

Joint segmentation of wind speed and direction using a hierarchical model

Nicolas Dobigeon^{*}, Jean-Yves Tournet

IRIT/ENSEEIH/TS&A, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France.

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. Extensions to other statistical models are also discussed. These models allow us to study other joint segmentation problems including segmentation of wave amplitude and direction. The performance of the proposed algorithms is illustrated with results obtained with synthetic and real data.

Key words: Change-points, Gibbs sampler, wind data.

1 Introduction

Segmentation has received considerable attention in the literature (Basseville and Nikiforov (1993); Brodsky and Darkhovsky (1993) and references therein). Segmentation methods use on-line or off-line algorithms. On-line algorithms quickly detect the occurrence of a change and ensure a fixed rate of false alarms before the change instant (Basseville and Nikiforov, 1993, p. 4). On-line algorithms observe the sample until a change is detected. This procedure explains the “on-line” or “sequential” terminology. Conversely, 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 Inflan and Tiao (1994); Rotondi (2002)). This paper focuses on the off-line class of algorithms. The main off-line segmentation strategies are based on the least squares principle (Lavielle and Moulines, 2000; Birgé and Massart, 2001), the maximum likelihood method or Bayesian strategies (Djurić, 1994; Fearnhead, 2005; Lavielle, 1998; Punsakaya et al., 2002). A common idea 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 Birgé and Massart (2001) 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 (Lavielle, 1998). The segmentation is then conducted using the posterior distribution of the change locations. Markov chain Monte Carlo (MCMC) methods are often considered when this 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 (Djurić, 1994) or minimum mean square error (MMSE) estimator (Lavielle, 1998; Tournet et al., 2003). Bayesian methods combined with MCMCs have useful properties for signal segmentation. In particular,

^{*} Corresponding author. Nicolas Dobigeon, IRIT/ENSEEIH/TS&A, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France. Tel.: +33 561 588 011; fax. +33 561 588 014.

Email address: nicolas.dobigeon@enseeiht.fr (Nicolas Dobigeon).

URL: <http://www.enseeiht.fr/~dobigeon> (Nicolas Dobigeon).

they allow estimation of the unknown model hyperparameters using hierarchical Bayesian algorithms (Dobigeon et al., 2007a).

This paper studies the problem of jointly segmenting the wind speed and direction, recently introduced in Reboul and Benjelloun (2006). 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. Note that the hyperparameters could also be estimated by coupling MCMCs with an expectation-maximization algorithm as in Kuhn and Lavielle (2004). The proposed hierarchical algorithm differs from the algorithms developed in Reboul and Benjelloun (2006). 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. 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 Bayesian 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 (Greene and Johnson, 1989; Skyum et al., 1996; Bogardi and Matyasovszky, 1996; Giraud and Salameh, 2001). Choice of one of these depends upon many factors, including site location or low/strong wind intensities. Unkasević et al. (1998); Garcia et al. (1998); Martin et al. (1999) describe more details. Note 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 (corresponding to a special case (see Section 7 for details)). For the lognormal assumption, the statistical properties of the wind speed on the k^{th} segment can be defined as follows:

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

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

- $\mathcal{LN}(m, \sigma^2)$ denotes a lognormal distribution with scale m and shape σ^2 (see Papoulis and Pillai (2002, p. 295)),
- K_1 is the number of segments in the speed sequence $\mathbf{y}_1 = [y_{1,1}, \dots, y_{1,n}]$.

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

$$y_{2,i} \sim \mathcal{M}(\psi_k, \kappa), \quad (2)$$

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

- $\mathcal{M}(\psi, \kappa)$ denotes a Von Mises distribution with mean direction ψ and concentration parameters κ (see Mardia and Jupp (2000, p. 36) for more details),
- K_2 is the number of segments in the direction sequence $\mathbf{y}_2 = [y_{2,1}, \dots, y_{2,n}]$.

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. The convergence properties of the sampler are investigated in Section 4. Some simulation results for synthetic and real wind data are presented in Sections 5 and 6. Extensions to other statistical models are discussed in Section 7. Conclusions are reported in Section 8.

2 Hierarchical Bayesian Model

The joint segmentation problem is based upon estimation of the unknown parameters (K_1, K_2) (numbers of segments), $\mathbf{l}_j = (l_{j,1}, \dots, l_{j,K_j})$, $j \in \{1, 2\}$ (change-point locations), $\mathbf{m} = [m_1, \dots, m_{K_1}]^\top$ (wind speed scale parameters), $\boldsymbol{\sigma}^2 = [\sigma_1^2, \dots, \sigma_{K_1}^2]^\top$ (wind speed shape parameters), $\boldsymbol{\psi} = [\psi_1, \dots, \psi_{K_2}]^\top$ (wind direction means) and κ (wind direction concentration parameter). A standard re-parameterization (Lavielle, 1998; Dobigeon et al., 2007a) introduces indicator variables $r_{j,i}$ ($j \in \{1, 2\}$, $i \in \{1, \dots, n\}$) such that:

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

with $r_{j,n} = 1$. This last condition ensures that the number of change-points and segments are equal for each signal, i.e. $K_j = \sum_{i=1}^n r_{j,i}$. 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 $\mathbf{r}_j = [r_{j,1}, \dots, r_{j,n}]$. Note that the parameter vector $\boldsymbol{\theta}$ belongs to a space whose dimension depends on K_1 and K_2 , i.e., $\boldsymbol{\theta} \in \Theta = \Theta_1 \times \Theta_2$ with $\Theta_1 = \{0, 1\}^n \times (\mathbb{R} \times \mathbb{R}^+)^{K_1}$ and $\Theta_2 = \{0, 1\}^n \times [-\pi, \pi]^{K_2} \times \mathbb{R}$. 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 $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2]^\top$. $f(\boldsymbol{\theta}|\mathbf{Y})$ 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:

$$\begin{aligned} f(\mathbf{y}_1|\mathbf{r}_1, \boldsymbol{\sigma}^2, \mathbf{m}) &= \prod_{k=1}^{K_1} \prod_{i \in I_{1,k}} \frac{1}{\sqrt{2\pi\sigma_k^2} y_{1,i}} e^{-\frac{(\log(y_{1,i}) - m_k)^2}{2\sigma_k^2}}, \\ &\propto \prod_{k=1}^{K_1} \left[\frac{1}{\sigma_k^2} \right]^{\frac{n_{1,k}}{2}} e^{-\frac{T_{1,k}^2 - 2m_k S_{1,k} + n_{1,k} m_k^2}{2\sigma_k^2}}, \end{aligned} \quad (3)$$

where \propto means ‘‘proportional to’’ and

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

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 given by:

$$\begin{aligned} f(\mathbf{y}_2 | \mathbf{r}_2, \boldsymbol{\psi}, \kappa) &= \prod_{k=1}^{K_2} \prod_{i \in I_{2,k}} \frac{1}{2\pi \mathcal{I}_0(\kappa)} e^{-\kappa \cos(y_{2,i} - \psi_k)}, \\ &\propto \prod_{k=1}^{K_2} \left[\frac{1}{\mathcal{I}_0(\kappa)} \right]^{n_{2,k}} e^{-\kappa \sum_{i \in I_{2,k}} \cos(y_{2,i} - \psi_k)}, \end{aligned} \quad (5)$$

where $\mathcal{I}_p(\cdot)$ is the modified Bessel function of the first kind and order p

$$\mathcal{I}_p(x) = \frac{1}{2\pi} \int_0^{2\pi} \cos(pu) e^{x \cos u} du, \quad (6)$$

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 can be expressed as:

$$f(\mathbf{Y} | \boldsymbol{\theta}) = f(\mathbf{y}_1 | \boldsymbol{\theta}_1) f(\mathbf{y}_2 | \boldsymbol{\theta}_2), \quad (7)$$

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

2.2 Parameter Priors

Correlations in the observed wind speed and direction signals are modeled by an appropriate prior distribution $f(\mathbf{R} | \mathbf{P})$, where $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2]^\top$ 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 $[\mathbf{r}_{1,i}, \mathbf{r}_{2,i}]^\top$ is independent of $[\mathbf{r}_{1,i'}, \mathbf{r}_{2,i'}]^\top$ for any $i \neq i'$. As a consequence, the indicator prior distribution expresses as:

$$\begin{aligned} f(\mathbf{R} | \mathbf{P}) &= \prod_{\boldsymbol{\epsilon} \in \mathcal{E}} P_\boldsymbol{\epsilon}^{S_\boldsymbol{\epsilon}(\mathbf{R})} \\ &= P_{00}^{S_{00}} P_{10}^{S_{10}} P_{01}^{S_{01}} P_{11}^{S_{11}}, \end{aligned} \quad (8)$$

where $P_\boldsymbol{\epsilon} = \Pr[\mathbf{r}_{1,i}, \mathbf{r}_{2,i}]^\top = \boldsymbol{\epsilon}$, $\mathbf{P} = \{P_{00}, P_{01}, P_{10}, P_{11}\} = \{P_\boldsymbol{\epsilon}\}_{\boldsymbol{\epsilon} \in \mathcal{E}}$ and $S_\boldsymbol{\epsilon}(\mathbf{R})$ is the number of times i such that $[\mathbf{r}_{1,i}, \mathbf{r}_{2,i}]^\top = \boldsymbol{\epsilon}$. 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 $[\mathbf{r}_{1,i}, \mathbf{r}_{2,i}]^\top = \boldsymbol{\epsilon}$ for all $i = 1, \dots, n$. 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.

Inverse-Gamma distributions are selected for shape parameter priors:

$$\sigma_k^2 | \nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right), \quad (9)$$

where $\mathcal{IG}(a, b)$ denotes the Inverse-Gamma distribution with parameters a and b , $\nu = 2$ (as in Punsakaya et al. (2002)) and γ is an adjustable hyperparameter. This is the *conjugate prior* for the wind speed variance which has been used successfully in Punsakaya et al. (2002) for Bayesian curve fitting.

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

$$m_k | \sigma_k^2, \delta_0^2 \sim \mathcal{N}(m_0, \sigma_k^2 \delta_0^2), \quad (10)$$

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 $f(\boldsymbol{\theta}|\mathbf{Y})$.

Bayesian inference for Von Mises distributions has been of great interest in applied statistics for several decades (Mardia and El-Atoum, 1976; Guttorp and Lockhart, 1988). The following priors were selected based on these results:

- conjugate Von Mises distributions are selected for wind mean directions:

$$\psi_k | \psi_0, R_0, \kappa \sim \mathcal{M}(\psi_0, \kappa R_0), \quad (11)$$

where R_0 and ψ_0 are fixed hyperparameters yielding a vague prior,

- an improper Jeffreys' prior is assigned to the concentration parameter κ :

$$f(\kappa) = \frac{1}{\kappa} \mathbf{1}_{\mathbb{R}^+}(\kappa), \quad (12)$$

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

2.3 Hyperpriors

The hyperparameter vector defined above and associated with the parameter priors is $\boldsymbol{\Phi} = (\mathbf{P}, \gamma, \delta_0^2)$. 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.

The prior distribution for the hyperparameter \mathbf{P} is a Dirichlet distribution with parameter vector $\boldsymbol{\alpha} = [\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}]^T$ defined on the simplex $\mathcal{P} = \{\mathbf{P} | \sum_{\epsilon \in \mathcal{E}} P_\epsilon = 1, P_\epsilon \geq 0\}$ 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 $\alpha_\epsilon = \alpha, \forall \epsilon \in \mathcal{E}$, the Dirichlet distribution reduces to the uniform distribution on \mathcal{P} .

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:

$$f(\gamma) = \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma), \quad \delta_0^2 | \xi, \beta \sim \mathcal{IG}(\xi, \beta). \quad (13)$$

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

$$\begin{aligned} f(\boldsymbol{\Phi} | \boldsymbol{\alpha}, \xi, \beta) &\propto \left(\prod_{\epsilon \in \mathcal{E}} P_\epsilon^{\alpha_\epsilon - 1} \right) \mathbf{1}_{\mathcal{P}}(\mathbf{P}) \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma) \\ &\times \frac{\beta^\xi}{\Gamma(\xi)(\delta_0^2)^\xi} \exp\left(-\frac{\beta}{\delta_0^2}\right) \mathbf{1}_{\mathbb{R}^+}(\delta_0^2), \end{aligned} \quad (14)$$

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:

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \int f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y}) d\boldsymbol{\Phi} \propto \int f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}) d\boldsymbol{\Phi}, \quad (15)$$

where

$$f(\boldsymbol{\theta}|\boldsymbol{\Phi}) = f(\mathbf{R}|\mathbf{P}) \left[\prod_{k=1}^{K_1} f(\sigma_k^2|\nu, \gamma) f(m_k|\sigma_k^2, \delta_0^2) \right] \times \left[\prod_{k=1}^{K_2} f(\psi_k|\kappa, R_0, \psi_0) \right] f(\kappa), \quad (16)$$

and $f(\mathbf{Y}|\boldsymbol{\theta})$ and $f(\boldsymbol{\Phi})$ are defined in (7) and (14). This hierarchical structure allows one to integrate out the nuisance parameters \mathbf{m} , σ^2 , $\boldsymbol{\psi}$ and \mathbf{P} from the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y})$, yielding:

$$f(\mathbf{R}, \gamma, \delta_0^2, \kappa|\mathbf{Y}) \propto \frac{1}{\kappa} \mathbf{1}_{\mathbb{R}^+}(\kappa) f(\delta_0^2|\xi, \beta) C(\mathbf{R}|\mathbf{Y}, \boldsymbol{\alpha}) \times \left[\frac{(\frac{\gamma}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \right]^{K_1} \left[\frac{1}{\mathcal{I}_0(\kappa)} \right]^N \times \prod_{k=1}^{K_1} \left[\frac{1}{2} \left(\frac{1}{1 + n_{1,k} \delta_0^2} \right)^{\frac{1}{2}} \Gamma\left(\frac{\nu + n_{1,k}}{2}\right) \right] \times \prod_{k=1}^{K_1} \left[\gamma + T_{1,k}^2 + \frac{m_0^2 - \mu_{1,k}^2 (1 + \delta_0^2 n_{1,k})}{\delta_0^2} \right]^{-\frac{n_{1,k}}{2}} \times \prod_{k=1}^{K_2} \left[\frac{\mathcal{I}_0(R_k \kappa)}{\mathcal{I}_0(R_0 \kappa)} \right], \quad (17)$$

with

$$\begin{cases} \mu_{1,k} = \frac{m_0 + \delta_0^2 S_{1,k}}{1 + \delta_0^2 n_{1,k}}, \\ R_k^2 = \left(R_0 \cos \psi_0 + \sum_{i \in I_{2,k}} \cos y_{2,i} \right)^2 \\ \quad + \left(R_0 \sin \psi_0 + \sum_{i \in I_{2,k}} \sin y_{2,i} \right)^2, \end{cases} \quad (18)$$

and

$$C(\mathbf{R}|\mathbf{Y}, \boldsymbol{\alpha}) = \frac{\prod_{\epsilon \in \{0,1\}^2} \Gamma(S_\epsilon(\mathbf{R}) + \alpha_\epsilon)}{\Gamma\left(\sum_{\epsilon \in \{0,1\}^2} (S_\epsilon(\mathbf{R}) + \alpha_\epsilon)\right)}. \quad (19)$$

The posterior distribution in (17) 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. Note that the proposed sampling method generates samples distributed according to the exact parameter posterior even if this posterior is known up to a normalization constant. This is a nice property of MCMCs since these constants are often difficult to compute (see Robert and Casella (1999, p. 234 or 329)).

3 A Gibbs Sampler for joint signal segmentation

Gibbs sampling is an iterative sampling strategy. It consists of generating random samples (denoted by $\cdot^{(t)}$ where t is the iteration index) distributed according to the conditional posterior distributions of each parameter. This paper

Algorithm 1: Gibbs sampler for abrupt change detection

- (1) For each time instant $i = 1, \dots, n - 1$, sample the local abrupt change configuration at time i $\left[\tilde{r}_{1,i}^{(t)}, \tilde{r}_{2,i}^{(t)} \right]^T$ from its conditional distribution given in (20),
 - (2) For segments $k = 1, \dots, K_1$, sample the wind speed shape parameters $\tilde{\sigma}_k^{2(t)}$ from their conditional posteriors given in (21),
 - (3) Sample the hyperparameter $\tilde{\gamma}^{(t)}$ from its posterior given in (23),
 - (4) For segments $k = 1, \dots, K_1$, sample the wind speed scale parameters $\tilde{m}_k^{(t)}$ from their conditional posteriors given in (24),
 - (5) Sample the hyperparameter $\tilde{\delta}_0^{2(t)}$ from its conditional posterior given in (25),
 - (6) For segments $k = 1, \dots, K_2$, sample the wind mean directions $\tilde{\psi}_k^{(t)}$ from their conditional posteriors given in (26),
 - (7) Sample the wind direction concentration parameter $\tilde{\kappa}^{(t)}$ from the conditional posterior given in (28) (see step 3.3 and Algorithm 2),
 - (8) *Optional step:* sample the hyperparameter $\tilde{\mathbf{P}}^{(t)}$ from the pdf in (33),
 - (9) Repeat steps (1)-(8).
-

samples the distribution $f(\mathbf{R}, \gamma, \delta_0^2, \kappa | \mathbf{Y})$ defined in (17) using the three step procedure outlined below. The main steps of Algorithm 1 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 $f\left(\left[r_{1,i}, r_{2,i}\right]^T | \gamma, \delta_0^2, \kappa, \mathbf{Y}\right)$. This vector is a random vector of Booleans in \mathcal{E} . Consequently, its distribution is fully characterized by the probabilities $\Pr\left[\left[r_{1,i}, r_{2,i}\right]^T = \epsilon | \gamma, \delta_0^2, \mathbf{Y}\right]$, $\epsilon \in \mathcal{E}$. Let \mathbf{R}_{-i} denote the matrix \mathbf{R} whose i^{th} column has been removed. Thus

$$\Pr\left[\left[r_{1,i}, r_{2,i}\right]^T = \epsilon | \mathbf{R}_{-i}, \gamma, \delta_0^2, \mathbf{Y}\right] \propto f(\mathbf{R}_i(\epsilon), \gamma, \delta_0^2 | \mathbf{Y}), \quad (20)$$

where $\mathbf{R}_i(\epsilon)$ is the matrix \mathbf{R} whose i^{th} column has been replaced by the vector ϵ . Equation in (20) yields a closed-form expression of the probabilities $\Pr\left[\left[r_{1,i}, r_{2,i}\right]^T = \epsilon | \mathbf{R}_{-i}, \gamma, \delta_0^2, \mathbf{Y}\right]$ after appropriate normalization.

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

To obtain samples distributed according to $f(\gamma, \delta_0^2 | \mathbf{R}, \mathbf{Y})$, it is very convenient to generate vectors distributed according to the joint distribution $f(\gamma, \delta_0^2, \boldsymbol{\sigma}^2, \mathbf{m} | \mathbf{R}, \mathbf{Y})$ by using Gibbs moves. Using the joint distribution (17), this step can be decomposed as follows:

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

Integrating the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi} | \mathbf{Y})$ with respect to the scale parameters m_k , the following results are obtained:

$$\sigma_k^2 | \mathbf{R}, \gamma, \delta_0^2, \mathbf{Y} \sim \mathcal{IG}(u_k, v_k), \quad k = 1, \dots, K_1 \quad (21)$$

with

$$\begin{cases} u_k = \frac{\nu + n_{1,k}}{2} \\ v_k = \frac{\gamma + T_{1,k}^2}{2} + \frac{m_0^2 - \mu_{1,k}^2 (1 + \delta_0^2 n_{1,k})}{2\delta_0^2}, \end{cases} \quad (22)$$

and

$$\gamma | \mathbf{R}, \boldsymbol{\sigma}^2 \sim \mathcal{G} \left(\frac{K_1 \nu}{2}, \sum_{k=1}^{K_1} \frac{1}{2\sigma_k^2} \right), \quad (23)$$

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:

$$m_k | \mathbf{R}, \boldsymbol{\sigma}^2, \delta_0^2, \mathbf{Y} \sim \mathcal{N} \left(\frac{m_0 + \delta_0^2 S_{1,k}}{1 + \delta_0^2 n_{1,k}}, \frac{\delta_0^2 \sigma_k^2}{1 + \delta_0^2 n_{1,k}} \right), \quad k = 1, \dots, K_1 \quad (24)$$

$$\delta_0^2 | \mathbf{R}, \mathbf{m}, \boldsymbol{\sigma}^2 \sim \mathcal{IG} \left(\xi + \frac{K_1}{2}, \beta + \sum_{k=1}^{K_1} \frac{(m_k - m_0)^2}{2\sigma_k^2} \right). \quad (25)$$

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 $f(\boldsymbol{\psi}, \kappa | \mathbf{R}, \mathbf{Y})$ by using Gibbs moves. This step can be done using the joint distribution of $f(\boldsymbol{\theta}, \boldsymbol{\Phi} | \mathbf{Y})$ and the following procedure:

- **Generate samples according to $f(\boldsymbol{\psi} | \mathbf{R}, \kappa, \mathbf{Y})$**

This is done as follows:

$$\psi_k | \mathbf{R}, \kappa, \mathbf{Y} \sim \mathcal{M}(\lambda_k, R_k), \quad k = 1, \dots, K_2 \quad (26)$$

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

$$\lambda_k = \arctan \left(\frac{R_0 \sin \psi_0 + \sum_{i \in I_{2,k}} \sin y_{2,i}}{R_0 \cos \psi_0 + \sum_{i \in I_{2,k}} \cos y_{2,i}} \right), \quad (27)$$

and R_k is the corresponding direction given in (18).

- **Generate samples according to $f(\kappa | \mathbf{R}, \boldsymbol{\psi}, \mathbf{Y})$**

The conditional posterior distribution of κ expresses as follows:

$$\begin{aligned} f(\kappa | \mathbf{R}, \boldsymbol{\psi}, \mathbf{Y}) &\propto \frac{1}{\kappa} \left[\frac{1}{\mathcal{I}_0(\kappa)} \right]^N \mathbf{1}_{\mathbb{R}^+}(\kappa) \\ &\times \prod_{k=1}^{K_2} \left[\frac{1}{2\pi \mathcal{I}_0(R_0 \kappa)} e^{\kappa R_k \cos(\psi_k - \lambda_k)} \right]. \end{aligned} \quad (28)$$

Therefore, samples can be taken from $f(\kappa | \mathbf{R}, \boldsymbol{\psi}, \mathbf{Y})$ (by a simple Metropolis-Hastings procedure) as summarized in Algorithm 2. The proposed distribution is a Gamma distribution $\kappa^* \sim \mathcal{G}(a, b)$. This choice is motivated (Mardia and Jupp (2000, 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 $x \rightarrow \infty$). The parameters a and b are chosen such that the mean of κ^* is the maximum likelihood estimate $\widehat{\kappa}_{\text{ML}}$ and β_κ^2 is an arbitrary variance:

$$a = \frac{\widehat{\kappa}_{\text{ML}}^2}{\beta_\kappa^2}, \quad b = \frac{\widehat{\kappa}_{\text{ML}}}{\beta_\kappa^2}. \quad (29)$$

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

$$\frac{\mathcal{I}_1(\widehat{\kappa}_{\text{ML}})}{\mathcal{I}_0(\widehat{\kappa}_{\text{ML}})} = \frac{1}{N} \sum_{k=1}^{K_2} \sum_{i \in I_{2,k}} \cos(y_{2,i} - \psi_k). \quad (30)$$

Algorithm 2: Sampling according to $f(\kappa|\mathbf{R}, \boldsymbol{\psi}, \mathbf{Y})$

- (1) Compute the maximum likelihood $\widehat{\kappa}_{\text{ML}}$ solution of (30),
 - (2) Propose κ^* according to $\mathcal{G}\left(\frac{\widehat{\kappa}_{\text{ML}}^2}{\widehat{\beta}_{\kappa}^2}, \frac{\widehat{\kappa}_{\text{ML}}}{\widehat{\beta}_{\kappa}}\right)$,
 - (3) if $(w_{\kappa} \sim \mathcal{U}_{[0,1]}) \leq \rho_{\kappa \rightarrow \kappa^*}$ (see (31)), then $\widetilde{\kappa}^{(t)} = \kappa^*$, else $\widetilde{\kappa}^{(t)} = \widetilde{\kappa}^{(t-1)}$.
-

The acceptance probability of the new state κ^* is:

$$\rho_{\kappa \rightarrow \kappa^*} = \min\left(1, \frac{f(\kappa^*|\mathbf{R}, \boldsymbol{\psi}, \mathbf{Y}) g_{a,b}(\kappa)}{f(\kappa|\mathbf{R}, \boldsymbol{\psi}, \mathbf{Y}) g_{a,b}(\kappa^*)}\right), \quad (31)$$

where $g_{a,b}(\cdot)$ denotes the pdf of the Gamma distribution with parameters (a, b) (Papoulis and Pillai, 2002, p. 87):

$$g_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbf{1}_{\mathbb{R}^+}(x). \quad (32)$$

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} :

$$\mathbf{P}|\mathbf{R}, \mathbf{Y} \sim \mathcal{D}_4(S_{\epsilon}(\mathbf{R}) + \alpha_{\epsilon}). \quad (33)$$

4 Convergence Diagnosis

The Gibbs sampler allows one to draw samples $(\mathbf{R}^{(t)}, \gamma^{(t)}, \delta_0^{2(t)}, \kappa^{(t)})$ asymptotically distributed according to $f(\mathbf{R}, \gamma, \delta_0^2, \kappa|\mathbf{Y})$. The change locations can then be estimated by the empirical average according to the MMSE principle:

$$\widehat{\mathbf{R}}_{\text{MMSE}} = \frac{1}{N_r} \sum_{t=1}^{N_r} \mathbf{R}^{(N_{\text{bi}}+t)}, \quad (34)$$

where N_{bi} is the number of burn-in iterations. However, two important questions have to be addressed: 1) When can one decide that the samples $\{\mathbf{R}^{(t)}\}$ are actually distributed according to the target distribution? 2) How many samples are necessary to obtain an accurate estimate of \mathbf{R} when using (34)? This section summarizes some works allowing us to determine appropriate values for parameters N_r and N_{bi} .

Multiple chains can be run with different initializations to define various convergence measures for MCMC methods (Robert and Richardson, 1998). The well-known between-within variance criterion has shown interesting properties for diagnosing convergence of MCMC methods. This criterion was initially studied by Gelman and Rubin (1992). It has been used in many studies (Robert and Richardson, 1998, p. 33), (Godsill and Rayner, 1998; Djurić and Chun, 2002). The main idea is to run M parallel chains of length N_r with different starting values. The dispersion of the estimates obtained from the different chains is then evaluated. The between-sequence variance B and within-sequence variance W for the M Markov chains are defined by

$$B = \frac{N_r}{M-1} \sum_{m=1}^M (\bar{\eta}_m - \bar{\eta})^2, \quad (35)$$

and

$$W = \frac{1}{M} \sum_{m=1}^M \frac{1}{N_r} \sum_{t=1}^{N_r} \left(\eta_m^{(t)} - \bar{\eta}_m \right)^2, \quad (36)$$

with

$$\begin{cases} \bar{\eta}_m = \frac{1}{N_r} \sum_{t=1}^{N_r} \eta_m^{(t)}, \\ \bar{\eta} = \frac{1}{M} \sum_{m=1}^M \bar{\eta}_m, \end{cases} \quad (37)$$

where η is the parameter of interest and $\eta_m^{(t)}$ is the t^{th} run of the m^{th} chain. The convergence of the chain can then be monitored by the so-called *potential scale reduction factor* $\hat{\rho}$ defined as (Gelman et al., 1995, p. 332):

$$\sqrt{\hat{\rho}} = \sqrt{\frac{1}{W} \left(\frac{N_r - 1}{N_r} W + \frac{1}{N_r} B \right)}. \quad (38)$$

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

The number of runs N_r , needed to obtain an accurate estimate of \mathbf{R} when using (34), is determined after the number of burn-in iterations has been adjusted. An *ad hoc* approach consists of assessing convergence via appropriate graphical evaluations (Robert and Richardson, 1998, p. 28). This paper computes a reference estimate (denoted as $\tilde{\mathbf{R}}$) from a large number of iterations. This approach ensures convergence of the sampler and good accuracy of the approximation in (34). The mean square error (MSE) between $\tilde{\mathbf{R}}$ and the estimate obtained after p iterations is then computed as

$$e_r^2(p) = \left\| \tilde{\mathbf{R}} - \frac{1}{p} \sum_{t=1}^p \mathbf{R}^{(N_{\text{bi}}+t)} \right\|^2. \quad (39)$$

Here N_r is calculated as the p ensuring the MSE $e_r^2(p)$ is below a predefined threshold.

5 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 $\mathbf{l}_1 = (80, 150)$ for the wind speed sequence and $\mathbf{l}_2 = (150)$ for the wind direction sequence. The parameters of the first sequence are $\mathbf{m} = [1.1, 0.6, 1.3]^T$ and $\boldsymbol{\sigma}^2 = [0.05, 0.1, 0.04]^T$. The parameters of the second sequence are $\boldsymbol{\psi} = \left[-\frac{\pi}{6}, \frac{\pi}{3}\right]^T$ and $\kappa = 2$. The fixed parameters and hyperparameters have been chosen as follows: $\nu = 2$ (as in Punsakaya et al. (2002)), $R_0 = 10^{-2}$ 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 (33). The total number of runs for each Markov chain is $N_{\text{MC}} = 6000$, including $N_{\text{bi}} = 1000$ burn-in iterations. Thus only the last 5000 Markov chain output samples are used for the estimations.

5.1 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 (left) and for joint segmentation (right). These results show that the Gibbs sampler (when applied to the second time-series) is slow to locate the change-point between close time indexes (left). However, this change-point is detected much more precisely when using the joint segmentation procedure (right).

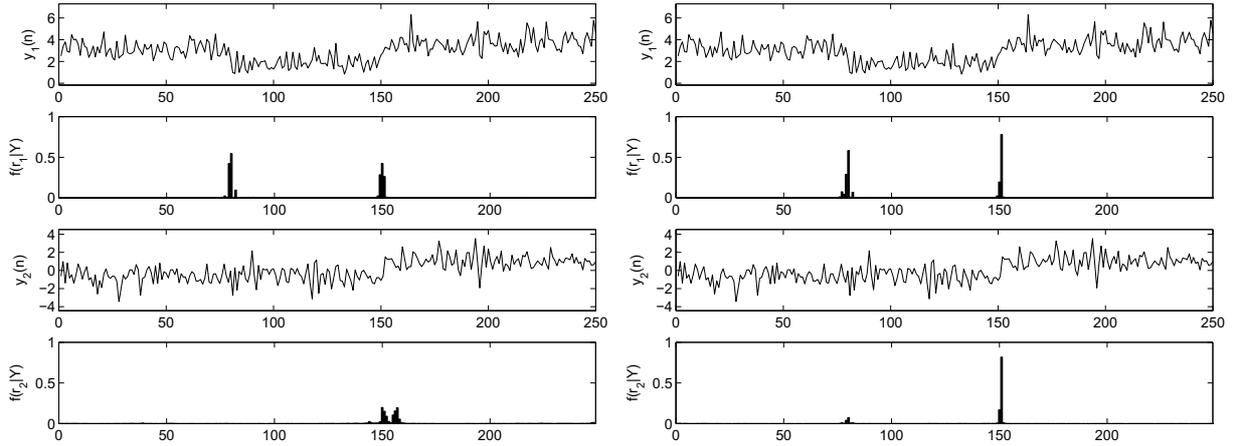


Fig. 1. Posterior distributions of the change-point locations for 1D (left) and joint segmentations (right) obtained after $N_{\text{bi}} = 1000$ burn-in iterations and $N_r = 5000$ iterations of interest.

5.2 Posterior distribution of the change-point numbers

An important problem is the estimation of the number of changes for the two time-series. The proposed algorithm generates samples $(\mathbf{R}^{(t)}, \gamma^{(t)}, \delta_0^{2(t)}, \kappa^{(t)})$, distributed according to the posterior distribution $f(\mathbf{R}, \gamma, \delta_0^2, \kappa | \mathbf{Y})$, and allows for model selection. Indeed, for each sample $\mathbf{R}^{(t)}$, the numbers of changes are $\widehat{K}_1^{(t)}(\mathbf{r}_1^{(t)}) = \sum_{i=1}^N r_{1,i}^{(t)}$ and $\widehat{K}_2^{(t)}(\mathbf{r}_2^{(t)}) = \sum_{i=1}^N r_{2,i}^{(t)}$. Figure 2 shows the means of \widehat{K}_1 and \widehat{K}_2 computed from the 5000 last Markov chain samples with the joint approach. The histograms have maximum values for $K_1 = 3$ and $K_2 = 2$ which correspond to the actual numbers of changes.

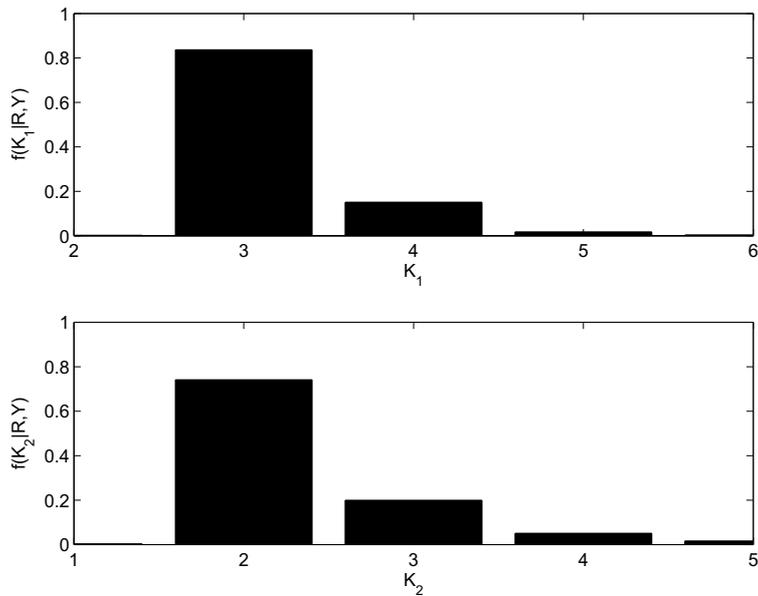


Fig. 2. Posterior distributions of the change-point numbers computed from $N_r = 5000$ iterations of interest.

5.3 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 figures plotted in this section all have been obtained by averaging the posteriors of $M = 20$ Markov chains (with $N_{\text{bi}} = 1000$ and $N_r = 5000$).

The posterior distributions of the parameters m_k and σ_k^2 ($k = 1, \dots, K_1$) (conditioned upon $K_1 = 3$) are depicted in figures 3 and 4 respectively. They are clearly in good agreement with the actual values of these parameters, $\mathbf{m} = [1.1, 0.6, 1.3]^T$ and $\boldsymbol{\sigma}^2 = [0.05, 0.1, 0.04]^T$. The posterior distributions of the mean directions ψ_k ($k = 1, \dots, K_2$) (conditioned upon $K_2 = 2$) and the concentration parameter κ are depicted in figures 5 and 6. These posteriors are also in good agreement with the actual values of the parameters.

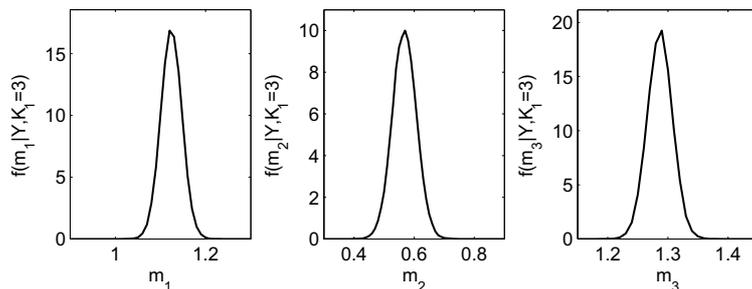


Fig. 3. Posterior distributions of the wind speed scale parameters m_k (for $k = 1, 2, 3$) conditioned on $K_1 = 3$.

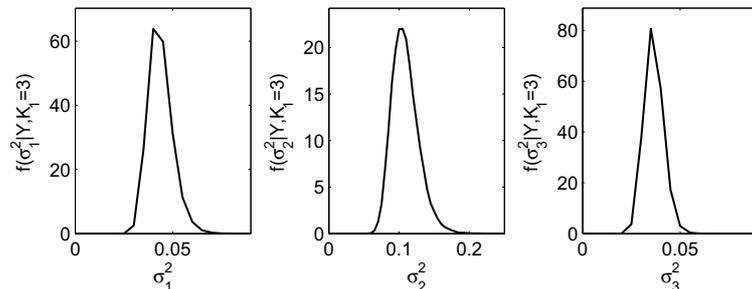


Fig. 4. Posterior distributions of the wind speed shape parameters σ_k^2 (for $k = 1, 2, 3$) conditioned on $K_1 = 3$.

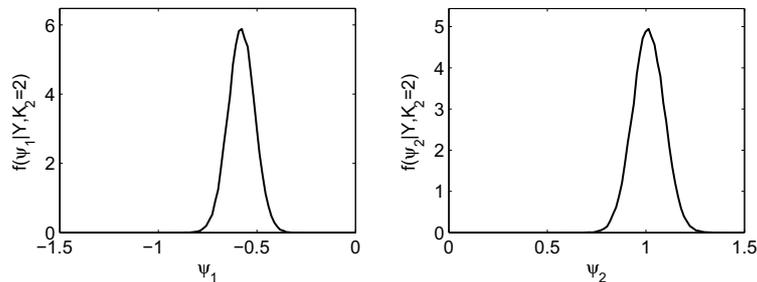


Fig. 5. Posterior distributions of the wind mean directions parameters ψ_k (for $k = 1, 2$) conditioned on $K_2 = 2$.

5.4 Hyperparameter estimation

The posterior distributions of P_ϵ in figure 7 correspond to the actual Dirichlet distribution in (33).

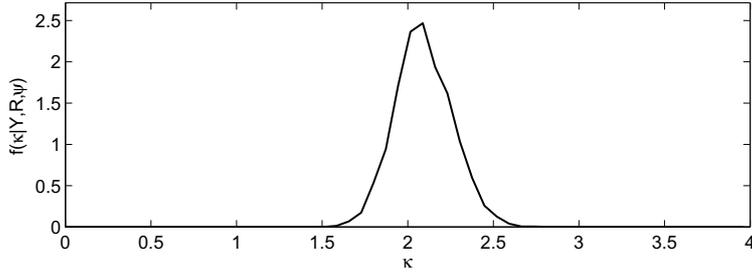


Fig. 6. Posterior distributions of the wind concentration parameter κ .

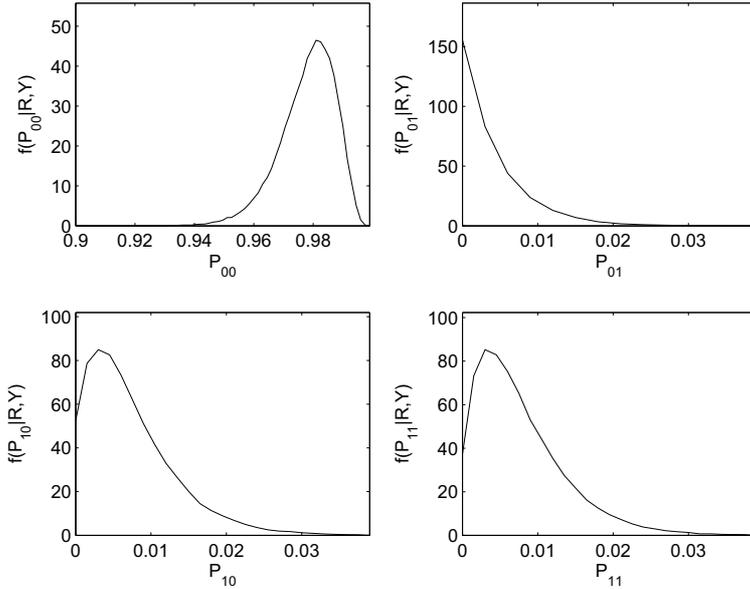


Fig. 7. Posterior distributions of the hyperparameters P_{00} , P_{01} , P_{10} and P_{11} .

5.5 Sampler convergence

The sampler convergence is monitored by checking the output of several parallel chains and by computing the potential scale reduction factor defined in (38). Different choices for parameter η can be considered for the proposed joint segmentation procedure. This paper monitors the convergence of the Gibbs sampler by checking the hyperparameter \mathbf{P} . As an example, the outputs of 5 chains for parameter P_{00} are depicted in figure 8. The chains clearly converge to similar values close to 1. This value agrees with the actual probability of the change configuration $\epsilon = (0, 0)$. The potential scale reduction factors are shown in Table 1 for parameters P_ϵ , $\epsilon \in \mathcal{E}$ and computed from Markov chains with $M = 20$. These values of $\sqrt{\hat{\rho}}$ confirm the good convergence of the sampler (a recommendation for convergence assessment is $\sqrt{\hat{\rho}} \leq 1.2$ (Gelman et al., 1995, p. 332)). The multivariate potential scale reduction factor (introduced in Brooks and Gelman (1998) and computed for the hyperparameter vector \mathbf{P}) is equal to $\sqrt{\hat{\rho}_{MV}} = 1.0008$. This result also confirms the convergence of the Gibbs sampler for the simulation example.

Table 1
Potential scale reduction factors of P_ϵ (computed from $M = 20$ Markov chains)

| P_ϵ | P_{00} | P_{01} | P_{10} | P_{11} |
|---------------------|----------|----------|----------|----------|
| $\sqrt{\hat{\rho}}$ | 0.9999 | 0.9999 | 0.9999 | 0.9999 |

The number of iterations N_r for an accurate estimate of \mathbf{R} (according to the MMSE principle in (34)) is determined by monitoring the MSE between a reference estimate $\tilde{\mathbf{R}}$ and the estimate obtained after p iterations (see (39)). Figure 9 shows the MSE between $\hat{\mathbf{R}}$ ($N_r = 10000$) and the estimate obtained after $N_r = p$ iterations (the number of burn-in

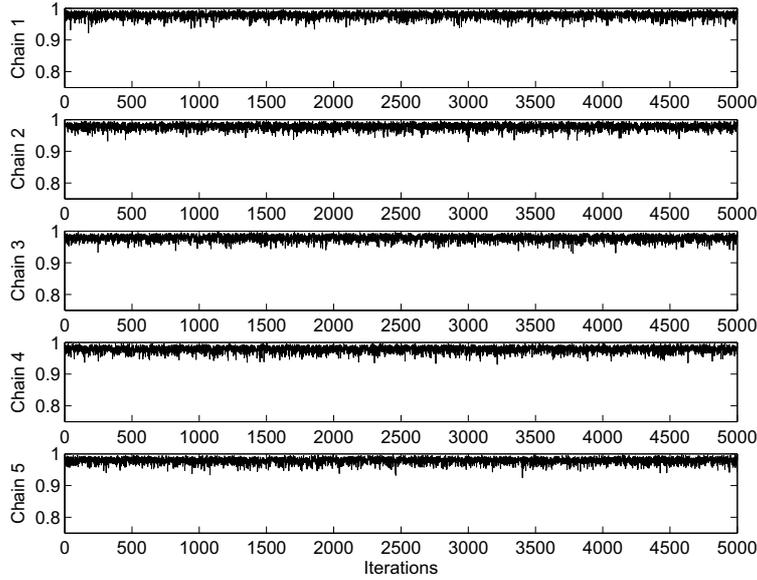


Fig. 8. Convergence assessment with five realizations of the Markov chain.

iterations is $N_{\text{bi}} = 1000$). Figure 9 indicates that $N_r = 5000$ is sufficient for this example to ensure an accurate estimate of the empirical average in (34).

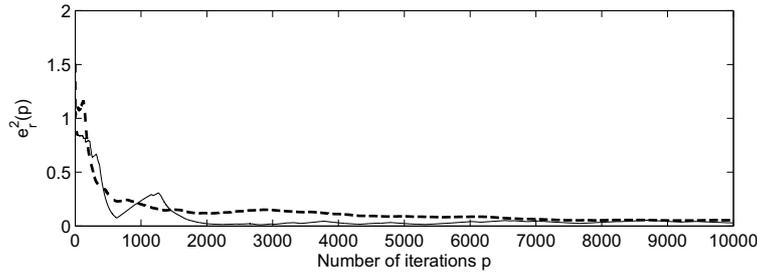


Fig. 9. MSE between the reference and estimated *a posteriori* change-point probabilities versus p (solid line). Averaged MSE computed from $M = 20$ chains (dashed line) ($N_{\text{bi}} = 1000$).

6 Segmentation of real wind data

This section demonstrates the performance of the algorithm using a real wind speed and direction record available online (Beardsley et al., 2001). The record has been extracted from a hydrographic survey of the Sea of Japan taken in 1999. This survey was carried out on the Research Vessel *R/V Professor Khromov*. The survey has already received much attention in the oceanography literature (Shcherbina et al., 2003; Talley et al., 2003, 2004). The analyzed data ($n = 166$) consists of a 1-minute sample of the wind speed and direction on July 25th, 1999. The $J = 2$ sequences are depicted in figure 10. The estimated number of change-points and their positions were obtained after $N_{\text{MC}} = 9000$ iterations (including a burn-in period of $N_{\text{bi}} = 1000$). The posterior distributions of the change-point locations are depicted in figure 10. Note the algorithm detects some close change-points. This behavior (also observed in Reboul and Benjelloun (2006)) can be explained by the slow evolution of the parameters from one segment to another.

Figure 11 shows the means of \hat{K}_1 and \hat{K}_2 computed from the 8000 last Markov chain samples with the joint approach. The histograms have maximum values for $K_1 = 10$ and $K_2 = 12$.

The segments for the two time-series can be obtained by keeping the largest values of the change-point posteriors

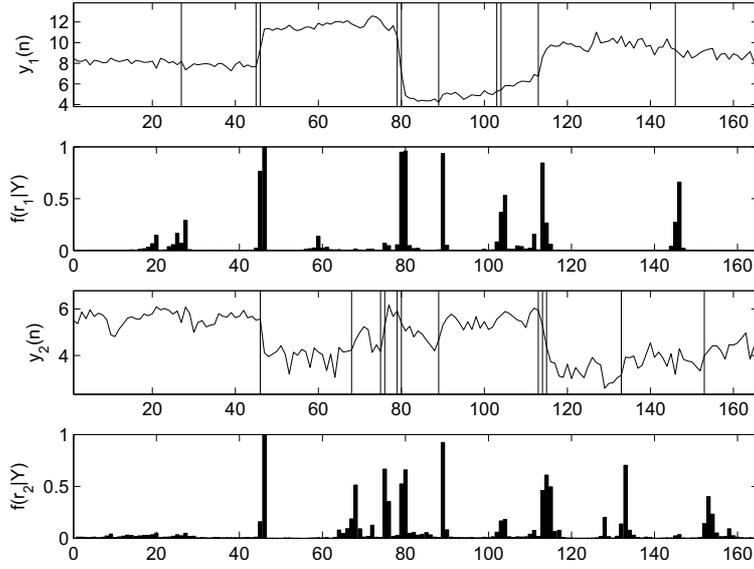


Fig. 10. Posterior distributions of the change-point locations and segmentation of real data obtained after $N_{\text{bi}} = 1000$ burn-in iterations and $N_r = 8000$ iterations of interest.

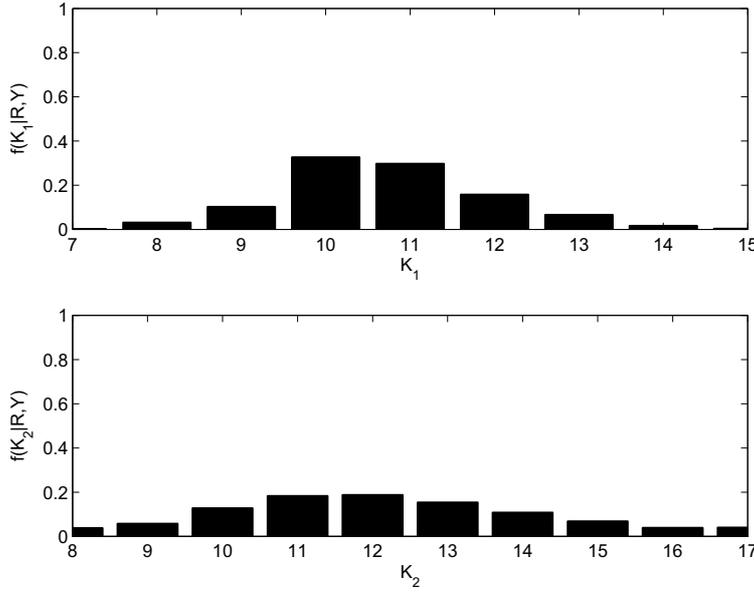


Fig. 11. Posterior distributions of the change-point numbers computed from $N_r = 8000$ iterations of interest.

(corresponding to the estimated change-point numbers \hat{K}_j , $j = 1, 2$). The results are represented by the vertical lines in figure 10.

7 Extension to other statistical models

This section discusses the application of the hierarchical Bayesian segmentation to other statistical models. Particular attention is given to the joint segmentation of time series distributed according to Rayleigh and Von Mises distributions. The use of the Rayleigh distribution has been recommended for wind turbine design (IEC 61400-1, 1994; Feijóo and Dornelas, 1999) and wave amplitude modeling (Liu et al., 1997). Thus, the algorithm studied here is appropriate for the

joint segmentation of wave amplitude/direction and wind turbine speed/direction.

7.1 Hierarchical model for wave amplitude/direction

The statistical properties of the wave amplitude can be modeled accurately by a Rayleigh distribution (e.g. see Papoulis and Pillai (2002, p. 148)) with piecewise constant parameters:

$$f(\mathbf{y}_1 | \mathbf{r}_1, \boldsymbol{\sigma}^2) \propto \prod_{k=1}^{K_1} \frac{1}{(\sigma_k^2)^{n_{1,k}}} \exp\left(-\frac{T_{1,k}^2}{2\sigma_k^2}\right), \quad (40)$$

where K_1 is the number of segments in the wave amplitude sequence $\mathbf{y}_1 = [y_{1,1}, \dots, y_{1,n}]$, $I_{1,k} = \{l_{1,k-1} + 1, \dots, l_{1,k}\}$ is the k^{th} segment, $T_{1,k}^2 = \sum_{i \in I_{1,k}} y_{1,i}^2$ and $\boldsymbol{\sigma}^2 = [\sigma_1^2, \dots, \sigma_{K_1}^2]^\top$.

The wave direction is assumed to be distributed according to a Von Mises distribution $\mathcal{M}(\psi_k, \kappa)$ whose mean direction ψ_k may change from a segment to another (similar to Section 1.1). Assume the wave amplitude and direction sequences are independent. Then the likelihood of the observed data matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2]^\top$ can be written:

$$f(\mathbf{Y} | \mathbf{R}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}, \kappa) \propto \prod_{k=1}^{K_1} \frac{1}{(\sigma_k^2)^{n_{1,k}}} \exp\left(-\frac{T_{1,k}^2}{2\sigma_k^2}\right) \times \prod_{k=1}^{K_2} \frac{1}{[\mathcal{I}_0(\kappa)]^{n_{2,k}}} \exp\left[-\kappa \sum_{i \in I_{2,k}} \cos(y_{2,i} - \psi_k)\right], \quad (41)$$

where \mathbf{R} stands for the indicator matrix introduced in Section 2.2 and $\boldsymbol{\psi} = [\psi_1, \dots, \psi_{K_2}]^\top$. Choose conjugate inverse-Gamma distributions $\mathcal{IG}(\nu, \frac{\gamma}{2})$ for the prior distributions for parameters σ_k^2 and an improper Jeffrey's prior distribution for γ . Then the nuisance parameters $\boldsymbol{\sigma}^2$, $\boldsymbol{\psi}$ and \mathbf{P} can be integrated out from the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi} | \mathbf{Y})$. Thus

$$f(\mathbf{R}, \gamma, \kappa | \mathbf{Y}) \propto \frac{1}{\kappa} \mathbf{1}_{\mathbb{R}^+}(\kappa) C(\mathbf{R} | \mathbf{Y}, \boldsymbol{\alpha}) \times \left[\frac{(\frac{\gamma}{2})^\nu}{\Gamma(\nu)}\right]^{K_1} \left[\frac{1}{\mathcal{I}_0(\kappa)}\right]^N \left[\prod_{k=1}^{K_2} \frac{\mathcal{I}_0(R_k \kappa)}{\mathcal{I}_0(R_0 \kappa)}\right] \times \left[\prod_{k=1}^{K_1} \Gamma(\nu + n_{1,k}) \left(\frac{\gamma + T_{1,k}^2}{2}\right)^{-n_{1,k}}\right], \quad (42)$$

where R_k^2 and $C(\mathbf{R} | \mathbf{Y}, \boldsymbol{\alpha})$ have been previously defined in (18) and (19) respectively.

The generation of samples $(\tilde{\mathbf{R}}^{(t)}, \tilde{\gamma}^{(t)}, \tilde{\kappa}^{(t)})$ (distributed according to the posterior of interest in (42)) can then be done using a Gibbs sampler algorithm similar to Algorithm 1.

7.2 Application to synthetic wave data

The previously discussed hierarchical Bayesian model has been used for segmenting jointly synthetic wave amplitude/direction data. The simulations presented here have been obtained for $J = 2$ sequences with sample size $n = 250$ as in Section 5. The change-point locations for the wave amplitude and direction are $\mathbf{l}_1 = (80, 150)$ and $\mathbf{l}_2 = (150)$. The Rayleigh parameters for the first sequence \mathbf{y}_1 are $\boldsymbol{\sigma}^2 = [0.09, 2.25, 0.49]^\top$. The wave direction sequence \mathbf{y}_2 has been generated as in Section 5 (i.e. with $\boldsymbol{\psi} = [-\frac{\pi}{6}, \frac{\pi}{3}]^\top$ and $\kappa = 2$). Figure 12 shows the estimated posterior distributions of the change-point locations for the two time-series. Note the estimates were computed with $N_r = 5000$ and with burn-in period $N_{\text{bi}} = 1000$. These posterior distributions are in very good agreement with the actual positions of the changes

in the two time series. The posterior distributions of the wave amplitude parameters σ_k^2 (for $k = 1, 2, 3$) conditioned on $K_1 = 3$ are depicted in figure 13. These posteriors illustrate the Gibbs sampler accuracy for estimating the wave amplitude parameters.

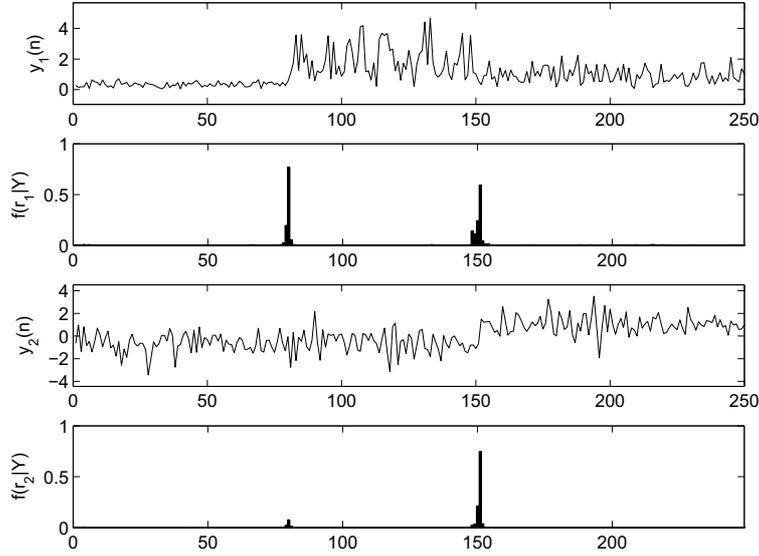


Fig. 12. Posterior distributions of the change-point locations obtained after $N_{\text{bi}} = 1000$ burn-in iterations and $N_r = 5000$ iterations of interest.

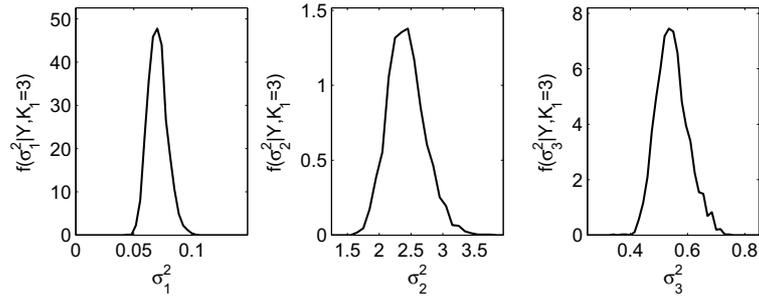


Fig. 13. Posterior distributions of the wave amplitude parameters σ_k^2 (for $k = 1, 2, 3$) conditioned on $K_1 = 3$.

7.3 Discussion

The results presented above show that the hierarchical Bayesian methodology (initially studied for the segmentation of wind speed and direction) can be extended to the segmentation of wave amplitude and direction. Other possible applications for which the Von Mises distribution has been used successfully include the analysis of directional movements of animals (such as fishes in Johnson et al. (2003)) or human observers (Graf et al., 2005). Finally, note that the joint segmentation of astronomical data used a similar strategy in Dobigeon et al. (2007b).

8 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 Von Mises distributions with piecewise constant parameters. Posterior distributions of the unknown parameters provided estimates of the unknown parameters 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 Von Mises distributions was discussed. Simulation results conducted on synthetic and real signals illustrated the performance of the proposed methodology.

The proposed 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.

Acknowledgments

The authors would like to thank the anonymous referees and the associate editor for their insightful comments improving the quality of the paper. They also would like to thank the Woods Hole Oceanographic Institution (Woods Hole, MA, USA) for supplying the meteorological measurements collected during the hydrographic cruise on the *R/V Professor Khromov*. Finally, the English grammar was enhanced with the help of an American colleague Professor Neil Bershad of UC Irvine in California.

References

- Basseville, M., Nikiforov, I. V., 1993. Detection of Abrupt Changes: Theory and Application. Prentice-Hall, Englewood Cliffs NJ.
- Beardsley, R., Dorman, C., Limeburner, R., Scotti, A., 2001. Japan/East Sea Home Page. Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.
URL http://www.whoi.edu/science/P0/japan_sea/
- Birgé, L., Massart, P., 2001. Gaussian model selection. *Jour. Eur. Math. Soc.* 3, 203–268.
- Bogardi, I., Matyasovszky, I., 1996. Estimating daily wind speed under climate change. *Solar Energy* 57 (3), 239–248.
- Brodsky, B., Darkhovsky, B., 1993. Nonparametric methods in change-point problems. Kluwer Academic Publishers, Boston MA.
- Brooks, S. P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7 (4), 434–455.
- Djurić, P. M., April 1994. A MAP solution to off-line segmentation of signals. In: *Proc. IEEE ICASSP-94*. Vol. 4. pp. 505–508.
- Djurić, P. M., Chun, J.-H., 2002. An MCMC sampling approach to estimation of nonstationary hidden markov models. *IEEE Trans. Signal Processing* 50 (5), 1113–1123.
- Dobigeon, N., Tournet, J.-Y., Davy, M., April 2007a. Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach. *IEEE Trans. Signal Processing* 55 (4), 1251–1263.
- Dobigeon, N., Tournet, J.-Y., Scargle, J. D., Feb. 2007b. Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model. *IEEE Trans. Signal Processing* 55 (2), 414–423.
- Fearnhead, P., June 2005. Exact Bayesian curve fitting and signal segmentation. *IEEE Trans. Signal Processing* 53 (6), 2160–2166.
- Feijóo, A. E., Dornelas, J. C. J. L. G., Dec. 1999. Wind speed simulation in wind farms for steady-state security assessment of electrical power systems. *IEEE Trans. Energy Conversion* 14 (4), 1582–1588.
- Garcia, A., Torres, J., Prieto, E., Francisco, A. D., 1998. Fitting wind speed distributions: A case study. *Solar Energy* 62 (2), 139–144.
- Gelman, A., Carlin, J. B., Stern, M. S., Rubin, D. B., 1995. *Bayesian Data Analysis*. Chapman & Hall, London.
- Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7 (4), 457–511.

- Giraud, F., Salameh, Z. M., March 2001. Steady-state performance of a grid-connected rooftop hybrid wind-PhotoVoltaic power system with battery storage. *IEEE Trans. Energy Conversion* 16 (1), 1–7.
- Godsill, S., Rayner, P., 1998. Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler. *IEEE Trans. Speech, Audio Proc.* 6 (4), 352–372.
- Graf, E. W., Warren, P. A., Maloney, L. T., 2005. Explicit estimation of visual uncertainty in human motion processing. *Vision Research* 45 (24), 3050–3059.
- Greene, D., Johnson, E., April 1989. A model of wind dispersal of winged or plumed seeds. *Ecology* 70 (2), 339–347.
- Guttorp, P., Lockhart, R. A., June 1988. Finding the location of a signal: a Bayesian analysis. *Journal of the American Statistical Association* 83 (402), 322–330.
- IEC 61400-1, 1994. Wind turbine generator systems. Part 1: Safety requirements.
- Inclan, C., Tiao, G. C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Am. Stat. Assoc.* 89, 913–923.
- Johnson, R. L., Simmons, M. A., McKinstry, C. A., Simmons, C. S., Cook, C. B., Brown, R. S., Tano, D. K., Thorsten, S. L., Faber, D. M., LeCaire, R., Francis, S., 2003. Strobe light deterrent efficacy test and fish behavior determination at Grand Coulee Dam third powerplant forebay. Tech. rep. for Bonneville Power Administration, U.S. Department of Energy (PNNL-14177), Pacific Northwest National Laboratory.
- Kuhn, E., Lavielle, M., 2004. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM Probability & Statistics* 8, 115–131.
- Lavielle, M., May 1998. Optimal segmentation of random processes. *IEEE Trans. Signal Processing* 46 (5), 1365–1373.
- Lavielle, M., Moulines, E., Jan. 2000. Least squares estimation of an unknown number of shifts in a time series. *Jour. of Time Series Anal.* 21 (1), 33–59.
- Liu, Y., Yan, X.-H., Liu, W., Hwang, P., May 1997. The probability density function of ocean surface slopes and its effects on radar backscatter. *J. of Physical Oceanography* 27, 782–797.
- Mardia, K. V., El-Atoum, S., April 1976. Bayesian inference for the Von-Mises distribution. *Biometrika* 63 (1), 203–206.
- Mardia, K. V., Jupp, P. E., 2000. *Directional statistics*. Wiley, Chichester, England.
- Martin, M., Cremades, L., Santabarbara, J., Feb. 1999. Analysis and modelling of time series of surface wind speed and direction. *Int. J. Climatol.* 19 (2), 197–209.
- Papoulis, A., Pillai, S. U., 2002. *Probability, random variables and stochastic processes*, 4th Edition. McGraw-Hill Editions, New-York.
- Punskaya, E., Andrieu, C., Doucet, A., Fitzgerald, W., March 2002. Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Trans. Signal Processing* 50 (3), 747–758.
- Reboul, S., Benjelloun, M., April 2006. Joint segmentation of the wind speed and direction. *Signal Processing* 86 (4), 744–759.
- Robert, C. P., Casella, G., 1999. *Monte Carlo Statistical Methods*. Springer-Verlag, New-York.
- Robert, C. P., Richardson, S., 1998. Markov Chain Monte Carlo methods. In: Robert, C. P. (Ed.), *Discretization and MCMC Convergence Assessment*. Springer Verlag, New York, pp. 1–25.
- Rotondi, R., Sept. 2002. On the influence of the proposal distributions on a reversible jump MCMC algorithm applied to the detection of multiple change-points. *Computational Statistics & Data Analysis* 40 (3), 633–653.
- Shcherbina, A. Y., Talley, L. D., Firing, E., Hacker, P., 2003. Near-surface frontal zone trapping and deep upward propagation of internal wave energy in the Japan/East Sea. *Journal of Physical Oceanography* 33 (4), 900–912.
- Skyum, P., Christiansen, C., Blaesild, P., 1996. Hyperbolic distributed wind, sea-level and wave data. *J. Coast. Res.* 12 (4), 883–889.
- Talley, L. D., Lobanov, V., Ponomarev, V., Salyuk, A., Tishchenko, P., Zhabin, I., Riser, S., Feb. 2003. Deep convection and brine rejection in the Japan Sea. *Geophysical research letters* 30 (4), 1159.
- Talley, L. D., Tishchenko, P., Luchin, V., Nedashkovskiy, A., Sagalaev, S., Kang, D.-J., Warner, M., Min, D.-H., May 2004. Atlas of Japan (East) Sea hydrographic properties in summer, 1999. *Progress in Oceanography* 61 (2–4), 277–348.
- Tourneret, J.-Y., Doisy, M., Lavielle, M., Sept. 2003. Bayesian retrospective detection of multiple changepoints corrupted by multiplicative noise. Application to SAR image edge detection. *Signal Processing* 83 (9), 1871–1887.
- Unkasević, M., Mališić, J., Tošić, I., 1998. On some new statistical characteristics of the wind “Koshava”. *Meteorol. Atmos. Phys.* 66 (1–2), 11–21.