# CHANGE-POINT DETECTION IN ASTRONOMICAL DATA BY USING A HIERARCHICAL MODEL AND A BAYESIAN SAMPLING APPROACH

*Nicolas Dobigeon*[†]*, Jean-Yves Tourneret*[†] *and Jeffrey D. Scargle*[*]

[†]IRIT/ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France
[*]Space Science Division, NASA, Ames Research Center, Moffett Field, CA, USA
Nicolas.Dobigeon@enseeiht.fr, Jean-Yves.Tourneret@tesa.prd.fr, Jeffrey.D.Scargle@nasa.gov

**ABSTRACT**

Detection of significant intensity variations in astronomical time-series can be achieved with 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 with real and synthetic photon counting data illustrate the performance of the proposed algorithm. An extension to joint segmentation of multiple time series is also discussed.

## 1. INTRODUCTION

The problem of segmenting signals or images has been considered with increasing interest in the literature (for instance see [1] and references therein). Many segmentation algorithms are based on Bayesian inference [2], [3], [4] or on the penalized maximum likelihood principle [5] (which can be viewed as a Bayesian strategy with appropriate prior knowledge). 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 [6]. 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, an appropriate stopping rule is required and the segmentation results strongly depends on this stopping rule. The multiple changepoint algorithm [7] 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. The current paper studies a new Bayesian time series segmentation algorithm that is not iterative and does not use a stopping rule. We emphasize posterior distributions for the various parameters, as these provide very practical statistical significance information.

As explained in [6], the arrival times of photons can be modeled accurately by a discrete time Poisson counting process. The numbers of photons counted in successive equally spaced intervals (bins) 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:

$$y_i \sim \mathcal{P}\left(\lambda_k\right),$$

where $i \in I_k = [l_{k-1}+1, ..., l_k[$, $k = 1, ..., K$ and the following notations have been used:

- $\mathcal{P}(\lambda)$ denotes a Poisson distribution with parameter $\lambda$,
- $K$ is the number of segments in the observed signal,
- $l_k$ is the sample point after which the $k^{th}$ change occurs in the signal (by convention $l_0 = 0$ and $l_K = n$ where $n$ is the number of observed samples). In other words, the actual change locations are $t_k = l_k T + \tau$ with $0 \leq \tau < T$, where $T$ is the sampling period.

Segmenting an astronomical time series consists of estimating the parameters $K$ and $l = (l_1, ..., l_{K-1})$ from the observations $y = (y_1, ..., y_n)$. This paper proposes a hierarchical model appropriate for the segmentation problem. This model assumes that appropriate prior distributions for the unknown parameters (change-point locations, Poisson parameters) 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 [2] [8]. Such methodology consists of assigning vague priors to the hyperparameters and to estimate these parameters from the observed data. The price to pay with the proposed hierarchical model is that the change-point Bayesian estimators such as the Maximum a posteriori (MAP) or minimum mean square (MMSE) estimators are difficult to derive. Markov Chain Monte Carlo

(MCMC) methods are then used to draw samples according to the posteriors of interest. The Bayesian estimators can be computed from these simulated samples.
This paper is organized as follows. Section 2 presents the hierarchical model used for the segmentation. Section 3 studies a Gibbs sampler for the posteriors of interest. Some simulation results on synthetic and real data are presented in Section 4. An extension to the joint segmentation of multiple time series is discussed in Section 5. Conclusions are reported in Section 6.

## 2. HIERARCHICAL BAYESIAN MODEL

The unknown parameters for the segmentation problem (introduced in the previous section) are $K$, $l = (l_1, ..., l_{K-1})$, $\lambda = (\lambda_1, ..., \lambda_K)$. A standard reparameterization consists of introducing indicators $r_i$, $i \in \{1, \ldots, n\}$ such that:

$$\begin{cases} r_i = 1 \text{ if there is a changepoint at lag } i, \\ r_i = 0 \text{ otherwise,} \end{cases}$$

with $r_n = 1$ (this condition ensures that the number of change-points and the number of steps are equal to $K = \sum_{i=1}^{n} r_i$). The unknown parameter vector resulting from this reparameterization is $\theta = (r, \lambda, K)$ with $r = (r_1, \ldots, r_n)$ and $\lambda = (\lambda_1, \ldots, \lambda_K)$. Note that the unknown parameter vector $\theta$ belongs to a space whose dimension depends on $K$ i. e. $\theta \in \Theta = \{0, 1\}^n \times \mathbb{R}^K$. This paper proposes estimating the unknown parameter vector $\theta$ by using Bayesian estimation theory. Bayesian inference on $\theta$ is based on the posterior distribution $f(\theta|y)$ (with $y = (y_1, \ldots, y_n)$). This posterior distribution is related to the likelihood of the observations and the parameter priors via Bayes' theorem $f(\theta|y) \propto f(y|\theta) f(\theta)$. The likelihood and priors for the segmentation problem are summarized below.

### 2.1. Likelihood

The likelihood of the observed vector $y$ can be expressed as follows:

$$\begin{aligned} f(y|\theta) &= \prod_{k=1}^{K} \prod_{i \in I_k} \frac{\lambda_k^{y_i} \exp(-\lambda_k)}{y_i!} \\ &= \frac{1}{\prod_{i=1}^{n} y_i!} \prod_{k=1}^{K} \lambda_k^{s_k(r)} \exp(-\lambda_k n_k(r)) \quad (1) \\ &\propto \prod_{k=1}^{K} \lambda_k^{s_k(r)} \exp(-\lambda_k n_k(r)), \end{aligned}$$

where $\propto$ means "proportional to", $s_k(r) = \sum_{i \in I_k} y_i$ and $n_k(r) = l_k - l_{k-1}$ (number of samples in the $k$th interval $I_k$).

### 2.2. Parameter Priors

1) **Indicator Vector**: we assume that the probabilities of having $r_i = 0$ and $r_i = 1$ do not depend on $i$. As a consequence, by assuming that the variables $r_i$ (for $i = 1, ..., n$) are *a priori* independent, the indicator prior distribution can be written

$$\begin{aligned} f(r|P) &= \prod_{i=1}^{n-1} P^{r_i} (1-P)^{1-r_i}, \\ &= P^{\sum_{i=1}^{n-1} r_i} (1-P)^{n-1-\sum_{i=1}^{n-1} r_i}, \end{aligned}$$

where $P$ is the probability of $r_i = 1$. See [3] [4] for a similar choice.
2) **Poisson parameters**: Gamma distributions are assigned to the Poisson parameters:

$$f(\lambda_k|\nu, \gamma) \sim \mathcal{G}(\nu, \gamma),$$

where $\nu = 1$ and $\gamma$ is an adjustable hyperparameter. By assuming the parameters $\lambda_k$ (for $k = 1, ..., K$) are *a priori* independent, the following prior distribution for $\lambda$ can be obtained:

$$\begin{aligned} f(\lambda|\gamma) &= \prod_{k=1}^{K} \frac{\gamma^\nu}{\Gamma(\nu)} \lambda_k^{\nu-1} e^{-\beta\lambda_k} \mathbb{I}_{\mathbb{R}^+}(\lambda_k), \\ &= \left(\frac{\gamma^\nu}{\Gamma(\nu)}\right)^K \prod_{k=1}^{K} \left(\lambda_k^{\nu-1} \mathbb{I}_{\mathbb{R}^+}(\lambda_k)\right) e^{-\gamma \sum_{k=1}^{K} \lambda_k}. \end{aligned}$$

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

$$\begin{aligned} &\mathbb{I}_{\mathbb{R}^+}(x) = 1 \text{ if } x \geq 0, \\ &\mathbb{I}_{\mathbb{R}^+}(x) = 0 \text{ otherwise.} \end{aligned}$$

The hyperparameter vector associated with the priors defined above is $\Phi = (P, \gamma)$. 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 [9]. However, in order to increase the robustness of the algorithm, hyperparameters can be considered as random variables with noninformative priors as in [2]. 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.

### 2.3. Hyperparameter priors

1) **Hyperparameter** $\gamma$: the prior distribution for $\gamma$ is a noninformative Jeffreys' prior (as in [2]) which reflects the absence of knowledge regarding this hyperparameter:

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

2) **Hyperparameter** $P$: the prior distribution for $P$ is a uniform distribution on $[0,1]$:

$$f(P) = \mathbb{I}_{[0,1]}(P).$$

This prior distribution indicates the absence of information regarding the hyperparameter $P$.
Assuming that the different hyperparameters are *a priori* independent, the prior distribution for the hyperparameter vector $\Phi$ can be written as follows:

$$f(\Phi) = \frac{1}{\gamma}\mathbb{I}_{\mathbb{R}^+}(\gamma)\mathbb{I}_{[0,1]}(P). \quad (2)$$

### 2.4. Posterior distribution of $\theta$

The posterior distribution of the unknown parameter vector $\theta = (\lambda, r)$ can be computed from the following hierarchical structure:

$$f(\theta|y) = \int f(\theta, \Phi|y)d\Phi \propto \int f(y|\theta)f(\theta|\Phi)f(\Phi)d\Phi,$$

where $f(y|\theta)$ and $f(\Phi)$ have been defined in (1) and (2). This hierarchical structure allows one to integrate out the nuisance parameters $\lambda_k$ and $P$ from the joint distribution $f(\theta, \Phi|y)$, yielding:

$$\frac{f(r,\gamma|y)}{C(r|y)} \propto \frac{1}{\gamma}\left(\frac{\gamma^\nu}{\Gamma(\nu)}\right)^K \prod_{k=1}^{K} \frac{\Gamma(n_k(r)+\nu)}{(s_k(r)+\gamma)^{n_k(r)+\nu}}, \quad (3)$$

with

$$C(r|y) = \Gamma\left(\sum_{i=1}^{n-1} r_i + 1\right)\Gamma\left(n - \sum_{i=1}^{n-1} r_i\right),$$

where $\Gamma(t)$ is the Gamma function. The posterior distribution (3) 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 which 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 [2]. However, it is important to note the following difference: our proposed Gibbs sampling strategy does not involve reversible jumps, since the unknown parameters in (3) belong to a space with fixed dimension.

## 3. GIBBS SAMPLER FOR CHANGE-POINT DETECTION

The Gibbs sampler 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 $f(r,\gamma|y)$ defined in (3) by the following two step procedure:

### 3.1. Generation of samples distributed according to $f(r|\gamma, y)$

This generation is achieved by using the Gibbs Sampler to draw $(n-1)$ samples distributed according to the distribution of the variable $r_i|\gamma, y$. This random variable is discrete and takes its values in $\mathcal{F} = \{0,1\}$. Consequently, its distribution is fully characterized by the probabilities $P(r_i = \epsilon|\gamma, y), \epsilon \in \mathcal{F}$. By using the notations $r_{j:k} = (r_j, ..., r_k)$ and $r_{-i} = (r_{1:i-1}, r_{i+1:n})$, the following result can be obtained:

$$P(r_i = \epsilon|r_{-i}, \gamma, y) \propto f(r_i(\epsilon), \gamma|y),$$

where $r_i(\epsilon) = (r_{1:i-1}, \epsilon, r_{i+1:n})$. This yields a closed-form expression of the probabilities $P(r_i = \epsilon|\gamma, y)$ after appropriate normalization.

### 3.2. Generation of samples distributed according to $f(\gamma|r, y)$

To obtain samples distributed according to $f(\gamma|r, y)$, it is very convenient to simulate vectors distributed according to the joint distribution $f(\gamma, \lambda|r, y)$ by using Gibbs moves. This step can be decomposed as follows:

- **Draw samples according to** $f(\lambda|\gamma, r, y)$
  Looking carefully at the joint distribution $f(\theta, \Phi|y)$, the following results can be obtained:

  $$\lambda_k|\gamma, r, y \sim \mathcal{G}(s_k(r) + \nu, n_k(r) + \gamma),$$

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

- **Draw samples according to** $f(\gamma|\lambda, r, y)$
  This is achieved as follows:

  $$\gamma|\lambda, r, y \sim \mathcal{G}\left(\nu K, \sum_{k=1}^{K} \lambda_k\right).$$

The hyperparameter $P$ carries information regarding the probability of having a change 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 $r$ and the vector of observed samples $y$ can be easily derived. Straightforward computations lead to:

$$f(P|r, y) \propto P^{\sum_{i=1}^{n-1} r_i}(1-P)^{n-1-\sum_{i=1}^{n-1} r_i}\mathbb{I}_{[0,1]}(P). \quad (4)$$

This is a Beta distribution with parameters $\sum_{i=1}^{n-1} r_i + 1$ and $n - \sum_{i=1}^{n-1} r_i$ denoted as $\mathcal{B}e\left(\sum_{i=1}^{n-1} r_i + 1, n - \sum_{i=1}^{n-1} r_i\right)$.

## 4. SIMULATION RESULTS

### 4.1. Synthetic data

The simulations presented in this paper have been obtained for a signal of $n = 120$ samples whose change-point locations are $l = (20, 50, 100, 120)$. The Poisson parameters for the different segments are $\lambda = (19, 9, 17, 7)$. 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 for the estimation). The MMSE estimates of the change locations (which are the estimated posterior probabilities to have changes at the different locations) are depicted on Fig. 1. The estimates are clearly in good agreement with the actual change locations.

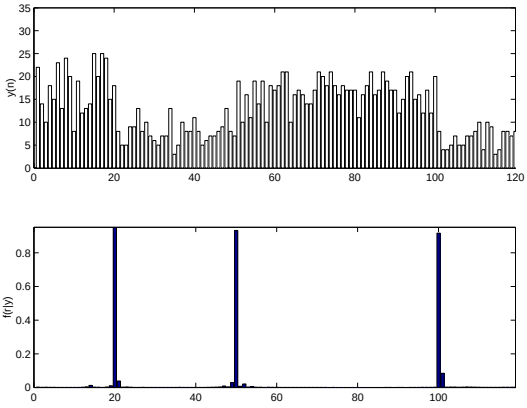


**Fig. 1**. Posterior distribution of the change-point locations.

The determination of the number of change-points is an important problem. The proposed algorithm generates samples $(r^n, \gamma^n)$ distributed according to the posterior distribution $f(r, \gamma|y)$, which allows for model selection. Indeed, for each sample $r^i$, the change-point number can be estimated by $\widehat{K}^i = \sum_{k=1}^{n} r_k^i$. Fig. 2 shows the histogram of $\widehat{K}^i$ computed from the last Markov chain samples. The histogram has a maximum value for $K = 4$, which corresponds to the actual number of changes.

The estimation of the Poisson parameters is interesting since it allows for signal reconstruction. The histograms of the estimated parameters $(\widehat{\lambda}_1, ..., \widehat{\lambda}_4)$ computed conditionally on $K^i = 4$ are presented on the Fig. 3. These histograms are in good agreement with the actual values of the Poisson parameters ($\lambda_1 = 19, \lambda_2 = 9, \lambda_3 = 17, \lambda_4 = 7$).

The last simulation results illustrate the performance of the hyperparameter estimation procedure. The estimated posterior of hyperparameter $P$ shown in Fig. 4 clearly coincides with the theoretical probability density function (4).

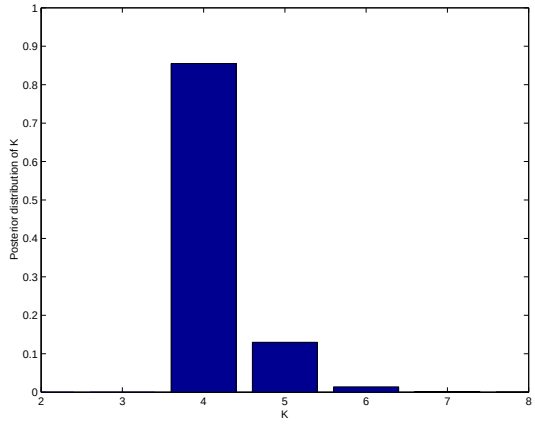


**Fig. 2**. Posterior distribution of the change-point number.

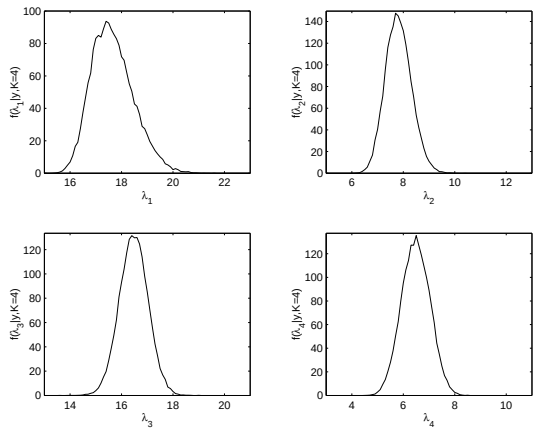


**Fig. 3**. Posterior distributions of the Poisson parameters $\lambda_i$ (for $i = 1, ..., 4$) conditioned on $K = 4$.

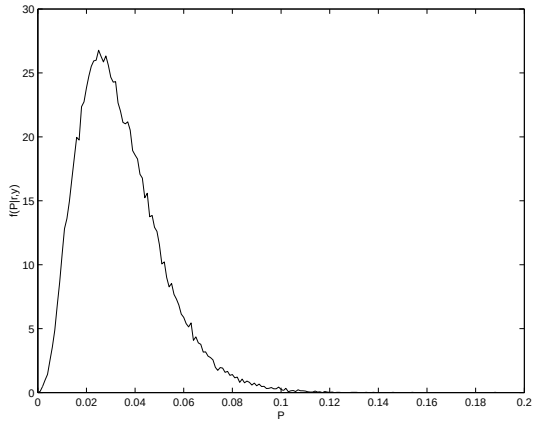


**Fig. 4**. Posterior distribution of the hyperparameter $P$.

### 4.2. Real astronomical data

The hierarchical method presented in this paper has been applied to the astronomical data studied in [6]. 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 $3.68ms$. All figures have been obtained after averaging the results of $64$ Markov Chains with $N_{MC} = 1550$ runs and $N_{bi} = 50$ burn-in iterations.

The first step of the analysis consists of estimating the number of change-points for the observed sequence depicted in Fig. $6.a$. The histogram of the estimated number of changes (computed from the last $1500$ Markov Chain output samples) is shown in Fig. $5$. The corresponding marginal MAP estimator is $\widehat{K} = 18$. Fig. $6.b$ shows the estimated posterior distribution of $r$.

The different Poisson intensities can be estimated on each segment from the change locations (by averaging the signal on each segment which corresponds to the intensity MAP estimator conditioned on $K = 18$). This procedure yields Bayesian blocks introduced in [6]. Fig. $6.a$ shows Bayesian blocks obtained after keeping $K = 18$ segments as suggested by Fig. $5$. However, 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 $[0.44; 0.47]$ appears in dotted lines on Fig. $6.b$. This high value could induce a modified segmentation including a change in this interval $[0.44; 0.47]$. These results are in good agreement with those of [6]. However, note again that the procedure used here does not require a stopping rule, other than that implicit in the assessment of the Markov chain's convergence.

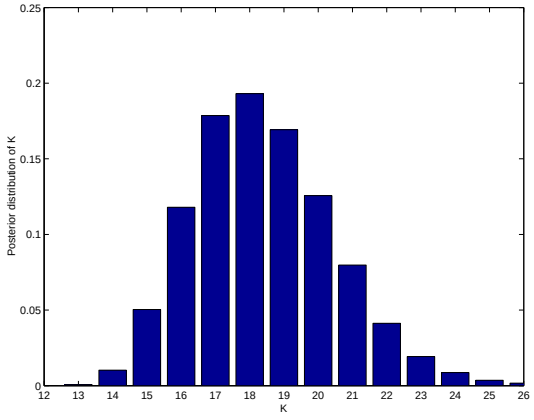


**Fig. 5**. Posterior distribution of the change-point number.

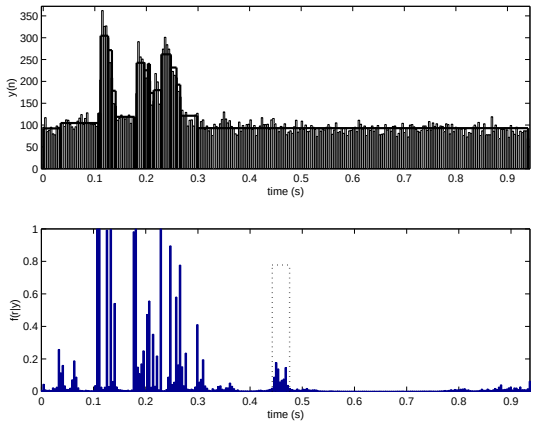


**Fig. 6**. Posterior distribution of the change-point locations.

## 5. EXTENSION TO JOINT SEGMENTATION

This section generalizes the previous analysis to the detection of changes in multiple astronomical time-series. This problem is important when signals coming from different sensors have to be jointly segmented. The change locations in the multiple time-series are supposed to be correlated: when a change is detected in one or several signals at a given position, the proposed algorithm allows favoring 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. Note that a similar strategy was proposed in [10] for joint detection of variance changes in Gaussian time-series. The observed sequences are assumed to be distributed according to Poisson distributions with piecewise constant parameters:

$$y_{j,i} \backsim \mathcal{P}\left(\lambda_{j,k}\right), i \in I_{j,k}, k = 1, ..., K_j \qquad (5)$$

where $I_{j,k} = [l_{j,k-1} + 1; l_{j,k}[$, $l_{j,k}$ is the sample point after which the $k^{th}$ change occurs in the $j^{th}$ sequence (with $l_{j,0} = 0$ and $l_{j,K_j} = n$, where $n$ is the number of samples of each observed sequence) and $K_j$ is the number of changes in the $j^{th}$ sequence. The sequences $y_k = (y_{k,1}, ..., y_{k,n})$ and $y_l = (y_{l,1}, ..., y_{l,n})$ are assumed to be independent for $k \neq l$. The presence or absence of changes in the $j^{th}$ time series ($j \in \{1, \ldots, J\}$) is governed by **indicators** $r_{j,i}$, where $i \in \{1, \ldots, n\}$. The different steps necessary to determine the posterior distribution of these change-point locations are described in [11].

This section presents some simulation results obtained for $J = 2$ sequences of $n = 120$ samples. The change-point locations for the two sequences are $l_1 = (20, 50, 100, 120)$, $l_2 = (50, 120)$ and the Poisson parameters are summarized in $\lambda_1 = (19, 9, 17, 7)$ and $\lambda_2 = (8, 11)$. The results have been obtained after averaging $64$ Monte Carlo runs. The total number of runs for each Markov Chain is $N_{MC} = 1000$,

including $N_{bi} = 200$ burn-in iterations. Fig. 7 shows the histograms of the estimated numbers of changes for the two time-series computed from the last Markov chain samples. The histograms have maximum values for $K_1 = 4$ and $K_2 = 2$, which correspond to the actual numbers of changes. The MMSE estimates of the change locations are depicted on Fig. 8 illustrating the performance of the joint segmentation procedure.

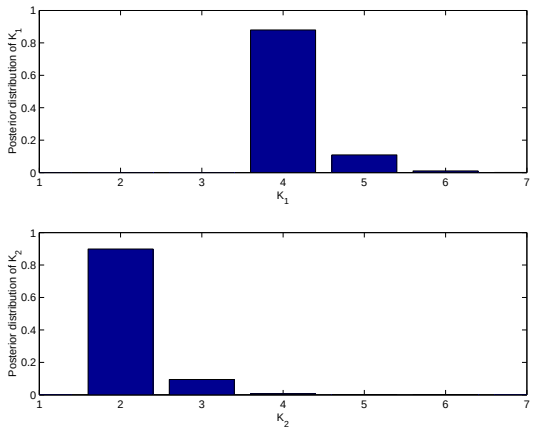


**Fig. 7**. Posterior distributions of the change-point numbers.

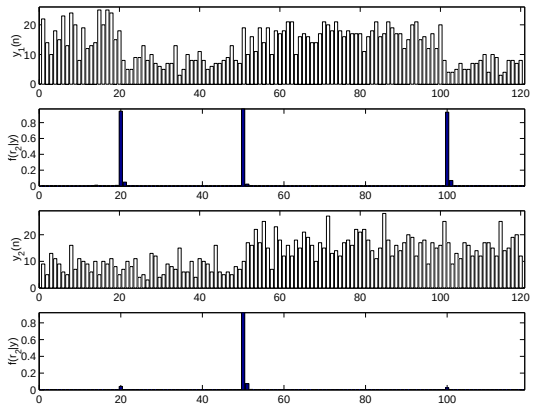


**Fig. 8**. Posterior distributions of the change-point locations.

## 6. CONCLUSIONS

This paper presents Bayesian sampling algorithms, for segmenting single and multiple time series obeying Poisson distributions with piecewise constant parameters, with appropriate hierarchical priors. Posterior distributions of the unknown parameters give estimates of the unknown parameters and their uncertainties. Simulation results conducted on synthetic and real signals illustrate the performance of the proposed methodologies.

## 7. REFERENCES


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

[2] 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, pp. 747–758, March 2002.

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

[4] 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, pp. 1871–1887, Sept. 2003.

[5] E. Lebarbier, "Detecting multiple change-points in the mean of gaussian process by model selection," *Signal Processing*, 2004. to appear.

[6] 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.

[7] 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.

[8] C. P. Robert, *The Bayesian Choice - A Decision Theoretic Motivation*. New-York: Springer Verlag, 1994.

[9] 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.

[10] 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*, (Saint-Louis, Missouri), pp. 593–596, Sept. 2003.

[11] N. Dobigeon, J.-Y. Tourneret, and J. Scargle, "Joint segmentation of astronomical data by using a hierarchical model and a Bayesian sampling approach," to be submitted.