameter γ: The prior distribution for γ is a noninformative Jeffreys' prior (as in [9]), which reflects the absence of knowledge regarding this hyperparameter:

$$f(\gamma) = \frac{1}{\gamma}\mathbb{I}_{\mathbb{R}^+}(\gamma). \tag{5}$$

2) Hyperparameter $\mathbf{P}$: The Dirichlet distribution is the usual prior for positive parameters summing to 1. It has the nice property of providing a vague or informative prior depending on its parameter values. It also allows us to integrate out the parameters P_{ϵ} from the joint posterior. This paper assumes that the prior distribution for $\mathbf{P}$ is a Dirichlet distribution with parameter vector $\boldsymbol{\alpha} = \left[\alpha_{0\ldots0}, \ldots, \alpha_{1\ldots1}\right]^{\mathsf{T}}$ denoted as:

$$\mathbf{P}|\boldsymbol{\alpha} \sim \mathcal{D}_{2^J}(\boldsymbol{\alpha}). \tag{6}$$

This distribution is defined on the simplex $\mathcal{P} = \left\{\mathbf{P}; \sum_{\boldsymbol{\epsilon}\in\mathcal{E}} \mathrm{P}_{\boldsymbol{\epsilon}} = 1, \mathrm{P}_{\boldsymbol{\epsilon}} > 0\right\}$.

Assuming that the different hyperparameters are *a priori* independent, the prior distribution for the hyperparameter vector $\mathbf{\Phi}$ can be written as follows:

$$f(\mathbf{\Phi}|\boldsymbol{\alpha}) \propto \left(\prod_{\boldsymbol{\epsilon}\in\mathcal{E}} \mathrm{P}_{\boldsymbol{\epsilon}}^{\alpha_{\boldsymbol{\epsilon}}-1}\right)\frac{1}{\gamma}\mathbb{I}_{\mathbb{R}^+}(\gamma)\mathbb{I}_{\mathcal{P}}(\mathbf{P}), \tag{7}$$

where $\alpha_{\boldsymbol{\epsilon}} \in \{\alpha_{0\ldots0}, \ldots, \alpha_{1\ldots1}\}$. This paper has assumed all values of $\alpha_{\boldsymbol{\epsilon}}$ are equal. In this case, the Dirichlet distribution reduces to the uniform distribution on $\mathcal{P}$.

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

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

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

where $f(\mathbf{Y}|\boldsymbol{\theta})$ and $f(\mathbf{\Phi})$ have been defined in (1) and (7). This hierarchical structure is shown on the the directed acyclic graph (DAG) of Fig. 1. It allows one to integrate out the nuisance parameters $\mathbf{\Lambda}$ and $\mathbf{P}$ from the joint distribution $f(\boldsymbol{\theta}, \mathbf{\Phi}|\mathbf{Y})$, yielding:

$$\frac{f(\mathbf{R}, \gamma|\mathbf{Y})}{C(\mathbf{R}|\mathbf{Y})} \propto \frac{1}{\gamma}\prod_{j=1}^{J}\left[\left(\frac{\gamma^{\nu}}{\Gamma(\nu)}\right)^{K_j}\prod_{k=1}^{K_j}\frac{\Gamma(s_{j,k}+\nu)}{(n_{j,k}+\gamma)^{s_{j,k}+\nu}}\right]\mathbb{I}_{\mathbb{R}^+}(\gamma), \tag{8}$$

with

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

and $\Gamma(t)$ is the Gamma function (i.e. $\Gamma(t) = \int_0^{+\infty} u^{t-1}e^{-u}du, \quad t > 0$).
The posterior distribution (8) is too complex to obtain closed-form expressions of the Bayesian estimators for the unknown parameters (such as the minimum mean square error (MMSE) estimator or the maximum *a posteriori* (MAP) estimator). In this case, it is quite common to apply MCMC methods to generate samples that are asymptotically distributed according to the posteriors of interest. The samples can then be used to estimate the unknown parameters. This paper studies a Gibbs sampling strategy similar to the segmenter in [9]. However, it is important to note the following differences. First, our proposed Gibbs sampling strategy does not involve reversible jumps. Indeed, due to the reparametrization introduced in the beginning of section III, the unknown parameters in (8) belong

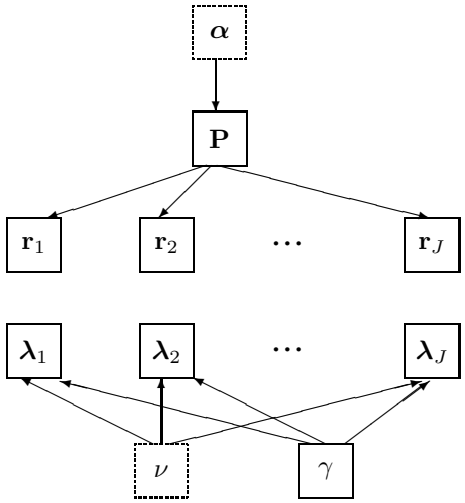

Fig. 1. DAG for the prior distributions; the fixed parameters appear dashed boxes.

to a space with a fixed dimension. Second, the proposed algorithm allows joint segmentation of multiple time-series.

IV. Gibbs Sampler for Change-point Detection

The Gibbs sampler is an iterative sampling strategy for generating samples distributed according to the full conditional distributions of each parameter. This paper proposes to sample according to the distribution $f(\mathbf{R}, \gamma|\mathbf{Y})$ defined in (8). The main steps of the algorithm are summarized in the table "Algorithm 1" and are detailed in Subsections IV-A to IV-C.

Algorithm 1: Gibbs sampling algorithm for joint segmentation

- Initialization:
  - Sample hyperparameter $\widetilde{\gamma}^0$ from the pdf in (5),
  - Sample hyperparameter $\widetilde{\mathbf{P}}^{(0)}$ from the pdf in (6),
  - For $i = 1, \ldots, n-1$ sample, $\left[\widetilde{\mathrm{r}}_{1,i}^{(0)}, \ldots, \widetilde{\mathrm{r}}_{J,i}^{(0)}\right]^{\mathsf{T}}$ from the pdf in (3),
  - For $j = 1, \ldots, J$, $k = 1, \ldots, K$, sample $\widetilde{\lambda}_{j,k}^{(0)}$ from the pdf's in (4),
  - Set $t \leftarrow 1$,
- Iterations: for $t = 1, 2, 3, \ldots,$ do
  - For $i = 1, \ldots, n-1$, sample $\left[\widetilde{\mathrm{r}}_{1,i}^{(t)}, \ldots, \widetilde{\mathrm{r}}_{J,i}^{(t)}\right]^{\mathsf{T}}$ according to the probabilities (9),
  - For $j = 1, \ldots, J$, $k = 1, \ldots, K$, sample $\widetilde{\lambda}_{j,k}^{(t)}$ from the pdf's in (10),
  - Sample $\widetilde{\gamma}^{(t)}$ from the pdf in (11),
  - *Optional step:* sample $\widetilde{\mathbf{P}}^{(t)}$ from the pdf in (12),
  - Set $t \leftarrow t+1$.

A. Generation of Samples Distributed According to $f(\mathbf{R}|\gamma, \mathbf{Y})$

This generation is achieved by using the Gibbs Sampler to draw $(n-1)$ samples distributed according to $f(r_{1,i}, \ldots, r_{J,i}|\gamma, \mathbf{Y})$. This random variable is discrete and takes its values in $\mathcal{E} = \{0,1\}^J$. Consequently, its distribution is fully characterized by the probabilities $P(\left[r_{1,i}, \ldots, r_{J,i}\right]^{\mathsf{T}} = \boldsymbol{\epsilon}|\gamma, \mathbf{Y}), \boldsymbol{\epsilon} \in \mathcal{E}$. By using the notations $\mathbf{R}_{-i}$ to denote the matrix $\mathbf{R}$ where the column at time i is supressed, the following result can be obtained:

$$P\left(\left[r_{1,i}, \ldots, r_{J,i}\right]^{\mathsf{T}} = \boldsymbol{\epsilon}|\mathbf{R}_{-i}, \gamma, \mathbf{Y}\right) \propto f\left(\mathbf{R}_i(\boldsymbol{\epsilon}), \gamma|\mathbf{Y}\right), \quad (9)$$

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

B. Generation of Samples Distributed According to $f(\gamma|\mathbf{R}, \mathbf{Y})$

To obtain samples distributed according to $f(\gamma|\mathbf{R}, \mathbf{Y})$, it is convenient to simulate vectors distributed according to the joint distribution distribution $f(\gamma, \boldsymbol{\Lambda}|\mathbf{R}, \mathbf{Y})$ by using Gibbs moves. This step can be decomposed as follows:

- **Draw samples according to** $f(\boldsymbol{\Lambda}|\mathbf{R}, \gamma, \mathbf{Y})$
  Looking carefully at the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y})$, the following result can be obtained:
  $$\lambda_{j,k}|\mathbf{R}, \gamma, \mathbf{Y} \sim \mathcal{G}\left(s_{j,k}\left(\mathbf{r}_j\right) + \nu, n_{j,k}\left(\mathbf{r}_j\right) + \gamma\right). \quad (10)$$
- **Draw samples according to** $f(\gamma|\mathbf{R}, \boldsymbol{\Lambda}, \mathbf{Y})$
  This is achieved as follows:
  $$\gamma|\mathbf{R}, \boldsymbol{\Lambda} \sim \mathcal{G}\left(\nu \sum_{j=1}^{J} K_j, \sum_{j=1}^{J}\sum_{k=1}^{K_j} \lambda_{j,k}\right). \quad (11)$$

C. Posterior Distribution of Hyperparameter $\mathbf{P}$

The hyperparameter $\mathbf{P}$ carries information regarding the probability of having simultaneous changes at a given location. As a consequence, its estimation may be interesting in practical applications. The posterior distribution of this parameter, conditioned upon the indicator vector $\mathbf{R}$, the vector of observed samples $\mathbf{Y}$ and the parameters $\boldsymbol{\alpha}$, can be easily derived. This is a 2^J-Dirichlet distribution with parameters $(\alpha_{\boldsymbol{\epsilon}} + S_{\boldsymbol{\epsilon}}(\mathbf{R}))_{\boldsymbol{\epsilon}\in\mathcal{E}}$:

$$\mathbf{P}|\mathbf{R}, \mathbf{Y}, \boldsymbol{\alpha} \sim \mathcal{D}_{2^J}(\alpha_{\boldsymbol{\epsilon}} + S_{\boldsymbol{\epsilon}}(\mathbf{R})). \quad (12)$$

V. Simulations

A. Synthetic Data

Many simulations have been conducted to validate the segmentation algorithm (cf. Section IV). The simulations presented in this section have been obtained for $J = 2$ signals of $n = 120$ samples. The change-point locations for these two sequences are $\boldsymbol{l}_1 = (20, 50, 100, 120)$ and $\boldsymbol{l}_2 = (50, 120)$. The parameters of the Poisson distributions are $\boldsymbol{\lambda}_1 = \left[19, 9, 16, 6\right]^{\mathsf{T}}$ and $\boldsymbol{\lambda}_2 = \left[8, 11\right]^{\mathsf{T}}$. The hyperparameters have been set to $\nu = 2$ and $\alpha_\epsilon = \alpha = 1, \forall\epsilon$. The hyperparameters α_ϵ are equal to insure the Dirichlet distribution reduces to a uniform distribution. Moreover, the common value of the hyperparameters α_ϵ has been set to $\alpha = 1 \ll n$ to reduce the influence of this parameter on the posterior (12). All figures have been obtained after averaging the results of 64 Markov chains. The total number of runs for each Markov chain is $N_{MC} = 1000$, including $N_{bi} = 200$ burn-in iterations. Thus, only the last 800 Markov chain output samples are used in

computing the estimates. The results provided by the joint segmentation procedure are compared with those provided by two 1D segmentations (which consists of performing the proposed algorithm on each of the two sequences). Note that running 100 iterations of the proposed algorithm for joint segmentation takes approximately 30 seconds for a MATLAB implementation on a 2.8Ghz Pentium IV.

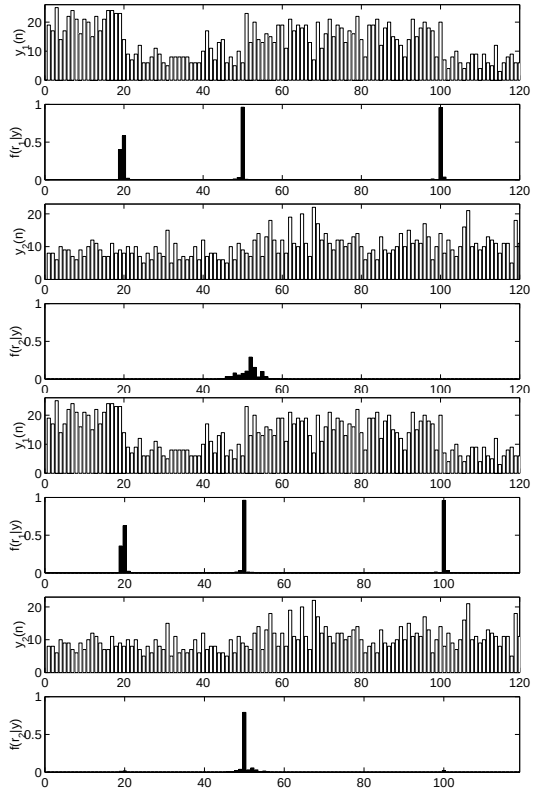

Fig. 2. Posterior distributions of the change-point locations for 1D (top) and joint segmentations (bottom).

1) Posterior Distribution of the Change-point Locations: The MMSE estimates of the change locations are depicted in Figure 2 for 1D and joint approaches. These estimates have been computed as follows:

$$\hat{\mathbf{R}}_{\text{MMSE}} = \frac{1}{N_r} \sum_{t=1}^{N_r} \mathbf{R}^{(N_{bi}+t)}, \tag{13}$$

where N_{bi} is the number of burn-in iterations. Note that the mean of a matrix $\mathbf{R}$ is distributed according to $f(\mathbf{R}|\mathbf{Y})$ and provides the *a posteriori* probabilities for changes at the different lags and signals (since $r_{j,i}$ is a binary random variable). For example, there is a very high posterior probability that a change occurred at lags $i = 50$ and $i = 100$ in the first sequence (with both 1D and 2D approaches). However, the two methods seem hesitant to locate the first change (at lag $i = 20$) in the first signal. The advantage of the joint segmentation procedure is illustrated by the bottom of Figure 2: the change at lag $i = 50$ in the second sequence is more clearly estimated by the joint approach (right figure) than by the 1D approach (left figure). Thus, in this example, the joint 2D segmentation procedure provides better results than two independent 1D segmentations.

2) Posterior Distribution of (K_1, K_2)*:* The proposed algorithm generates samples $\left(\mathbf{R}^{(t)}, \gamma^{(t)}\right)$ distributed according to the posterior distribution $f(\mathbf{R}, \gamma|\mathbf{Y})$, which allows for model selection. Indeed, the change-point number in each sequence can be estimated by $K_j^{(t)} = \sum_{i=1}^{n} r_{j,i}^{(t)}$. Figure 3 shows the estimated posteriors of K_j in each sequence (computed from the last 800 Markov chain samples) for the 1D and joint segmentation algorithms. The maximum values of these posteriors provide the MAP estimates of the changepoint numbers $\widehat{K}_1 = 4$ and $\widehat{K}_2 = 2$, which corresponds to the actual numbers of changes. (Remember that there is a change at lag $n = 120$ in all signals, by convention.)

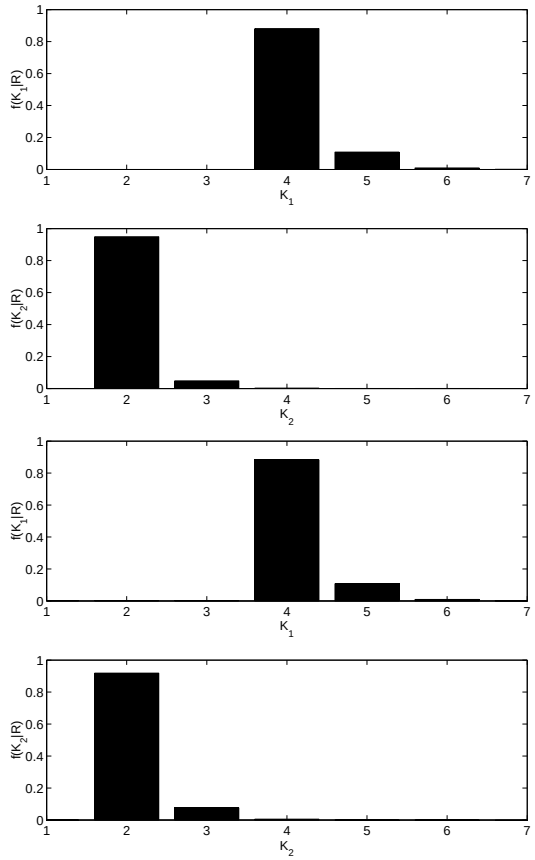

Fig. 3. Posterior distributions of K_1 and K_2 for 1D (top) and joint segmentations (bottom).

3) Poisson Parameter Estimation: The estimation of the Poisson parameters is interesting since it allows for signal reconstruction. The posterior distributions of the parameters $\lambda_{1,k}$ and $\lambda_{2,k}$, conditioned upon $K_1 = 4$ and $K_2 = 2$, are depicted in Figures 4 and 5. They are clearly in good agreement with the actual values of the parameters, $\boldsymbol{\lambda}_1 = \left[19, 9, 16, 6\right]^{\mathsf{T}}$ and $\boldsymbol{\lambda}_2 = \left[8, 11\right]^{\mathsf{T}}$.

4) Hyperparameter Estimation: The posterior distributions of $\mathrm{P}_{\boldsymbol{\epsilon}}$ are depicted in Figure 6. They are clearly in agreement with the actual posterior distributions given by the Dirichlet distribution $\mathcal{D}_4\left(n-3, 1, 3, 2\right)$ defined in (12).

5) Sampler Convergence: A crucial issue when using MCMC methods is convergence assessment. The Gibbs sampler allows us to draw samples $\mathbf{R}^{(t)}$ asymptotically distributed according to $f(\mathbf{R}|\gamma, \mathbf{Y})$. The change-point posterior

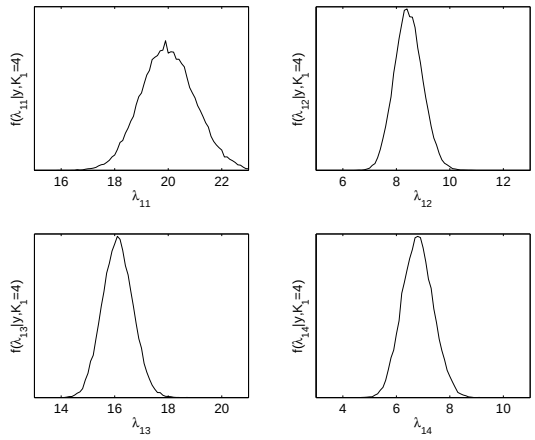

Fig. 4. Posterior distributions of the Poisson parameters $\lambda_{1,i}$ (for $i = 1, \ldots, 4$) conditioned on $K_1 = 4$.

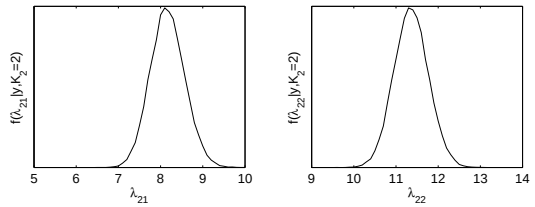

Fig. 5. Posterior distributions of the Poisson parameters $\lambda_{2,i}$ (for $i = 1, 2$) conditioned on $K_2 = 2$.

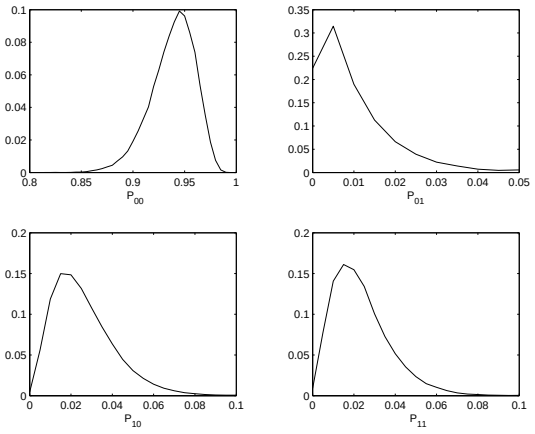

Fig. 6. Posterior distributions of the hyperparameters P_{00}, P_{01}, P_{10} and P_{11}.

probabilities are then be estimated by (13). However, two important questions have to be answered: 1) When can we decide whether the simulated samples $\left\{\mathbf{R}^{(t)}\right\}$ are distributed according to the target distribution? 2) How many samples do we need to obtain an accurate estimate of $\mathbf{R}$ when using (13)? This section presents some details about determining appropriate values for parameters N_r and N_{bi}.

A first *ad hoc* approach consists of assessing convergence via appropriate graphical evaluations [15, p. 28]. Here, a reference estimate denoted as $\widetilde{\mathbf{R}}$ has been computed for a large number of iterations $\widetilde{N_r} = 10000$ and $\widetilde{N_{bi}} = 10000$ (to ensure convergence of the sampler and good accuracy of the approximation (13)). Figure 7 shows the mean square error (MSE) between this reference estimate $\widetilde{\mathbf{R}}$ and the estimate obtained after $N_r = p$ runs and $\widetilde{N_{bi}} = 10000$ burn-in iterations:

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

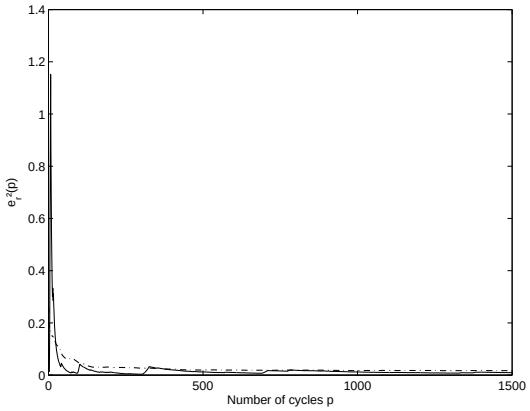

Fig. 7. MSE between the reference and estimated *a posteriori* change-point probabilities versus p (solid line). Averaged MSE computed from 64 chains (dotted line) ($N_{bi} = 200$).

Figure 7 shows that a number of runs equal to $N_r = 800$ is sufficient to ensure an accurate estimation of the empirical average (13). Similarly, the MSE e_r^2 versus the number of burn in iterations N_{bi} (for a fixed number of iterations $p = 800$) is depicted in Figure 8. This figure indicates a short burn-in period is sufficient. We have chosen $N_{bi} = 200$ for the next simulations.

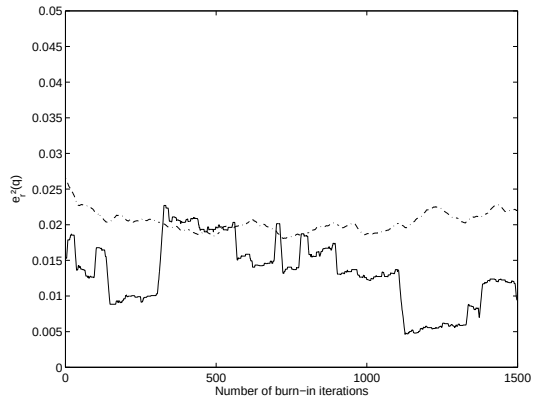

Fig. 8. MSE between the reference and estimated *a posteriori* change-point probabilities versus N_{bi} (solid line). Averaged MSE computed from 64 chains (dotted line) ($N_r = 800$).

Running multiple chains with different over-dispersed initializations allows us to define various convergence measures for MCMC methods [15]. We use the popular "between-within variance criterion" to confirm our previous convergence diagnosis. This method was initially studied by Gelman and Rubin in [16] and has been often used to monitor convergence (see, for example, [17], [18] or [15, p. 33]). This criterion requires running M parallel chains of length N_r with different

starting values. The between-sequence variance B and within-sequence variance W for the M Markov chains are defined by

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

and

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

with

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

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 [16]:

$$\sqrt{\hat{\rho}} = \sqrt{\frac{N_r-1}{N_r} + \frac{M+1}{MN_r}\frac{B}{W}}. \tag{14}$$

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

Different choices for parameter κ could be considered for our proposed joint segmentation procedure. We propose to monitor the convergence of the Gibbs sampler via 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 Figure 9. The chains clearly converge to similar values. The potential scale reduction factors for all parameters $\mathrm{P}_{\boldsymbol{\epsilon}}$ are given in Table I. These values of $\sqrt{\hat{\rho}}$ confirm the good convergence of the sampler (a value of $\sqrt{\hat{\rho}}$ below 1.2 is recommended in [19, p. 332]).

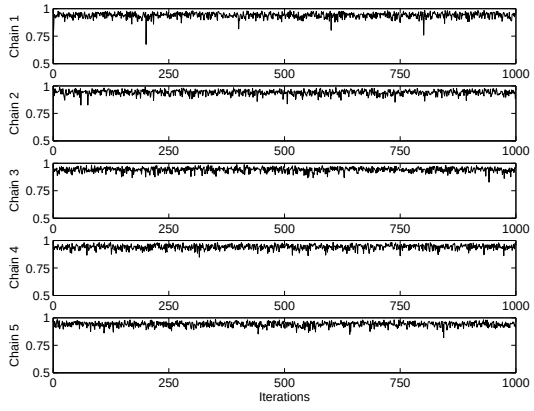

Fig. 9. Convergence assessment with five different Markov chains.

TABLE I
POTENTIAL SCALE REDUCTION FACTORS OF $\mathrm{P}_{\boldsymbol{\epsilon}}$ (COMPUTED FROM $M = 5$ MARKOV CHAINS)

$\mathrm{P}_{\boldsymbol{\epsilon}}$	$\sqrt{\hat{\rho}}$	$\mathrm{P}_{\boldsymbol{\epsilon}}$	$\sqrt{\hat{\rho}}$	$\mathrm{P}_{\boldsymbol{\epsilon}}$	$\sqrt{\hat{\rho}}$	$\mathrm{P}_{\boldsymbol{\epsilon}}$	$\sqrt{\hat{\rho}}$
P_{00}	0.9995	P_{01}	1.0000	P_{10}	0.9998	P_{11}	1.0001

B. Real Astronomical Data

1) 1D Data: This section presents the analysis of a small sample of data obtained by the NASA Compton Gamma Ray Observatory's BATSE (Burst and Transient Source Experiment) [20]. By the nature of this photon-counting experiment, the time series can be accurately modeled as Poisson processes. The Poisson rate parameter varies as determined by the actual changes in brightness of the gamma-ray burst (GRB) source. The only significant departure from this picture is that the recorded photons are not quite independent, due to a small *dead time* in the detectors.

The intensity of the GRB as a function of time often consists of a series of short-time-scale structures, called pulses. The goal of the analysis is to determine parameters such as the rise and decay times of the pulses, and other quantities that can be derived from a piecewise-constant representation.

The hierarchical method presented in this paper has been applied to the astronomical 1D-data studied in [12]. The raw counting data (which consists of about 29000 photons) have been transformed into binned data by counting the number of photons distributed in 256 time bins of width $3.68ms$. Note that $J = 1$ in this example. The results have been averaged from 64 Markov chains with $N_{MC} = 1550$ runs and $N_{bi} = 50$ burn-in iterations. These values of N_{MC} and N_{bi} have been chosen in order to obtain appropriate values of the potential scalar reduction factor for parameters P_0 and P_1 (see the end of this section).

The first step of the analysis consists of estimating the posterior distribution of $\mathbf{R}$ for the observed sequence plotted in Figure 11(a). This estimated posterior distribution is depicted in Figure 11(b).

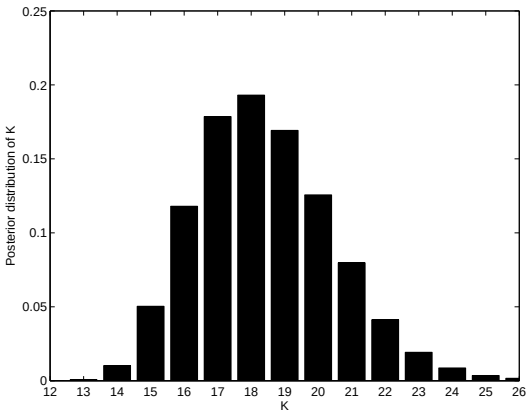

Fig. 10. Posterior distribution of the change-point number (1D astronomical data).

In the second step of the analysis, we estimate the number of change-points for the observed sequence. The posterior of the number of changes (computed from the last 1500 Markov chain output samples) is depicted in Figure 10. The corresponding MAP estimator is $\widehat{K} = 18$.

In the last step of the analysis, the different Poisson intensities are estimated for each segment from the change locations. More precisely, segments are obtained from the 18 largest values of the posterior depicted in Figure 11(b). The MMSE

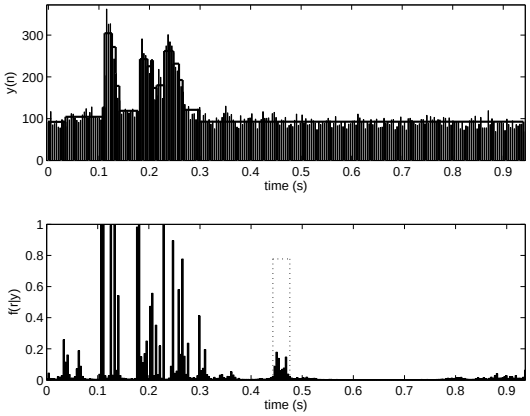

Fig. 11. (a) Original and segmented data, (b) Posterior distribution of the change-point locations (1D astronomical data).

Poisson estimates are then obtained by averaging the signal on each segment (which corresponds to the intensity MMSE estimator conditioned on $K = 18$). This procedure yields Bayesian blocks, which are introduced in [12]. Figure 11(a) shows Bayesian blocks obtained after keeping $K = 18$ segments, as suggested by Figure 10. It is also possible to compute the probability of having changes within a given interval. For instance, the probability of having at least one change-point in the interval $\left[0.44; 0.47\right]$ appears in dotted lines in Figure 11(b). This high value could induce a modified segmentation including a change in this interval $\left[0.44; 0.47\right]$. These results are in good agreement with those of [12].

The convergence of the Gibbs sampler for segmenting the real astronomical data of Figure 11(a) has been studied. The potential scalar reduction factors for parameters P_0 and P_1 (obtained from 5 parallel chains and $N_r = 1500$) are both equal to $\sqrt{\hat{\rho}} = 0.9996$. The convergence criterion $\sqrt{\hat{\rho}} < 1.2$ (see [19, p. 332]) is satisfied for this example. Note again that our segmentation procedure does not require any stopping rule, other than what is implicit in the assessment of the Markov chain's convergence.

TABLE II

POTENTIAL SCALE REDUCTION FACTORS OF $\mathrm{P}_{\boldsymbol{\epsilon}}$ (COMPUTED FROM $M = 5$ MARKOV CHAINS)

P_ϵ	$\sqrt{\hat{\rho}}$	P_ϵ	$\sqrt{\hat{\rho}}$
P_{0000}	0.9999	P_{0100}	1.0008
P_{0001}	1.0004	P_{0101}	1.0015
P_{0010}	0.9999	P_{0110}	1.0002
P_{0011}	1.0000	P_{0111}	1.0013
P_{1000}	1.0013	P_{1100}	1.0021
P_{1001}	1.0000	P_{1101}	1.0000
P_{1010}	0.9999	P_{1110}	0.9999
P_{1011}	0.9998	P_{1111}	0.9999

2) Multidimensional Data: The dependence of the GRB variability on the energy of the radiation is of considerable interest. In the data mode analyzed here, BATSE recorded the energies of the photons in four energy channels, which are analogous to four colors in ordinary visible radiation. The unit of energy is keV (thousand electron volts), and the nominal energy channels are $25-60keV$, $60-110keV$, $110-325keV$, and $> 325keV$. The variability curves at low and high energies are typically very similar, but there can be a delay or lag between them.

The thousands of recorded GRB light curves form an extremely heterogeneous collection. As with snowflakes, no two are alike. They range in duration from a few tens of milliseconds to a few hundred seconds. Their shapes range from simple rise-and-fall forms to complex multiple-pulse structures. The study of these objects is still much in the exploratory phase, and the kind of multivariate analysis described here is an important part of the exploration.

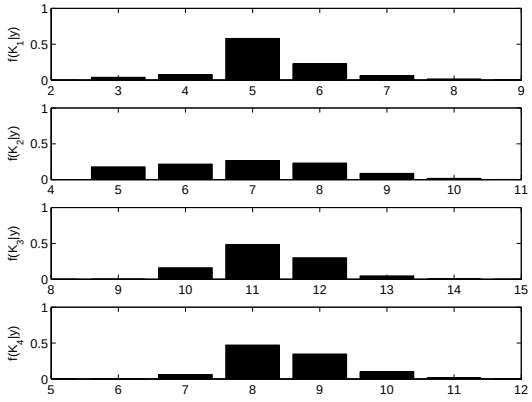

Fig. 12. Posterior distribution of the change-point number (4D astronomical data).

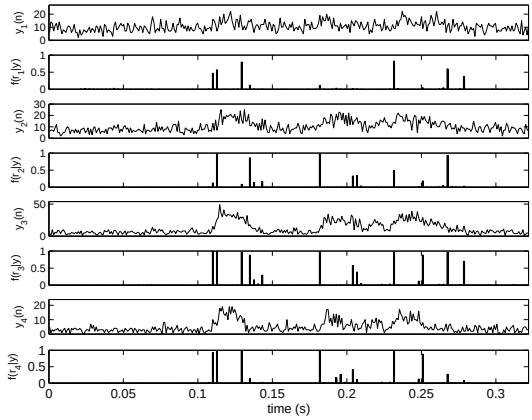

Fig. 13. Posterior distribution of the change-point locations (4D astronomical data).

The observed data corresponding to the four channels have been processed by the proposed joint segmentation algorithm. The estimated number of change-points and their positions are obtained after 3500 iterations including a burn-in period of 200 runs. Figure 12 shows that the MAP estimates of parameters K_j are $\widehat{K_1} = 5$, $\widehat{K_2} = 7$, $\widehat{K_3} = 11$ and $\widehat{K_4} = 8$. The estimated posterior distribution of $\mathbf{R}$ depicted in Figure 13 can then be used to estimate the change locations in each channel (as explained in the previous section). The resulting Bayesian blocks are shown in Figure 14. Note that the time scales are not the same for figures 13 and 14 (i.e. [0, 0.33s] for figure 13

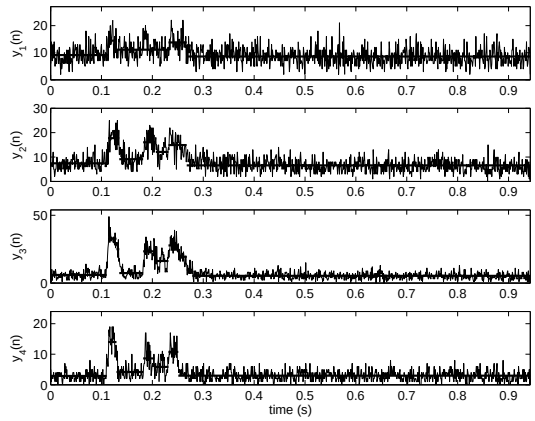

Fig. 14. Block representation (4D astronomical data).

whereas it is [0, 0.94s] for figure 14) for clarity. Most results are in good agreement with those presented in [21]. However, the proposed joint approach makes it possible to find out changes that were not initially detected by the iterative method. For example, the second and third changepoints $l_{1,2}$ and $l_{1,3}$ in the first channel (respectively at $0.1294s$ and $0.2316s$) are detected by the joint approach and not by the 1D approach. The presence of changes at the same position in the other channels explains this detection.

The convergence of the Gibbs sampler for the joint segmentation of the real astronomical data of figure 13 has been studied. The potential scalar reduction factors for parameters $\mathrm{P}_{\boldsymbol{\epsilon}}$, $\boldsymbol{\epsilon} \in \{0,1\}^4$ are provided in Table II. The convergence criterion $\sqrt{\hat{\rho}} < 1.2$ is satisfied for this example.

VI. Conclusions

This paper studied Bayesian sampling algorithms for segmenting single and multiple time series obeying Poisson distributions with piecewise constant parameters. Posterior distributions of the unknown parameters gave estimates of the unknown parameters and their uncertainties. Simulation results conducted on synthetic and real signals illustrated the performance of the proposed methodologies.

One of the two most important aspects of this work is its treatment of possible relationships between the observed times series. In many scientific areas, astronomy in particular, one has incomplete knowledge ahead of time – indeed, the main goal of the data analysis is typically to uncover such interrelationships. On the other hand, one typically has some information, for example, that the different time series are more or less similar. This kind of vague but important knowledge is naturally expressed in a Bayesian context by the prior distribution adopted for the models and their parameters.

The second important aspect is that information about the uncertainties of the parameter estimates arises as a matter of course from the sampling strategy. This is typical of Markov chain Monte Carlo methods but, for example, the methods in [12], [13] do not explore the relevant parameter space and provide no direct variance estimation.

Finally, it is interesting to note that the hierarchical Bayesian algorithm developed in this paper could be modified to handle other data sets including time-tagged event (TTE) data and time-to-spill (TTS) data (see [12] for more details). This study is currently under investigation.

Acknowledgements

This work was supported by the CNRS under MathSTIC Action No. 80/0244. Jeffrey Scargle would like to thank the NASA Applied Information Systems Research Program and Joseph Bredekamp, and the kind hospitality of the Institute for Pure and Applied Mathematics. The authors also thank David van Dyk for helpful comments.

References

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

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

[3] M. Lavielle and E. Moulines, "Least squares estimation of an unknown number of shifts in a time series," *Jour. of Time Series Anal.*, vol. 21, pp. 33–59, Jan. 2000.

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

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

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

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

[8] J.-Y. Tourneret, M. Doisy, and M. Lavielle, "Bayesian retrospective detection of multiple changepoints corrupted by multiplicative noise. application to SAR image edge detection," *Signal Processing*, vol. 83, no. 9, pp. 1871–1887, Sept. 2003.

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

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

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

[12] J. D. Scargle, "Studies in astronomical time series analysis: v. Bayesian blocks, a new method to analyze structure in photon counting data," *The Astrophysical Journal*, vol. 504, pp. 405–418, Sept. 1998.

[13] B. Jackson, J. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai, "An algorithm for optimal partitioning of data on an interval," *IEEE Signal Processing Lett.*, vol. 12, pp. 105–108, Feb. 2005.

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

[15] 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.

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

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

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

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

[20] W. S. Paciesas *et al.*, "The fourth BATSE burst revised catalog," *Astrophysical Journal Supplement Series*, vol. 122, pp. 465–497, June 1999, the data are available at `http://cossc.gsfc.nasa.gov/batse/`.

[21] J. D. Scargle, J. Norris, and B. Jackson, "Studies in astronomical time series analysis: vi. optimal segmentation : blocks, histograms and triggers," to be submitted.

Nicolas Dobigeon 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, both in June 2004. He is currently pursuing the Ph.D. degree with the Signal and Communication group at the IRIT Laboratory, Toulouse. 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 received the ingénieur degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique in Toulouse (ENSEEIHT). He received the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992. He is currently a professor in the university of Toulouse, France (ENSEEIHT). 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 non-stationary processes. He was the program chair of the European conference on signal processing (EUSIPCO), which was held in Toulouse (France) in 2002. He is also member of the ICASSP'06 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.

Jeffrey D. Scargle graduated from Pomona College, Summa Cum Laude, and received a Ph.D. in astrophysics from the California Institute of Technology. Following a series of positions at Lick Observatory and the University of California, at Berkley and Santa Cruz, he is now a research astrophysicist in the Astrobiology and Space Science Division at the U.S. National Aeronautics and Space Administration, at the Ames Research Center. His research centers on the study of variability of astronomical objects, especially at high energies – gamma-rays and x-rays. He has developed a number of techniques for analysis of time series data, including a widely used periodogram for unevenly sampled data, time domain modeling techniques, and most recently a general approach to optimal partitioning of data in spaces of arbitrary dimension.