MCMC SAMPLING FOR JOINT SEGMENTATION OF WIND SPEED AND DIRECTION

Nicolas Dobigeon and Jean-Yves Tourneret

University of Toulouse, IRIT/INP-ENSEEIHT, BP 7122, 31071 Toulouse cedex 7, FRANCE
`{nicolas.dobigeon, jean-yves.tourneret}@enseeiht.fr`

ABSTRACT

The problem of detecting changes in wind speed and direction is considered. Bayesian priors, with various degrees of certainty, are used to represent relationships between the two time series. Segmentation is then conducted using a hierarchical Bayesian model that accounts for correlations between the wind speed and direction. A Gibbs sampling strategy overcomes the computational complexity of the hierarchical model and is used to estimate the unknown parameters and hyperparameters. The performance of the proposed algorithm is illustrated with results obtained with synthetic data.

Index Terms— Monte Carlo methods, Bayesian inference, hierarchical model, joint segmentation, wind data.

1. INTRODUCTION

Segmentation has received considerable attention in the literature (see for instance [1] and references therein). Off-line segmentation algorithms assume a collection of observations is available. One or multiple change-points instants are estimated from this collection. The change instants are estimated retrospectively (or equivalently detected retrospectively). This procedure explains the terminology "off-line" or "retrospective" used in the literature (see for instance [2]). The most common off-line segmentation strategies are based on the least squares principle [3], the maximum likelihood method or Bayesian strategies [4]. A common idea to all these strategies optimizes a criterion that depends upon the observed data, penalized by an appropriate cost function for preventing over-segmentation. The choice of the appropriate penalty for signal segmentation remains a fundamental problem. The penalty can be selected using the ideas of [3] in the framework of Gaussian model selection. However, this approach is difficult in practical applications where the noise level is unknown. Another idea views the penalty as prior information about the change-point locations within the Bayesian paradigm [5]. The segmentation is then conducted using the posterior distribution of the change locations. The choice of the prior distribution is of course very important since the segmentation performance strongly depends on this choice. Markov chain Monte Carlo (MCMC) methods are often considered when the posterior distribution is too complicated to be used to derive the Bayesian estimators of the change parameters, such as the maximum a posteriori (MAP) estimator or minimum mean square error (MMSE) estimator. It is important to mention here that Bayesian methods (possibly combined with MCMCs) have useful properties for signal segmentation. In particular, they allow estimation of the unknown model hyperparameters using hierarchical Bayesian algorithms [6, 7].

This paper studies the problem of jointly segmenting the wind speed and direction, recently introduced in [8]. A hierarchical Bayesian model is proposed for the wind vector, assuming that appropriate prior distributions for the unknown parameters are available. Vague priors are assigned to the corresponding hyperparameters, that are then integrated out from the joint posterior distribution. An interesting property of the proposed approach is that it allows the estimation of the unknown model hyperparameters from the data (the hyperparameters could also be estimated by coupling MCMCs with an expectation-maximization algorithm as in [9]). Note that the proposed hierarchical algorithm differs from the algorithms developed in [8]. There the hyperparameters were fixed (by assuming the true value of the average of detected change is known) or adjusted from prior information regarding the tide hours. The hierarchical model developed in this paper estimates jointly the unknown parameters and hyperparameters from the observed data. There is a price to pay with the proposed hierarchical model. The change-point posterior distribution is too complicated to compute closed form expressions for the MAP and MMSE change-point estimators. This problem is solved by an appropriate Gibbs sampling strategy. The strategy draws samples according to the posteriors of interest and computes the Bayesian change-point estimators by using these samples.

1.1. Notation and problem formulation

Determination of the statistical properties of the wind vector amplitude has been previously considered in the literature. The Weibull and lognormal distributions have been reported as the best distributions for wind speeds [10]. Choice of one of these depends upon many factors, including site location or low/strong wind intensities. The choice of these distributions is supported by many experiments with real data sets. This paper concentrates on the lognormal distribution. However, a similar analysis could be conducted for the Weibull distribution or the Rayleigh distribution (see [11] for more details). For the lognormal assumption, the statistical properties of the wind speed on the kth segment can be defined as follows:

$$y_{1,i} \sim \mathcal{LN}\left(m_k, \sigma_k^2\right), \tag{1}$$

where $k = 1, ..., K_1$, $i \in I_{1,k} = \{l_{1,k-1} + 1, ..., l_{1,k}\}$, and the following notations have been used:

- $\mathcal{LN}\left(m, \sigma^2\right)$ denotes a lognormal distribution with scale m and shape σ^2,
- K_1 is the number of segments in the speed sequence $\mathbf{y}_1 = [y_{1,1}, \ldots, y_{1,n}]$ (n is the sample length).

The wind direction is a circular random variable, described by a Von Mises (VM) distribution whose mean direction parameter varies from one segment to another:

$$y_{2,i} \sim \mathcal{M}\left(\psi_k, \kappa\right), \tag{2}$$

where $k = 1, ..., K_2$, $i \in I_{2,k} = \{l_{2,k-1} + 1, ..., l_{2,k}\}$, and the following notations have been used:

- $\mathcal{M}(\psi, \kappa)$ denotes a VM distribution with mean direction ψ and concentration parameters κ [12, p. 36],
- K_2 is the number of segments in the direction sequence $\mathbf{y}_2 = [y_{2,1}, \ldots, y_{2,n}]$.

For both the wind speed and direction sequences, the segment #k in the signal #j ($j \in \{1, 2\}$) has boundaries denoted by $[l_{j,k-1} + 1, l_{j,k}]$, $l_{j,k}$ is the time index immediately after which a change occurs with the convention that $l_{j,0} = 0$ and $l_{j,K_j} = n$. Finally, the wind speed $\mathbf{y}_1$ and direction $\mathbf{y}_2$ are assumed to be independent sequences.

1.2. Paper organization

This paper proposes a Bayesian framework and an efficient algorithm for estimating the change-point locations $l_{j,k}$ from the two observed time series $\mathbf{y}_j$, $j \in \{1, 2\}$. The Bayesian model requires appropriate priors for the change-point locations, wind speed and direction parameters. The Bayesian model also needs to adjust the corresponding hyperparameters. The proposed methodology assigns vague priors to the unknown hyperparameters. The priors are then integrated out from the joint posterior or estimated from the observed data. This results in a hierarchical Bayesian model described in Section 2. A Gibbs sampling strategy is studied in Section 3 and allows the generation of samples distributed according to the posterior of interest. Some simulation results for synthetic wind data are presented in Section 4. Conclusions are reported in Section 5.

2. HIERARCHICAL BAYESIAN MODEL

The joint segmentation problem is based upon estimation of the following unknown parameters

- (K_1, K_2) (numbers of segments),
- $\mathbf{l}_j = (l_{j,1}, \ldots, l_{j,K_j})$ for $j \in \{1, 2\}$ (change-point locations),
- $\mathbf{m} = [m_1, \ldots, m_{K_1}]^\mathsf{T}$ (wind speed scale parameters),
- $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_{K_1}^2]^\mathsf{T}$ (wind speed shape parameters),
- $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_{K_2}]^\mathsf{T}$ (wind direction means),
- κ (wind direction concentration parameter).

A standard re-parameterization [5, 7] introduces indicator variables $\mathrm{r}_{j,i}$ ($j \in \{1, 2\}$, $i \in \{1, \ldots, n\}$) such that:

$$\begin{cases} \mathrm{r}_{j,i} = 1 \text{ if there is a change-point at time } i \text{ in signal } \#j, \\ \mathrm{r}_{j,i} = 0 \text{ otherwise,} \end{cases}$$

with $\mathrm{r}_{j,n} = 1$. This last condition ensures that the number of change-points and segments are equal for each signal, i.e. $K_j = \sum_{i=1}^{n} \mathrm{r}_{j,i}$. Using these indicator variables, the unknown parameter vector is $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$, where $\boldsymbol{\theta}_1 = \{\mathbf{r}_1, \mathbf{m}, \boldsymbol{\sigma}^2\}$, $\boldsymbol{\theta}_2 = \{\mathbf{r}_2, \boldsymbol{\psi}, \kappa\}$ and $\mathbf{r}_j = [\mathrm{r}_{j,1}, \ldots, \mathrm{r}_{j,n}]$. Note that the parameter vector $\boldsymbol{\theta}$ belongs to a space whose dimension depends on K_1 and K_2, i.e., $\boldsymbol{\theta} \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2$ with $\boldsymbol{\Theta}_1 = \{0, 1\}^n \times (\mathbb{R} \times \mathbb{R}^+)^{K_1}$ and $\boldsymbol{\Theta}_2 = \{0, 1\}^n \times [-\pi, \pi]^{K_2} \times \mathbb{R}$.

This paper proposes a Bayesian approach for estimating the unknown parameter vector $\boldsymbol{\theta}$. Bayesian inference on $\boldsymbol{\theta}$ is based on the posterior distribution $f(\boldsymbol{\theta}|\mathbf{Y})$, with $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2]^\mathsf{T}$, where $f(\boldsymbol{\theta}|\mathbf{Y})$ is related to the observation likelihoods and the parameter priors via Bayes rule $f(\boldsymbol{\theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})$. The likelihood and priors are presented below for the joint segmentation of wind speed and direction.

2.1. Likelihoods

The likelihood function of the wind speed sequence $\mathbf{y}_1$ can be expressed as follows:

$$f(\mathbf{y}_1|\boldsymbol{\theta}_1) \propto \prod_{k=1}^{K_1} \left(\sigma_k^2\right)^{-\frac{n_{1,k}}{2}} e^{-\frac{T_{1,k}^2 - 2m_k S_{1,k} + n_{1,k} m_k^2}{2\sigma_k^2}}, \quad (3)$$

where $\boldsymbol{\theta}_1 = (\mathbf{r}_1, \boldsymbol{\sigma}^2, \mathbf{m})$, $\propto$ means "proportional to",

$$T_{1,k}^2 = \sum_{i \in I_{1,k}} \log^2(y_{1,i}), \qquad S_{1,k} = \sum_{i \in I_{1,k}} \log(y_{1,i}), \quad (4)$$

and $n_{1,k}(\mathbf{r}_1) = l_{1,k} - l_{1,k-1}$ denotes the length of segment #k in the wind speed sequence $\mathbf{y}_1$.

The likelihood function of the wind direction sequence $\mathbf{y}_2$ is:

$$f(\mathbf{y}_2|\boldsymbol{\theta}_2) \propto \prod_{k=1}^{K_2} [\mathcal{I}_0(\kappa)]^{-n_{2,k}} e^{-\kappa \sum_{i \in I_{2,k}} \cos(y_{2,i} - \psi_k)}, \quad (5)$$

where $\boldsymbol{\theta}_2 = (\mathbf{r}_2, \boldsymbol{\psi}, \kappa)$, $\mathcal{I}_p(\cdot)$ is the modified Bessel function of the first kind and order p and $n_{2,k}(\mathbf{r}_2) = l_{2,k} - l_{2,k-1}$ denotes the length of segment #k in the wind direction sequence $\mathbf{y}_2$.

By assuming independence between the wind speed and direction sequences, conditioned on the change-point locations, the full likelihood function is:

$$f(\mathbf{Y}|\boldsymbol{\theta}) = f(\mathbf{y}_1|\boldsymbol{\theta}_1) f(\mathbf{y}_2|\boldsymbol{\theta}_2), \quad (6)$$

where the two densities appearing in the right hand side have been defined in (3) and (5).

2.2. Parameter Priors

2.2.1. Indicators

Correlations in the observed wind speed and direction signals are modeled by an appropriate prior distribution $f(\mathbf{R}|\mathbf{P})$, where $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2]^\mathsf{T}$ and $\mathbf{P}$ is an hyperparameter defined below. A *global abrupt change configuration* is defined as a specific value of the matrix $\mathbf{R}$ composed of 0's and 1's, corresponding to a specific solution of the joint segmentation problem. Conversely, a *local abrupt change configuration*, denoted $\boldsymbol{\epsilon} \in \mathcal{E} = \{0, 1\}^2$, is a particular value of a column of $\mathbf{R}$ which corresponds to the presence/absence of changes at a given time in the two signals. Denote as $P_{\boldsymbol{\epsilon}}$ the probability of having a local change configuration $\boldsymbol{\epsilon}$ at time i. First assume that $P_{\boldsymbol{\epsilon}}$ does not depend on the time index i and that $[\mathrm{r}_{1,i}, \mathrm{r}_{2,i}]^\mathsf{T}$ is independent of $[\mathrm{r}_{1,i'}, \mathrm{r}_{2,i'}]^\mathsf{T}$ for any $i \neq i'$. As a consequence, the indicator prior distribution expresses as:

$$f(\mathbf{R}|\mathbf{P}) = \prod_{\boldsymbol{\epsilon} \in \mathcal{E}} P_{\boldsymbol{\epsilon}}^{S_{\boldsymbol{\epsilon}}(\mathbf{R})} = P_{00}^{S_{00}} P_{10}^{S_{10}} P_{01}^{S_{01}} P_{11}^{S_{11}}, \quad (7)$$

where $P_{\boldsymbol{\epsilon}} = \Pr\left[[\mathrm{r}_{1,i}, \mathrm{r}_{2,i}]^\mathsf{T} = \boldsymbol{\epsilon}\right]$, $\mathbf{P} = \{P_{00}, P_{01}, P_{10}, P_{11}\} = \{P_{\boldsymbol{\epsilon}}\}_{\boldsymbol{\epsilon} \in \mathcal{E}}$ and $S_{\boldsymbol{\epsilon}}(\mathbf{R})$ is the number of times i such that $[\mathrm{r}_{1,i}, \mathrm{r}_{2,i}]^\mathsf{T} = \boldsymbol{\epsilon}$. Some comments regarding the prior distribution $f(\mathbf{R}|\mathbf{P})$ are appropriate. A value of $P_{\boldsymbol{\epsilon}}$ near one indicates a very likely configuration $[\mathrm{r}_{1,i}, \mathrm{r}_{2,i}]^\mathsf{T} = \boldsymbol{\epsilon}$ for all $i = 1, \ldots, n$. For instance, choosing $P_{00} \approx 1$ (resp. $P_{11} \approx 1$) favors a simultaneous absence (resp. presence) of changes in the two observed signals. This choice introduces correlation between the change-point locations.

2.2.2. Wind speed parameter priors

Inverse-Gamma distributions are elected as shape parameter priors:

$$\sigma_k^2 \mid \nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right), \qquad (8)$$

where $\mathcal{IG}(a, b)$ denotes the Inverse-Gamma distribution with parameters a and b, $\nu = 2$ (as in [4]) and γ is an adjustable hyperparameter. This is the *conjugate prior* for the wind speed variance which has been used successfully in [4] for Bayesian curve fitting.

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

$$m_k \mid \sigma_k^2, \delta_0^2 \sim \mathcal{N}\left(m_0, \sigma_k^2 \delta_0^2\right), \qquad (9)$$

where $m_0 = 0$ and δ_0^2 is an adjustable hyperparameter. The conjugate priors allow one to easily integrate out the shape and scale parameters in the posterior $f(\boldsymbol{\theta}|\mathbf{Y})$.

2.2.3. Wind direction parameter priors

Bayesian inference for VM distributions has been studied in [13], where appropriate priors for the VM parameters were defined. This paper uses the same prior, i.e, a conjugate VM distributions for wind mean directions:

$$\psi_k \mid \psi_0, R_0, \kappa \sim \mathcal{M}\left(\psi_0, \kappa R_0\right), \qquad (10)$$

where R_0 and ψ_0 are fixed hyperparameters and a Jeffreys' prior for the concentration parameter κ:

$$f(\kappa) \propto \frac{1}{\kappa} \mathbf{1}_{\mathbb{R}^+}(\kappa), \qquad (11)$$

where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ is the indicator function on $\mathbb{R}^+$. This prior describes vague knowledge regarding the VM concentration parameter κ.

2.3. Hyperpriors

The hyperparameter vector defined above and associated with the parameter priors is $\boldsymbol{\Phi} = \left(\mathbf{P}, \gamma, \delta_0^2\right)$. Of course, the performance of the Bayesian segmenter for detecting changes in the two signals $\mathbf{y}_1$ and $\mathbf{y}_2$ depends strongly on these hyperparameter values. The proposed hierarchical Bayesian model allows estimation of these hyperparameters from the observed data. The hierarchical Bayesian model requires definition of the hyperparameter priors (hyperpriors). Summarized below are the hyperpriors for the joint segmentation of wind speed and direction.

2.3.1. Hyperparameter $\mathbf{P}$

The prior distribution for the hyperparameter $\mathbf{P}$ is a Dirichlet distribution with parameter vector $\boldsymbol{\alpha} = \left[\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}\right]^{\mathsf{T}}$ defined on the simplex $\mathcal{P} = \{\mathbf{P} \mid \sum_{\boldsymbol{\epsilon} \in \mathcal{E}} P_{\boldsymbol{\epsilon}} = 1, P_{\boldsymbol{\epsilon}} \geq 0\}$ denoted as $\mathbf{P} \sim D_4(\boldsymbol{\alpha})$. This distribution is a classical prior for positive parameters summing to one. It allows marginalization of the posterior distribution $f(\boldsymbol{\theta}|\mathbf{Y})$ with respect to $\mathbf{P}$. Moreover, by choosing $\alpha_{\boldsymbol{\epsilon}} = \alpha, \forall \boldsymbol{\epsilon} \in \mathcal{E}$, the Dirichlet distribution reduces to the uniform distribution on $\mathcal{P}$.

2.3.2. Hyperparameters γ and δ_0^2

The priors for hyperparameters γ and δ_0^2 are selected as a noninformative Jeffreys' prior and a vague conjugate Inverse-Gamma distribution (i.e, with large variance). This choice reflects lack of precise knowledge of these hyperparameters:

$$f(\gamma) = \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma), \qquad \delta_0^2 \mid \xi, \beta \sim \mathcal{IG}\left(\xi, \beta\right). \qquad (12)$$

Assume that the individual hyperparameters are independent. The full hyperparameter $\boldsymbol{\Phi}$ prior distribution can then be written (up to a normalizing constant):

$$\begin{aligned}
f\left(\boldsymbol{\Phi} \mid \boldsymbol{\alpha}, \xi, \beta\right) &\propto \left(\prod_{\boldsymbol{\epsilon} \in \mathcal{E}} P_{\boldsymbol{\epsilon}}^{\alpha_{\boldsymbol{\epsilon}} - 1}\right) \mathbf{1}_{\mathcal{P}}(\mathbf{P}) \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma) \\
&\times \frac{\beta^{\xi}}{\Gamma(\xi)(\delta_0^2)^{\xi+1}} \exp\left(-\frac{\beta}{\delta_0^2}\right) \mathbf{1}_{\mathbb{R}^+}\left(\delta_0^2\right),
\end{aligned} \qquad (13)$$

where $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ is the Gamma function.

2.4. Posterior distribution of $\boldsymbol{\theta}$

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

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \int f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y}) d\boldsymbol{\Phi} \propto \int f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}) d\boldsymbol{\Phi}, \qquad (14)$$

where

$$\begin{aligned}
f\left(\boldsymbol{\theta}|\boldsymbol{\Phi}\right) &= f\left(\mathbf{R}|\mathbf{P}\right) \left[\prod_{k=1}^{K_1} f\left(\sigma_k^2|\nu, \gamma\right) f\left(m_k|\sigma_k^2, \delta_0^2\right)\right] \\
&\times \left[\prod_{k=1}^{K_2} f\left(\psi_k|\kappa, R_0, \psi_0\right)\right] f(\kappa),
\end{aligned} \qquad (15)$$

and $f(\mathbf{Y}|\boldsymbol{\theta})$ and $f(\boldsymbol{\Phi})$ are defined in (6) and (13). This hierarchical structure allows one to integrate out the nuisance parameters $\mathbf{m}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}$ and $\mathbf{P}$ from the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y})$, yielding:

$$\begin{aligned}
f\left(\mathbf{R}, \gamma, \delta_0^2, \kappa|\mathbf{Y}\right) &\propto \frac{1}{\kappa} \mathbf{1}_{\mathbb{R}^+}(\kappa) f\left(\delta_0^2|\xi, \beta\right) C\left(\mathbf{R}|\mathbf{Y}, \boldsymbol{\alpha}\right) \\
&\times \left[\frac{\left(\frac{\gamma}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\right]^{K_1} \left[\frac{1}{\mathcal{I}_0(\kappa)}\right]^N \prod_{k=1}^{K_2} \left[\frac{\mathcal{I}_0\left(R_k \kappa\right)}{\mathcal{I}_0\left(R_0 \kappa\right)}\right] \\
&\times \prod_{k=1}^{K_1} \left[\frac{1}{2}\left(\frac{1}{1 + n_{1,k}\delta_0^2}\right)^{\frac{1}{2}} \Gamma\left(\frac{\nu + n_{1,k}}{2}\right)\right] \\
&\times \prod_{k=1}^{K_1} \left[\gamma + T_{1,k}^2 + \frac{m_0^2 - \mu_{1,k}^2\left(1 + \delta_0^2 n_{1,k}\right)}{\delta_0^2}\right]^{-\frac{n_{1,k}}{2}},
\end{aligned} \qquad (16)$$

with

$$\begin{cases}
\mu_{1,k} = \dfrac{m_0 + \delta_0^2 S_{1,k}}{1 + \delta_0^2 n_{1,k}}, \\[2ex]
R_k^2 = \left(R_0 \cos \psi_0 + \displaystyle\sum_{i \in I_{2,k}} \cos y_{2,i}\right)^2 \\[3ex]
\qquad + \left(R_0 \sin \psi_0 + \displaystyle\sum_{i \in I_{2,k}} \sin y_{2,i}\right)^2,
\end{cases} \qquad (17)$$

and

$$C\left(\mathbf{R}|\mathbf{Y},\boldsymbol{\alpha}\right) = \frac{\prod_{\boldsymbol{\epsilon}\in\{0,1\}^2}\Gamma\left(S_{\boldsymbol{\epsilon}}(\mathbf{R})+\alpha_{\boldsymbol{\epsilon}}\right)}{\Gamma\left(\sum_{\boldsymbol{\epsilon}\in\{0,1\}^2}(S_{\boldsymbol{\epsilon}}(\mathbf{R})+\alpha_{\boldsymbol{\epsilon}})\right)}. \quad (18)$$

The posterior distribution in (16) is too complex to derive closed-form expressions for the MAP or MMSE estimators for the unknown parameters. In this case, MCMC methods are implemented to generate samples that are asymptotically distributed according to the posteriors of interest. The samples can then be used to estimate the unknown parameters by replacing integrals with empirical averages over the MCMC samples.

3. A GIBBS SAMPLER FOR JOINT SIGNAL SEGMENTATION

Gibbs sampling is an iterative sampling strategy. It consists of generating random samples (denoted by $\tilde{\cdot}^{(t)}$ where t is the iteration index) distributed according to the conditional posterior distributions of each parameter. This paper samples the distribution $f(\mathbf{R},\gamma,\delta_0^2,\kappa|\mathbf{Y})$ defined in (16) using the three step procedure outlined below. The main steps of the algorithm and the key equations are detailed in subsections 3.1 to 3.4.

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

This step is achieved by using a Gibbs sampler to generate Monte Carlo samples distributed according to $f\left([r_{1,i},r_{2,i}]^\mathsf{T}|\gamma,\delta_0^2,\kappa,\mathbf{Y}\right)$. This vector is a random vector of Booleans in $\mathcal{E}$. Consequently, its distribution is fully characterized by the probabilities

$$\Pr\left[[r_{1,i},r_{2,i}]^\mathsf{T}=\boldsymbol{\epsilon}|\gamma,\delta_0^2,\mathbf{Y}\right],\ \boldsymbol{\epsilon}\in\mathcal{E}.$$

Let $\mathbf{R}_{-i}$ denote the matrix $\mathbf{R}$ whose i^{th} column has been removed. Thus

$$\Pr\left[[r_{1,i},r_{2,i}]^\mathsf{T}=\boldsymbol{\epsilon}|\mathbf{R}_{-i},\gamma,\delta_0^2,\mathbf{Y}\right] \propto f(\mathbf{R}_i(\boldsymbol{\epsilon}),\gamma,\delta_0^2|\mathbf{Y}), \quad (19)$$

where $\mathbf{R}_i(\boldsymbol{\epsilon})$ is the matrix $\mathbf{R}$ whose i^{th} column has been replaced by the vector $\boldsymbol{\epsilon}$. Equation in (19) yields a closed-form expression of the probabilities $\Pr\left[[r_{1,i},r_{2,i}]^\mathsf{T}=\boldsymbol{\epsilon}|\mathbf{R}_{-i},\gamma,\delta_0^2,\mathbf{Y}\right]$ after appropriate normalization.

3.2. Generation of samples according to $f\left(\gamma,\delta_0^2|\mathbf{R},\mathbf{Y}\right)$

To obtain samples distributed according to $f\left(\gamma,\delta_0^2|\mathbf{R},\mathbf{Y}\right)$, it is very convenient to generate vectors distributed according to the joint distribution $f\left(\gamma,\delta_0^2,\boldsymbol{\sigma}^2,\mathbf{m}|\mathbf{R},\mathbf{Y}\right)$ by using Gibbs moves. Using the joint distribution (16), this step can be decomposed as follows:

- **Generate samples according to** $f\left(\gamma,\boldsymbol{\sigma}^2|\mathbf{R},\delta_0^2,\mathbf{Y}\right)$
  Integrating the joint distribution $f(\boldsymbol{\theta},\boldsymbol{\Phi}|\mathbf{Y})$ with respect to the scale parameters m_k, the following results are obtained:

$$\sigma_k^2|\mathbf{R},\gamma,\delta_0^2,\mathbf{Y} \sim \mathcal{IG}\left(u_k,v_k\right), \quad k=1,\dots,K_1 \quad (20)$$

  with

$$\begin{cases} u_k = \dfrac{\nu+n_{1,k}}{2} \\ v_k = \dfrac{\gamma+T_{1,k}^2}{2} + \dfrac{m_0^2-\mu_{1,k}^2\left(1+\delta_0^2 n_{1,k}\right)}{2\delta_0^2}, \end{cases} \quad (21)$$

and

$$\gamma|\mathbf{R},\boldsymbol{\sigma}^2 \sim \mathcal{G}\left(\frac{K_1\nu}{2},\sum_{k=1}^{K_1}\frac{1}{2\sigma_k^2}\right), \quad (22)$$

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

- **Generate samples according to** $f(\delta_0^2,\mathbf{m}|\mathbf{R},\boldsymbol{\sigma}^2,\mathbf{Y})$
  This is achieved as follows:

$$m_k|\mathbf{R},\boldsymbol{\sigma}^2,\delta_0^2,\mathbf{Y} \sim \mathcal{N}\left(\frac{m_0+\delta_0^2 S_{1,k}}{1+\delta_0^2 n_{1,k}},\frac{\delta_0^2\sigma_k^2}{1+\delta_0^2 n_{1,k}}\right), \quad (23)$$

$$\delta_0^2|\mathbf{R},\mathbf{m},\boldsymbol{\sigma}^2 \sim \mathcal{IG}\left(\xi+\frac{K_1}{2},\beta+\sum_{k=1}^{K_1}\frac{(m_k-m_0)^2}{2\sigma_k^2}\right). \quad (24)$$

3.3. Generation of samples according to $f(\kappa|\mathbf{R},\mathbf{Y})$

To obtain samples distributed according to $f(\kappa|\mathbf{R},\mathbf{Y})$, it is convenient to generate vectors distributed according to the joint distribution $f(\boldsymbol{\psi},\kappa|\mathbf{R},\mathbf{Y})$ with Gibbs moves. This step can be done using the joint distribution $f(\boldsymbol{\theta},\boldsymbol{\Phi}|\mathbf{Y})$ and the following procedure:

- **Generate samples according to** $f(\boldsymbol{\psi}|\mathbf{R},\kappa,\mathbf{Y})$
  This is done as follows:

$$\psi_k|\mathbf{R},\kappa,\mathbf{Y} \sim \mathcal{M}\left(\lambda_k,R_k\right), \quad k=1,\dots,K_2 \quad (25)$$

  where λ_k is the resultant length on the segment $\#k$:

$$\lambda_k = \arctan\left(\frac{R_0\sin\psi_0+\sum_{i\in I_{2,k}}\sin y_{2,i}}{R_0\cos\psi_0+\sum_{i\in I_{2,k}}\cos y_{2,i}}\right), \quad (26)$$

  and R_k is the corresponding direction given in (17).

- **Generate samples according to** $f(\kappa|\mathbf{R},\boldsymbol{\psi},\mathbf{Y})$
  The conditional posterior distribution of κ is:

$$f\left(\kappa|\mathbf{R},\boldsymbol{\psi},\mathbf{Y}\right) \propto \frac{1}{\kappa}\left[\frac{1}{\mathcal{I}_0(\kappa)}\right]^N \mathbf{1}_{\mathbb{R}^+}(\kappa)$$
$$\times \prod_{k=1}^{K_2}\left[\frac{1}{2\pi\mathcal{I}_0(R_0\kappa)}e^{\kappa R_k\cos(\psi_k-\lambda_k)}\right]. \quad (27)$$

Therefore, samples can be taken from $f(\kappa|\mathbf{R},\boldsymbol{\psi},\mathbf{Y})$ (by a simple Metropolis-Hastings procedure). The proposed distribution is a Gamma distribution $\kappa^\star \sim \mathcal{G}(a,b)$. This choice is motivated ([12, p. 40]) by the asymptotic behavior of the modified Bessel function of order 0 (i.e. $\mathcal{I}_0(x) \approx (2\pi x)^{-\frac{1}{2}}e^x$ when $x\to\infty$). The parameters a and b are chosen such that the mean of $\kappa^\star$ is the maximum likelihood estimate $\widehat{\kappa}_{\text{ML}}$ and β_κ^2 is an arbitrary variance:

$$a = \frac{\widehat{\kappa}_{\text{ML}}^2}{\beta_\kappa^2}, \qquad b = \frac{\widehat{\kappa}_{\text{ML}}}{\beta_\kappa^2}. \quad (28)$$

Using the likelihood function $f(\mathbf{y}_2|\mathbf{r}_2,\kappa,\boldsymbol{\psi})$, $\widehat{\kappa}_{\text{ML}}$ satisfies:

$$\frac{\mathcal{I}_1(\widehat{\kappa}_{\text{ML}})}{\mathcal{I}_0(\widehat{\kappa}_{\text{ML}})} = \frac{1}{N}\sum_{k=1}^{K_2}\sum_{i\in I_{2,k}}\cos\left(y_{2,i}-\psi_k\right). \quad (29)$$

The acceptance probability of the new state $\kappa^\star$ is:

$$\rho_{\kappa\to\kappa^\star} = \min\left(1,\frac{f\left(\kappa^\star|\mathbf{R},\boldsymbol{\psi},\mathbf{Y}\right)g_{a,b}(\kappa)}{f\left(\kappa|\mathbf{R},\boldsymbol{\psi},\mathbf{Y}\right)g_{a,b}(\kappa^\star)}\right). \quad (30)$$

3.4. Posterior distribution of P_ϵ

The hyperparameters P_ϵ, $\epsilon \in \mathcal{E}$, provide information about the correlation between the change locations in the two time series. It is useful to estimate them from their posterior distribution for practical applications. Straightforward computations yield the following Dirichlet posterior distribution for $\mathbf{P}$:

$$\mathbf{P}|\mathbf{R}, \mathbf{Y} \sim \mathcal{D}_4(S_\epsilon(\mathbf{R}) + \alpha_\epsilon). \tag{31}$$

4. SEGMENTATION OF SYNTHETIC WIND DATA

The simulations presented in this section have been obtained for 2 sequences with sample size $n = 250$. The change-point locations are $l_1 = (140)$ for the wind speed sequence and $l_2 = (60, 140)$ for the wind direction sequence. The parameters of the first sequence are $\mathbf{m} = [1.1, 0.6]^\top$ and $\sigma^2 = [0.05, 0.1]^\top$. The parameters of the second sequence are $\psi = \left[-\frac{\pi}{2}, -\frac{\pi}{6}, \frac{\pi}{3}\right]^\top$ and $\kappa = 2$. The fixed parameters and hyperparameters have been chosen as follows: $\nu = 2$ (as in [4]), $R_0 = 10^{-2}$ and $\psi_0 = 0$ (vague prior), $\xi = 1$ and $\beta = 100$ (vague hyperprior), $\alpha_\epsilon = \alpha = 1, \forall \epsilon \in \mathcal{E}$. The hyperparameters α_ϵ are equal, ensuring the Dirichlet distribution reduces to a uniform distribution. Moreover, the common value to the hyperparameters α_ϵ has been set to $\alpha = 1 \ll n$ in order to reduce the influence of this parameter in the posterior (31). The total number of runs for each Markov chain is $N_{\mathrm{MC}} = 6000$, including $N_{\mathrm{bi}} = 1000$ burn-in iterations. Thus only the last 5000 Markov chain output samples are used for the estimations. These values of N_{MC} and N_{bi} ensure convergence of the sampler as shown in [11].

4.1. Posterior distributions of the change-point locations

The first simulation compares joint segmentation and signal-by-signal segmentations. Figure 1 shows the posterior distributions of the change-point locations in the two time-series, obtained for 1D segmentation (top) and for joint segmentation (bottom). These results show that the Gibbs sampler (when applied to the second time-series) has not been able to detect the second change in the second sequence (top). Conversely, this change-point is detected much more precisely when using the joint segmentation procedure (bottom). This result can be explained as follows. The presence of a change in the first time series favors the detection of this change in the second time series. Other simulations are available in [11] and confirms this behavior.

4.2. Wind parameter estimation

The estimates of the wind parameters (speed scale and shape parameters, mean directions and concentration parameters) are useful in practical applications requiring signal reconstruction. This section studies the estimated posterior distributions for the parameters corresponding to the previous synthetic wind data.

The posterior distributions of the parameters m_k $(k = 1, \ldots, K_1)$ (conditioned upon $K_1 = 2$) are depicted in figures 2[1]. They are clearly in good agreement with the actual values of these parameters $\mathbf{m} = [1.1, 0.6]^\top$. The posterior distributions of the mean directions ψ_k $(k = 1, \ldots, K_2)$ (conditioned upon $K_2 = 3$) and the concentration parameter κ are depicted in figures 3 and 4. These posteriors are also in good agreement with the actual values of the parameters.

---

[1]Similar results have been obtained for the parameters σ_k^2, see [11].
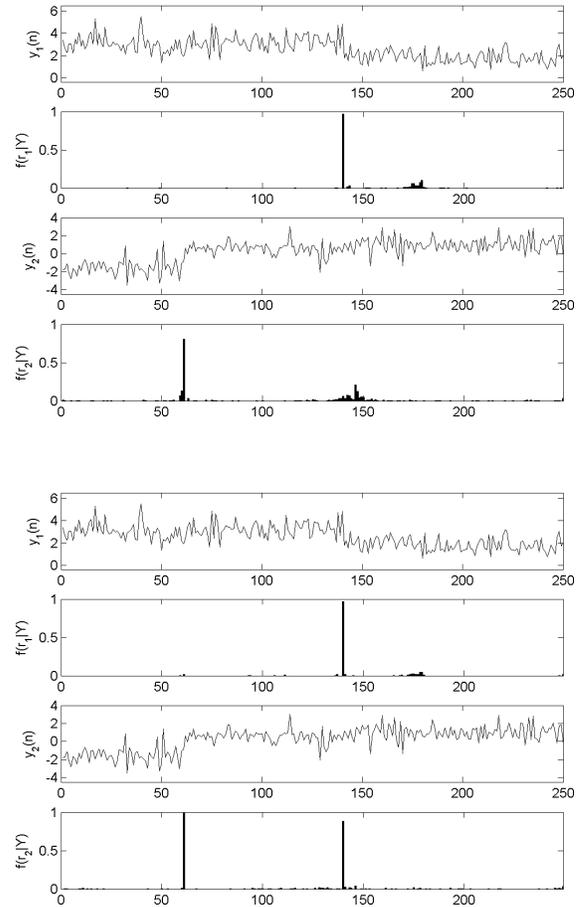

Fig. 1. Posterior distributions of the change-point locations for 1D (top) and joint segmentations (bottom) obtained after $N_{\mathrm{bi}} = 1000$ burn-in iterations and $N_r = 5000$ iterations of interest.
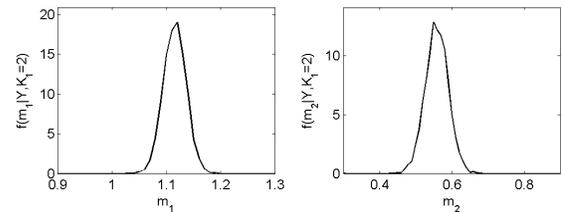

Fig. 2. Posterior distributions of the wind speed scale parameters m_k (for $k = 1, 2$) conditioned on $K_1 = 2$.

5. CONCLUSIONS

This paper studied a Bayesian sampling algorithm for segmenting wind speed and direction. The statistical properties of the wind speed and direction were described by lognormal and VM distributions with piecewise constant parameters. Posterior distributions of the unknown parameters provided estimates of the unknown param-
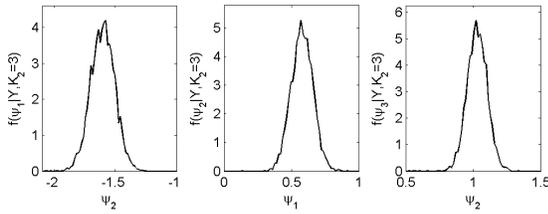
Fig. 3. Posterior distributions of the wind mean directions parameters ψ_k (for $k = 1, 2, 3$) conditioned on $K_2 = 3$.
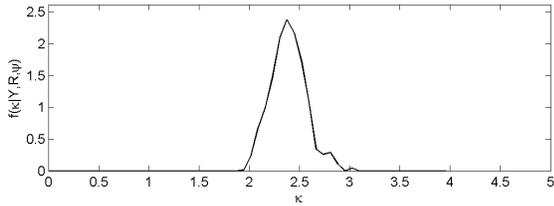

Fig. 4. Posterior distribution of the wind concentration parameter κ.

eters and their uncertainties. The proposed algorithm can be easily extended to other statistical models. For instance, joint segmentation of wave amplitude and direction described by Rayleigh and VM distributions was discussed. Simulation results conducted on synthetic signals illustrated the performance of the proposed methodology. This hierarchical Bayesian algorithm can handle other possible relationships between the two observed times series. For example, one typically has information 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 distributions adopted for the models and their parameters. Another important point is that information about the parameter uncertainties can be naturally extracted from the sampling strategy. This is typical of MCMC methods which explore the relevant parameter space by generating samples distributed according to the interesting posteriors. These samples can then be used to obtain confidence intervals and variances for the estimates of the unknown parameters.

6. REFERENCES

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

[2] C. Inclan and G. C. Tiao, "Use of cumulative sums of squares for retrospective detection of changes of variance," *J. Am. Stat. Assoc.*, vol. 89, pp. 913–923, 1994.

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

[4] 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.

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

[6] N. Dobigeon, J.-Y. Tourneret, and J. D. Scargle, "Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model," *IEEE Trans. Signal Processing*, vol. 55, no. 2, pp. 414–423, Feb. 2007.

[7] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1251–1263, April 2007.

[8] S. Reboul and M. Benjelloun, "Joint segmentation of the wind speed and direction," *Signal Processing*, vol. 86, no. 4, pp. 744–759, April 2006.

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

[10] F. Giraud and Z. M. Salameh, "Steady-state performance of a grid-connected rooftop hybrid wind-PhotoVoltaic power system with battery storage," *IEEE Trans. Energy Conversion*, vol. 16, no. 1, pp. 1–7, March 2001.

[11] N. Dobigeon and J.-Y. Tourneret, "Joint segmentation of wind speed and direction using a hierarchical model," *Computational Statistics & Data Analysis*, vol. 51, no. 12, pp. 5603–5621, Aug. 2007.

[12] K. V. Mardia and P. E. Jupp, *Directional statistics*. Chichester, England: Wiley, 2000.

[13] K. V. Mardia and S. El-Atoum, "Bayesian inference for the Von-Mises distribution," *Biometrika*, vol. 63, no. 1, pp. 203–206, April 1976.