HIERARCHICAL BAYESIAN SEGMENTATION OF SIGNALS CORRUPTED BY MULTIPLICATIVE NOISE

Jean-Yves Tourneret[*], *S. Suparman*[†] *and Michel Doisy*[*†]

[*] ENSEEIHT/IRIT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France
[†] LSP-UPS, 118 route de Narbonne, 31062 Toulouse cedex 4, France
jean-yves.tourneret@tesa.prd.fr, doisy@len7.enseeiht.fr, suparman@cict.fr

ABSTRACT

The paper addresses the important problem of signal segmentation, when signals are corrupted by multiplicative noise. A hierarchical Bayesian analysis is proposed to estimate the change-point locations and amplitudes. However, closed form expressions of the change-point parameter estimators are difficult to obtain. The proposed methodology draws samples distributed according to the change-point parameter posteriors by a Metropolis-within-Gibbs algorithm. The main advantage of the algorithm is that it allows joint estimation of the parameters and hyperparameters of the hierarchical model.

1. INTRODUCTION AND PROBLEM FORMULATION

Change-point detection has received considerable attention in signal and image processing applications. These applications include signal segmentation, fault detection, monitoring (for an overview see [1] and references therein) or image segmentation [2]. Most change-point detection strategies are based on the signal (or image) characteristics and on the statistical properties of the noise. This paper focuses on signals corrupted by multiplicative non Gaussian speckle noise as in [2, 3, 4]. Such signals have been shown to model accurately lines or columns of SAR images. As a consequence, their segmentation is interesting for SAR image edge detection. It is also interesting to note that these signals could model with minor changes (regarding the noise statistics) the Gaussian signals with piecewise-constant variances studied in [5].

The complex SAR reflectivity of a scene constitutes the primary geophysical data. However, a variety of other images can be obtained from the complex reflectivity. For instance, *since phase provides no information, we can get rid of it and use only amplitude, log, or intensity data* [2, p. 91]. This paper studies the segmentation of SAR image intensities as in [2, chapter 7], [3]. A given line (or column) of the complex reflectivity SAR image will be modeled as:

$$v_n = b_n m_n, \qquad n = 1, ..., N \tag{1}$$

where T is the sampling period, $b_n = b(nT)$, $m_n = m(nT)$, $v_n = v(nT)$ are the multiplicative speckle noise, the uncorrupted and corrupted line of the SAR image intensity respectively. Note that the present study focuses on the detection of vertical edges. However, two-dimensional implementations similar to those presented in [3, 6] can easily be developed using the proposed methodology. The properties of the signal and noise sequences b_n and m_n can be defined as follows:

• the autocovariance function of the speckle may decrease very rapidly. In this case, the speckle noise sequence b_n can be approximated by an *independent identically distributed* (iid) sequence of random variables with Gamma distribution whose parameters are L and L [3], [2, p. 99].

• The uncorrupted line of the SAR image intensity m_n can be modeled by K steps, when K fields with different reflectivities are considered. This model referred to as the Cartoon Model [2, p. 197] is a good approximation for important scene types such as agricultural fields.

Denote l_{i-1} (with $l_0 = 0$ and $l_K = N$) as the sample point after which there is the ith sudden change in the signal ($i = 1, ..., K$). In the following, the integers l_{i-1}will be referred to as change locations and the corresponding actual change locations are $t_{i-1} = l_{i-1}T + \tau$, with $0 < \tau < T$. The uncorrupted line of the SAR image can then be defined by:

$$m_n = A_i, \quad n \in]l_{i-1}, l_i], \quad i = 1, ..., K \tag{2}$$

where $A_i > 0$ is the ith step amplitude.

This paper addresses the problem of estimating the change-point locations l_i for $i \in \{1, ..., K-1\}$ from the observed data v_n. This edge detection problem is crucial in image segmentation. Once the change-point locations have been estimated, the line of the SAR image can be recovered by estimating the change-point amplitudes A_i. Consequently, the edge detection problem can also be used to recover the radar reflectivity (ideal image without speckle) (problem usually referred to as speckle filtering).

Many change-point detection strategies have already been studied in the literature for the segmentation of SAR images. Fjortoft *et als* [3] have proposed to model the change locations as random Poisson points and have derived the minimum mean square error (MMSE) estimator of the image reflectivity. By comparing the reflectivity MMSE estimator on opposite sides of the central pixel, they have obtained the so-called Ratio Of Exponentially Weighted Averages (ROEWA) detector. The ROEWA detector has shown interesting properties for the detection of multiple changes corrupted by multiplicative speckle noise [3]. Bayesian strategies studied in [7] have been adapted to the segmentation of piecewise constant processes corrupted by multiplicative speckle noise [8]. Appropriate priors for the change locations and amplitudes allow to estimate the change locations and amplitudes. However, the parameters of these priors, referred to as hyperparameters, have to be adjusted very carefully since they control the resolution level of the segmentation [7]. The estimation of hyperparameters was addressed in [8] by using an empirical Bayes analysis. Empirical Bayes analysis consists of estimating the unknown hyperparameters from the observed data, by using an appropriate estimation technique such as the moment method or the maximum likelihood method [9, p.

307]. The unknown hyperparameters are then replaced by their estimated values in the Bayesian model. However, empirical Bayes analysis may suffer from several problems. In particular, the approximated posterior obtained after replacing the hyperparameters by their estimates is only acceptable for large sample sizes (see [9, p. 309], for more details). The main contribution of this paper is to study a hierarchical Bayesian analysis (HBA) for the change-point detection problem. Such analysis allows to estimate the change-point locations and amplitudes as well as the hyperparameters.

The Bayesian change-point detector is studied in section 2. Section 3 addresses the problem of hyperparameter estimation by using HBA. Simulation results and conclusions are reported in sections 4 and 5.

2. BAYESIAN CHANGE-POINT DETECTION

Bayesian estimators are based on the posterior distribution of the unknown parameters. The unknown parameters for the problem defined in section 1 are the change-point locations l_i, the change-point amplitudes A_i and the number of change-points K. A standard reparametrization consists of defining indicators r_i such that:

$$\begin{cases} r_j = 1 & \text{if there is a change-point at instant } j \\ r_j = 0 & \text{otherwise} \end{cases} \quad (3)$$

for $j = 1, ..., N-1$. Conventionally, $r_N = 1$ such that the number of step changes equals the number of steps denoted $K(r) = \sum_{j=1}^{N} r_j$ with $r = (r_1, ..., r_{N-1})^T$. When using the indicators r_i, the unknown parameter vector for the change-point detection problem is $\theta = (r, A)^T$, where $A = (A_1, ..., A_K)^T$. Any Bayesian inference is based on the posterior distribution of θ conditional upon the observations $v_n, n = 1, ..., N$. This posterior distribution is usually determined from the Bayes rule, which requires to know the likelihood of the observations and to define appropriate priors for the unknown parameters.

2.1. Likelihood

Since the multiplicative speckle noise is distributed according to a Gamma distribution, the likelihood of the observed data can be expressed as:

$$f\left(v|\, K, A, l\right) \propto \prod_{i=1}^{N} {v_i}^{L-1} \prod_{k=1}^{K} \frac{1}{A_k^{L n_k}} \exp\left(-\sum_{k=1}^{K} \frac{L S_k}{A_k}\right) \quad (4)$$

with $n_k = l_k - l_{k-1}$ and $S_k = \sum_{i=l_{k-1}+1}^{l_k} v_i$. Equivalently, the likelihood can be expressed as a function of the indicators r_i and the change amplitudes:

$$f\left(v|\,\theta\right) \propto \exp\left(-L \sum_{k=1}^{K(r)} \left\{ \frac{S_k(r)}{A_k} + n_k(r) \log A_k \right\}\right) \quad (5)$$

2.2. Parameter Priors

In Bayesian inference, the choice of parameter priors is important and has received much attention in the literature [9]. This study uses the following priors for the change-point detection problem:

• Independent *Bernoulli priors* are chosen for the change-point locations:

$$f(r|\lambda) = \lambda^{K(r)-1} \left(1-\lambda\right)^{N-K(r)} \quad r \in \{0,1\}^{N-1} \quad (6)$$

The parameter $\lambda \in]0, 1[$ is the Bernoulli parameter which is the *a priori* probability of having a change-point at a given position.

• Independent Inverse Gamma (IG) priors (denoted $A_i \sim \mathcal{IG}(\alpha, \gamma)$) are chosen for the step amplitudes:

$$f\left(A|\, r, \alpha, \gamma\right) = \prod_{i=1}^{K(r)} \frac{\gamma^{\alpha} \exp\left(-\frac{\gamma}{A_i}\right)}{\Gamma\left(\alpha\right) {A_i}^{\alpha+1}} I_{\mathbb{R}^+}\left(A_i\right) \quad (7)$$

where $\alpha > 0$ and $\gamma > 0$ are two constants, $\Gamma(t)$ is the standard Gamma function $\Gamma\left(t\right) = \int_0^{+\infty} u^{t-1} e^{-u} du$ and $I_{\mathbb{R}^+}(.)$ is an indicator function ($I_{\mathbb{R}^+}(t) = 1$ if $t \in \mathbb{R}^+$ and $I_{\mathbb{R}^+}(t) = 0$ if $t \notin \mathbb{R}^+$). A suitable choice of parameters α and γ allow to incorporate either very vague or more specific prior information about the step amplitude. A motivation for choosing the IG prior is that the IG belongs to the conjugate family of priors for A with respect to the likelihood $f\left(v|\,\theta\right)$. In other words, $f\left(A|\, r, \alpha, \gamma\right)$ has the same "structure" as $f\left(v|\,\theta\right)$, when $f\left(v|\,\theta\right)$ is viewed as a function of A. This yields analytically tractable integration of $f\left(\theta|\, v, \lambda, \alpha, \gamma\right)$ with respect to A_i, i.e., allows marginalization.

To summarize, the prior distribution of the unknown parameter vector θ is defined as follows

$$f\left(\theta|\,\Phi\right) = f\left(A|\, r, \alpha, \gamma\right) f(r|\lambda) \quad (8)$$

where $\Phi = (\lambda, \alpha, \gamma)^T$.

2.3. Posterior distribution

Using Bayes' theorem, we can express the parameter posterior pdf as:

$$f\left(\theta|\, v, \Phi\right) \propto f\left(v|\,\theta\right) f(\theta|\Phi) \quad (9)$$

where $f\left(v|\,\theta\right)$ and $f\left(\theta|\,\Phi\right)$ have been defined in (5) and (8). Change-point detection may only require the estimation of the change-point vector r. In this case, the so called "nuisance parameters" A_i can be eliminated by integrating out A_i from the posterior pdf (9). Some straightforward computations allow to obtain the marginal posterior of r:

$$f(\, r|\, v, \Phi) = C(v, L) f(r|\lambda) \prod_{k=1}^{K(r)} \frac{\Gamma\left(\alpha + L n_k(r)\right)}{\left(\gamma + L S_k(r)\right)^{L n_k(r)+\alpha}} \quad (10)$$

with $C(v, L) = \left(\frac{L^L}{(L-1)!}\right)^N \prod_{i=1}^{N} v_i^{L-1}$. Equivalently, the marginal pdf of r can be written as $f\left(\, r|\, v, \Phi\right) \propto \exp\left(-U\left(\, r|\, v, \Phi\right)\right)$ where

$$U\left(\, r|\, v, \Phi\right) = \beta K(r) + \sum_{k=1}^{K(r)} \log \frac{\left(\gamma + L S_k(r)\right)^{L n_k(r)+\alpha}}{\Gamma\left(\alpha + L n_k(r)\right)} \quad (11)$$

is referred to as the energy function and $\beta = \log \frac{1-\lambda}{\lambda} - \log \frac{\gamma^{\alpha}}{\Gamma(\alpha)}$. Note that the parameter β is a decreasing function of λ. Consequently, the smaller β, the higher the *a priori* probability of a change and the fewer the omissions. On the other hand, the bigger β, the fewer the false alarms. The parameter β controls the resolution level of the segmentation: changes with small amplitudes will be detected for small values of β.

3. HYPERPARAMETER ESTIMATION USING HBA

The Bayesian analysis summarized in the previous section assumes that the hyparameters λ, α and γ are known a priori. This section proposes to estimate these parameters by using HBA. More precisely, we assume that the Bayesian statistical model $(f(v|\theta), f(\theta))$

is such that $f(\theta)$ is the marginal distribution of $f(\theta,\Phi)$, where $\Phi=(\lambda,\alpha,\gamma)^T$ contains the hyperparameters:

$$f(\theta)=\int f(\theta,\Phi)d\Phi=\int f(\theta|\Phi)f(\Phi)d\Phi \qquad (12)$$

In (12), $f(\Phi)$ is referred to as the (*a priori*) hyperparameter distribution. In the proposed implementation, the hyperparameter distribution has been chosen as follows:

$$f(\Phi)=f(\lambda)f(\gamma)f(\alpha)=\frac{1}{\gamma}I_{\mathbb{R}^+}(\gamma)I_{]0,c[}(\lambda)\delta(\alpha-1) \qquad (13)$$

where $\delta(.)$ is the Dirac delta function. In other words, α is not updated and set a priori to $\alpha=1$ (as in [10]), the prior probability to have a change-point at a given position λ is supposed to be uniform on $]0,c[$ and the hyperparameter γ is a priori distributed according to a non-informative Jeffreys prior. Note that parameter c is the only parameter which is adjusted in the algorithm ($c=0.1$ in our implementation). As explained in [9, p. 291], by decomposing the prior into several conditional levels of distributions, we improve the robustness of the resulting Bayes estimator. As a consequence, HBA has been extensively studied in the literature (for motivations, see [9] and references therein).

As a result of HBA, the joint distribution of (θ,Φ) conditional upon the observations v_n, $n=1,...,N$, can be determined by using the hierarchical structure:

$$f(\theta,\Phi|v)\propto f(v|\theta)f(\theta|\Phi)f(\Phi) \qquad (14)$$

Any Bayesian inference regarding the parameter vector θ is based on the posterior distribution $f(\theta|v)$. Unfortunately, a closed form expression of the marginal distribution $f(\theta|v)=\int f(\theta,\Phi|v)d\Phi$ is difficult to obtain. In such situation, Markov Chain Monte Carlo (MCMC) methods are classically used to draw samples according to the density of interest. This paper proposes to generate samples $(\theta^n,\Phi^n)=(A^n,r^n,\gamma^n,\lambda^n)$ distributed according to $f(\theta,\Phi|v)$ by using an hybrid MCMC sampler which combines Gibbs steps and Metropolis-Hastings steps. Such algorithm sometimes referred to as Metropolis-within-Gibbs algorithm has received much attention in the literature (see for instance [11]). After a sufficiently long *burn-in*, the MMSE estimator of the change-point parameters (locations and amplitudes) can be determined by computing the time average of the Markov chain output samples. The different steps of the Metropolis-within-Gibbs algorithm are detailed below:

3.1. Initialization

Set $(\theta^0,\Phi^0)=(r^0,A^0,\lambda^0,\gamma^0)$ and $n=1$.

3.2. Generation of samples distributed according to $f(\theta|\Phi^{n-1},v)$

In order to simplify notations, we drop the subscript $^{n-1}$ from variables at iteration n. The distribution of $\theta=(A,r)$ given the hyperparameter vector Φ can be expressed as

$$\begin{aligned} f(\theta|\Phi,v) &= f(A|r,\Phi,v)f(r|\Phi,v) \\ &= f(A|r,\Phi,v)\int f(A,r|\Phi,v)dA \end{aligned} \qquad (15)$$

Equation (15) shows that the determination of $f(\theta|\Phi,v)$ requires to determine $f(A|r,\Phi,v)$ and $f(r|\Phi,v)$.

Determination of $f(r|\Phi,v)$:
The pdf of r conditioned upon the hyperparameter vector Φ and the observation vector v has been obtained in (10). A possibility to sample from the non standard full conditional distribution of r consists of introducing a Metropolis-Hastings step. Here, we propose to apply the methodology described in [8]. Different kernels are used in turn to increase the algorithm convergence (the resulting hybrid strategy is called a *cycle*). The Markov chain state space and current state are denoted by $\Omega=\{0,1\}^{N-1}$ and $r^n=(r_i^n)_{i=1,...,N-1}\in\Omega$, respectively. The Markov chain moves are defined as follows:

a) The candidate $z^{n+1}\in\Omega$ is drawn independently of the current location r^n yielding *the independence sampler* [11] defined by $q(z^{n+1}|r^n)=q(z^{n+1})$, where q is an instrumental distribution. For our experiment, q is a Bernoulli distribution with parameter λ. In this procedure, the candidate z^{n+1} is selected using the classical acceptance probability,

b) Local changes are made following the *one-variable-at-a-time* MH algorithm. This variable-at-a-time step was suggested for instance in [11, p. 10] to increase the convergence speed. A random permutation of $\{1,...,N-1\}$ is uniformly drawn. According to this permutation, each component is flipped from 0 to 1 or from 1 to 0. The move is then accepted with the usual acceptance probability. This move visits each site randomly and all sites are visited in each scan.

The decomposition of the target distribution as a function of the energy function $f(r|v,\Phi)\propto\exp(-U(r|v,\Phi))$ allows to obtain the following acceptance probability (the instrumental distribution q in moves a) and b) is symmetric):

$$\begin{cases} r^{n+1}=z^{n+1} & \text{if } \ln(\text{rand})<-U(z^{n+1}|v,\Phi)+U(r^n|v,\Phi) \\ r^{n+1}=r^n & \text{otherwise} \end{cases} \qquad (16)$$

where rand is drawn according to a uniform distribution on $[0,1]$.

Determination of $f(A|r,\Phi,v)$:
The pdf of the change amplitude vector A conditioned upon the change location vector r, the hyperparameter vector Φ and the observation vector v can be expressed in closed form as follows

$$f(A|r,\Phi,v)\propto\prod_{k=1}^{K(r)}\frac{\exp\left(-\frac{\gamma+LS_k(r)}{A_k}\right)}{A_k^{\alpha+Ln_k(r)+1}} \qquad (17)$$

Equivalently, $f(A|r,\Phi,v)$ is the product of $K(r)$ IG pdfs:

$$f(A|r,\Phi,v)=\prod_{k=1}^{K(r)}\mathcal{IG}\left(\alpha+Ln_k(r),\gamma+LS_k(r)\right) \qquad (18)$$

As a consequence, samples distributed according to $f(A|r,\Phi,v)$ can be generated with standard IG generators.

3.3. Generation of samples distributed according to $f(\Phi|\theta^{n-1},v)$

As previously, in order to simplify notations, we drop the subscript $^{n-1}$ from all variables at iteration n. Eqs (5), (6), (7) and (13) show that

$$f(\Phi|\theta,v)\propto\lambda^{K(r)-1}(1-\lambda)^{N-K(r)}\gamma^{\alpha K(r)-1}\exp\left(-\gamma\sum_{i=1}^{K(r)}\frac{1}{A_i}\right) \qquad (19)$$

As a consequence, the second step of the Gibbs sampler consists of sampling λ and γ as follows

$$\begin{aligned} \lambda &\sim \mathcal{B}e(K(r),N-K(r)+1) \\ \gamma &\sim \mathcal{G}\left(\alpha K(r),\textstyle\sum_{i=1}^{K(r)}\frac{1}{A_i}\right) \end{aligned} \qquad (20)$$

where $\mathcal{B}e(a,b)$ and $\mathcal{G}(a,b)$ denote the Beta and Gamma distributions with parameters a and b (see [9, p. 381]).

4. SIMULATION RESULTS

Many simulations have been performed to illustrate the previous theoretical results. This papers considers a line of synthetic 4 look image ($L = 4$) depicted in Fig. 1. The change-point parameters are $A = (33.74, 16.73, 66.93, 16.73)$, $l = (0, 21, 112, 132, 174)$ and the number of samples is $N = 174$.

Indicators distributed according to $f(r|v)$ are simulated with the algorithm detailed in this paper. The minimum mean square error (MMSE) estimator of r is then computed by the time average of the last Markov chain output samples (convergence is guaranteed by the ergodic theorem for Markov chains). Fig. 1 shows the estimated posterior distribution $\hat{f}(r|v)$ computed after 1000 iterations (i.e. 1000 cycles) of the algorithm and a burn-in period of 1000 cycles. The MMSE estimator of r provides the posterior change-point probabilities which are in good agreement with the true change locations.

The algorithm studied in this paper draws vectors r^i distributed according to the distribution $f(r|v)$. For each vector r^i, the number of change-points is $K(r^i) = \sum_{j=1}^{N} r_j^i$. As a consequence, the posterior distribution of K can be easily estimated from the vectors r^i. Fig. 2 shows the estimated posterior distribution of $K(r)$. The histogram has a maximum value for $K(r) = 4$, which is in good agreement with the actual number of change-points (indeed, we have assumed that there is a change at $N = 174$).

Once the change-point locations been determined, the signal amplitudes can be estimated for signal reconstruction. Fig. 3 shows the histograms of the signal amplitude A_i conditioned upon $K = 4$. These histograms have to be compared with the true values of parameters A_i. Note that the noise has more influence on large amplitudes, which explains the bias for the estimate of A_3.

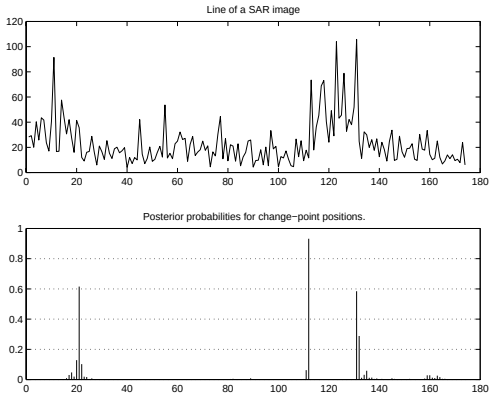

Fig. 1. Line of a SAR image and the estimated posterior change-point probabilities.

5. CONCONCLUSIONS

This paper studied the problem of detecting change-points in signals corrupted by multiplicative speckle noise. A hierarchical model similar to the model studied in [10] allowed to estimate jointly the change-point parameters and the model hyperparameters. The application of this model to SAR image edge detection requires to define an appropriate edge strength map [3]. This application is currently under investigation.

6. REFERENCES

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

[2] C. J. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Artech House, 1998.

[3] R. Fjortoft, A. Lopès, P. Marthon, and E. Cubero-Castan, "An optimal multiedge detector for SAR image segmentation," *IEEE Trans. Geosci. and Remote Sensing*, vol. 36, pp. 793–802, May 1998.

[4] S. Suparman, M. Doisy, and J. Tourneret, "Changepoint detection using reversible jump MCMC methods," in *Proc. IEEE ICASSP-02*, (Orlando, Florida), pp. 1569–72, May 2002.

[5] Z. Wang and P. Willett, "Fast and accurate variance segmentation of white Gaussian data," in *Proc. IEEE ICASSP-02*, (Orlando, Florida), pp. 1569–72, May 2002.

[6] R. Fjortoft, A. Lopès, J. Bruniquel, and P. Marthon, "Optimal edge detection and edge localization in complex SAR images with correlated speckle," *IEEE Trans. Geosci. and Remote Sensing*, vol. 37, pp. 2272–2281, Sept. 1999.

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

[8] J.-Y. Tourneret, M. Doisy, and M. Lavielle, "Bayesian off-line detection of multiple change-points corrupted by multiplicative noise. application to SAR image edge detection," *Signal Processing*, 2002. submitted.

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

[10] E. Punskaya, C. Andrieu, A. Doucet, and W. J. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Processing*, vol. 50, pp. 747–758, March 2002.

[11] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing Markov Chain Monte Carlo," in *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds.), pp. 1–19, London: Chapman & Hall, 1996.

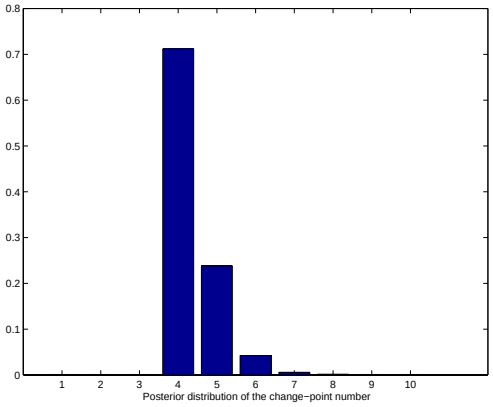

Fig. 2. Posterior distribution of the number of change-points.

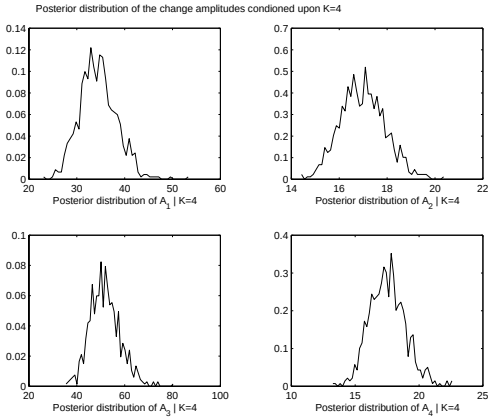

Fig. 3. Marginal posterior distributions of the change amplitudes.