Nonlinear regression using smooth Bayesian estimation

Abderrahim Halimi, Corinne Mailhes, Jean-Yves Tourneret

HAL Id: hal-01485021
https://hal.science/hal-01485021

Submitted on 8 Mar 2017

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 17108

The contribution was presented at ICASSP 2015 :
http://icassp2015.org/

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

NONLINEAR REGRESSION USING SMOOTH BAYESIAN ESTIMATION

Abderrahim Halimi, Corinne Mailhes and Jean-Yves Tourneret

[1] University of Toulouse, IRIT/INP-ENSEEIHT/TéSA, Toulouse, France
e-mail: {Abderrahim.Halimi, Corinne.Mailhes, Jean-Yves.Tourneret}@enseeiht.fr.

ABSTRACT

This paper proposes a new Bayesian strategy for the estimation of smooth parameters from nonlinear models. The observed signal is assumed to be corrupted by an independent and non identically (colored) Gaussian distribution. A prior enforcing a smooth temporal evolution of the model parameters is considered. The joint posterior distribution of the unknown parameter vector is then derived. A Gibbs sampler coupled with a Hamiltonian Monte Carlo algorithm is proposed which allows samples distributed according to the posterior of interest to be generated and to estimate the unknown model parameters/hyperparameters. Simulations conducted with synthetic and real satellite altimetric data show the potential of the proposed Bayesian model and the corresponding estimation algorithm for nonlinear regression with smooth estimated parameters.

Index Terms— Bayesian algorithm, Hamiltonian Monte-Carlo, MCMC, Parameter estimation, Radar altimetry.

1. INTRODUCTION

In many applications, the observed data are well described by a nonlinear function of a vector of parameters [1–4]. This paper aims at estimating these parameters from the observed data and a non-linear regression model. This can be achieved by using maximum likelihood based methods [5] or, equivalently, in the case of Gaussian noise, by using nonlinear least squares algorithms such as the Levenberg-Marquardt algorithm [6] and the natural gradient algorithm [7]. However, the resulting estimated parameters may be noisy and not convenient for physical interpretation. It is particularly true when the estimation procedure is applied to signals acquired at consecutive time instants and when the parameters of these signals have small variations from one instant to another. In this case, a lot of effort has been made to propose methods smoothing the estimated parameters. This smoothing is generally achieved by adding some time correlation prior resulting in the so-called Bayesian smoothing algorithms [8]. Many kinds of numerical approximations have also been proposed in the literature to handle the model nonlinearity while smoothing the parameters. For example, we can mention the extended and unscented Kalman filters [9, 10] and the Markov

chain Monte Carlo (MCMC) simulation based methods that will be considered in this paper (see [8] for more details about these methods).

The main contribution of this paper is the elaboration of a hierarchical Bayesian model that allows smooth estimation of parameters associated with different temporal signals. The observed signals are assumed to be corrupted by an additive, independent and colored Gaussian noise. The parameter of interest are assigned a prior enforcing smooth evolution between consecutive signals which improves their estimation. This prior is defined from the discrete Laplacian of the different parameters. It has shown increasing interest for many problems such as image deconvolution [11, 12], hyperspectral unmixing [13], medical imaging [14] and spectroscopy applications [15]. An algorithm is then proposed for estimating the unknown model parameters. However, the minimum mean square error (MMSE) and maximum a posteriori (MAP) estimators cannot be easily computed from the obtained joint posterior. The proposed algorithm alleviates this problem by generating samples distributed according to this posterior using Markov chain Monte Carlo (MCMC) methods. More precisely, we use a Hamiltonian Monte Carlo (HMC) algorithm since it has shown good mixing property for high-dimensional vectors [16].

The proposed estimation strategy is validated using synthetic signals as well as real satellite altimetric echoes. Altimetric radar echoes are defined as a nonlinear function of physical parameters [2, 4] (the epoch τ related to the distance satellite-observed scene, the significant wave height SWH and the signal amplitude P_u). Moreover, the noise corrupting these echoes is known to have an approximate Gaussian distribution which has been exploited to derive unweighted least squares (ULS) techniques [17–19] for parameter estimation. Note finally that the parameters of altimetric signals are often estimated echo by echo independently. However, recent works [20, 21] have shown the interest of considering echo's correlation which motivates the study of the proposed Bayesian approach.

The paper is organized as follows. Section 2 presents the hierarchical Bayesian model for the joint estimation of parameters varying smoothly from one observed signal to another while Section 3 details the proposed estimation algorithm. Simulation results performed on synthetic and real signals are

presented in Sections 4 and 5. Conclusions and future work are finally reported in Section 6.

2. HIERARCHICAL BAYESIAN MODEL

2.1. Observation model

In this work, we consider M successive signals $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_M)$ defined as noisy nonlinear functions of unknown parameters $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1^T, ..., \boldsymbol{\Theta}_M^T)^T$ following the model

$$\boldsymbol{y}_m = \boldsymbol{s}_m\left(\boldsymbol{\Theta}_m\right) + \boldsymbol{e}_m, \text{ with } \boldsymbol{e}_m \sim \mathcal{N}\left(\mu_m \boldsymbol{1}_K, \boldsymbol{\Sigma}_m\right) \quad (1)$$

where $\boldsymbol{y}_m$ and $\boldsymbol{s}_m$ are $(K \times 1)$ vectors representing the mth observed and noiseless signals, $\boldsymbol{\Theta}_m = [\theta_1(m), \cdots, \theta_H(m)]$ is a $1 \times H$ vector containing the H parameters of the mth signal, $\boldsymbol{e}_m$ is a Gaussian noise vector with a mean $\mu_m \boldsymbol{1}_K$, where $\boldsymbol{1}_K$ is a $(K \times 1)$ vector of 1, and a diagonal covariance matrix $\boldsymbol{\Sigma}_m = \text{diag}\left(\boldsymbol{\sigma}_m^2\right)$ with $\boldsymbol{\sigma}_m^2 = \left(\sigma_{m1}^2, \cdots, \sigma_{mK}^2\right)^T$ a $(K \times 1)$ vector. The proposed nonlinear regression method aims at estimating both signal and noise parameters with smothness constraints using the observation model (1).

2.2. Likelihood

The observation model defined in (1) and the Gaussian properties of the noise sequence $\boldsymbol{e}_m$ yield

$$f(\boldsymbol{y}_m|\boldsymbol{\Theta}_m, \mu_m, \boldsymbol{\Sigma}_m) \propto \frac{1}{\sqrt{\prod_{k=1}^K \sigma_{mk}^2}} \exp\left\{-\frac{1}{2}\boldsymbol{x}_m^T \boldsymbol{\Sigma}_m^{-1} \boldsymbol{x}_m\right\}$$
(2)

where $\propto$ means "proportional to", $\boldsymbol{x}_m = \boldsymbol{y}_m - \boldsymbol{s}_m - \mu_m \boldsymbol{1}_K$ and $\boldsymbol{s}_m\left(\boldsymbol{\Theta}_m\right)$ has been denoted by $\boldsymbol{s}_m$ for brevity. Assuming independence between the observations leads to

$$f(\boldsymbol{Y}|\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \propto \prod_{m=1}^M f(\boldsymbol{y}_m|\boldsymbol{\Theta}_m, \mu_m, \boldsymbol{\Sigma}_m). \quad (3)$$

The unknown parameters of the observation model (1) include the noise mean represented by an $(M \times 1)$ vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_M)^T$, the $(K \times M)$ matrix $\boldsymbol{\Lambda} = \left(\boldsymbol{\sigma}_1^2, \cdots, \boldsymbol{\sigma}_M^2\right)$ containing the noise variances associated with the considered M signals, and the $(M \times H)$ matrix $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_H]$ gathering the H parameters of the M signals.

2.3. Prior for signal parameters

The prior used for each parameter $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}$ enforces some smoothness property for the time evolution of this parameter. This can be done by constraining the derivative of this parameter to be small. In this paper, we propose to assign a Gaussian prior distribution to the second derivative of $\boldsymbol{\theta}_i$ as follows

$$f(\boldsymbol{\theta}_i|\epsilon_i^2) \propto \left(\frac{1}{\epsilon_i^2}\right)^{M/2} \exp\left(-\frac{1}{2\epsilon_i^2}\|\boldsymbol{D}\boldsymbol{\theta}_i\|^2\right) \quad (4)$$

for $i \in \{1, \cdots, H\}$, where ϵ_i^2 is an hyperparameter, $\|\cdot\|$ denotes the standard l_2 norm such that $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^T \boldsymbol{x}$ and $\boldsymbol{D}$ is the discrete Laplacian operator. This prior has been referred to as simultaneous autoregression (SAR) or conditional autoregression (CAR) models when used for image deconvolution [11, 12]. It has also been used for the spectral unmixing of hyperspectral images [13] or for medical imaging applications [14].

2.4. Prior for the noise parameters

The absence of knowledge about the noise mean can be considered by choosing the following Jeffreys prior $f(\boldsymbol{\mu}) \propto \boldsymbol{1}_M$. Considering the noise variances, one could estimate a diagonal matrix $\boldsymbol{\Sigma}_m$ for each observed signal. However, for the sake of simplicity, we assume that r consecutive signals have the same variances, i.e., $\sigma_{(n-1)r+1,k}^2 = \cdots = \sigma_{nr,k}^2$ for $n \in \{1, \cdots, N\}$, with $N = \frac{M}{r}$ (note that the general case is obtained by considering $r = 1$). Assuming prior independence between the noise variances $\sigma_{nr,k}^2$, the Jeffreys prior of $\boldsymbol{\Lambda}$ is defined as

$$f\left(\boldsymbol{\Lambda}\right) = \prod_{n=1}^N \prod_{k=1}^K \frac{1}{\sigma_{nr,k}^2} \mathbb{I}_{\mathbb{R}+}\left(\sigma_{nr,k}^2\right) \quad (5)$$

where $\mathbb{I}_A(.)$ is the indicator of the set A.

2.5. Hyperparameter priors

The hyperparameters $\epsilon_i^2, i \in \{1, \cdots, H\}$ are assigned a Jeffreys prior given by

$$f\left(\epsilon_i^2\right) = \frac{1}{\epsilon_i^2}\mathbb{I}_{\mathbb{R}+}\left(\epsilon_i^2\right) \quad (6)$$

which reflects the absence of knowledge about these coefficients [22]. Moreover, these hyperparameters are supposed to be a priori independent leading to

$$f\left(\boldsymbol{\epsilon}^2\right) = \prod_{i=1}^H f(\epsilon_i^2). \quad (7)$$

with $\boldsymbol{\epsilon}^2 = \left(\epsilon_1^2, \cdots, \epsilon_H^2\right)$.

2.6. Posterior distribution

The proposed Bayesian model depends on the parameters $\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}$ and hyperparameters $\boldsymbol{\epsilon}^2$. The joint posterior distribution of the unknown parameters and hyperparameter can be computed from the following hierarchical structure

$$f\left(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}|\boldsymbol{Y}\right) \propto f(\boldsymbol{Y}|\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda})f\left(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}\right) \quad (8)$$

with $f\left(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}\right) = f\left(\boldsymbol{\mu}\right)f\left(\boldsymbol{\Lambda}\right)\prod_{i=1}^H f(\boldsymbol{\theta}_i|\epsilon_i^2)$, after assuming a priori independence between the model parameters. The MMSE and MAP estimators associated with the posterior (8) are not easy to determine mainly because of the nonlinearity of the observation model. The next section presents an MCMC estimation algorithm that can be used to compute these MMSE and MAP estimators.

3. ESTIMATION ALGORITHM

The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest [23]. In this paper, we propose to use this principle to sequentially sample the parameters $\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}$ and $\boldsymbol{\epsilon}$. When a conditional distribution cannot be sampled directly, sampling techniques such as the HMC algorithm can be applied. This algorithm has shown better mixing properties than independent or random walk Metropolis-Hasting moves especially for high-dimensional problems [16, 24]. Therefore, it will be considered in the present paper since the variable to be sampled are of size $(M \times 1)$. The interested reader is invited to consult [16, 24] for more details about the HMC algorithm.

3.1. Sampling $\boldsymbol{\Theta}$

Using the likelihood (3) and the prior (4) leads to the following conditional distribution

$$f(\boldsymbol{\theta}_i | \boldsymbol{Y}, \boldsymbol{\Omega}_{\backslash i}) \propto \exp\left(-\sum_{m=1}^{M} \frac{\boldsymbol{x}_m^T \boldsymbol{\Sigma}_m^{-1} \boldsymbol{x}_m}{2} - \frac{\|\boldsymbol{D}\boldsymbol{\theta}_i\|^2}{2\epsilon_i^2}\right) \quad (9)$$

where $\boldsymbol{\Omega}_{\backslash i} = \{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \cdots, \boldsymbol{\theta}_H, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \epsilon_i^2\}$. The conditional distribution (9) has a complex form mainly because of the nonlinearity of the theoretical model with respect to the parameters $\boldsymbol{\Theta}$. This distribution is sampled using a HMC algorithm.

3.2. Noise parameters

Using (3) and the Jeffreys prior for the noise mean $\boldsymbol{\mu}$ defined in Section 2.4, it can be easily shown that the conditional distribution of $\boldsymbol{\mu}$ is the following Gaussian distribution

$$\mu_m | \boldsymbol{y}_m, \boldsymbol{\Theta}_m, \boldsymbol{\Sigma}_m \sim \mathcal{N}\left(\overline{\mu_m}, \frac{1}{\sum_{k=1}^{K} \sigma_{mk}^{-2}}\right) \quad (10)$$

with $\overline{\mu_m} = \frac{\sum_{k=1}^{K} \frac{y_{mk} - s_{mk}}{\sigma_{mk}^2}}{\sum_{k=1}^{K} \sigma_{mk}^{-2}}$. Similarly, using (3) and (5), it can be shown that

$$f(\boldsymbol{\Lambda} | \boldsymbol{Y}, \boldsymbol{\Theta}, \boldsymbol{\mu}) = \prod_{n=1}^{N} \prod_{k=1}^{K} f\left(\sigma_{nr,k}^2 | \boldsymbol{Y}_{:k}, \boldsymbol{\Theta}, \boldsymbol{\mu}\right) \quad (11)$$

and that $\sigma_{nr,k}^2 | \boldsymbol{Y}_{:k}, \boldsymbol{\Theta}, \boldsymbol{\mu}$ is distributed according to the following inverse-gamma distribution

$$\sigma_{nr,k}^2 | \boldsymbol{Y}_{:k}, \boldsymbol{\Theta}, \boldsymbol{\mu} \sim \mathcal{IG}\left(\frac{r}{2}, \beta\right) \quad (12)$$

with $\beta = \sum_{m=(n-1)r+1}^{nr} \frac{x_{mk}^2}{2}$. Note finally that the distributions (10) and (12) are easy to sample.

3.3. Hyperparameters

The conditional distribution of the hyperparameters ϵ_i^2 is an inverse-gamma distribution defined by $\epsilon_i^2 | \boldsymbol{\theta}_i \sim \mathcal{IG}\left(\frac{M}{2}, \frac{\|\boldsymbol{D}\boldsymbol{\theta}_i\|^2}{2}\right)$ that is easy to sample.

4. SIMULATION RESULTS

4.1. Altimetric signals

This section first considers synthetic altimetric signals defined by the physical Brown model [2] defined by

$$\begin{aligned} s_k &= \frac{P_u}{2}\left[1 + \text{erf}\left(\frac{kT - \tau T - \alpha\sigma_c^2}{\sqrt{2}\sigma_c}\right)\right] \\ &\times \exp\left[-\alpha\left(kT - \tau T - \frac{\alpha\sigma_c^2}{2}\right)\right] \end{aligned} \quad (13)$$

where $s_k = s(kT)$ is the kth data sample of the received signal, $\sigma_c^2 = \left(\frac{\text{SWH}}{2c}\right)^2 + \sigma_p^2$, $\text{erf}(t) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-z^2} dz$ is the Gaussian error function, α and σ_p^2 are two known satellite parameters, τ is the epoch expressed by samples (1 sample ≈ 46 cm), c is the speed of light and T is the time resolution. Note that the discrete altimetric echo is gathered in the $(K \times 1)$ vector $\boldsymbol{s} = (s_1, \cdots, s_K)^T$, where $K = 128$ samples.

The altimetric echoes are corrupted by a speckle noise whose influence is reduced by averaging (on-board the satellite) a sequence of L consecutive echoes. Considering the central limit theorem and using the fact that the averaging is conducted on a large number of echoes, the resulting noise sequence is approximated by a Gaussian distribution. This approximation is largely adopted in the altimetric community as shown in [25–27] and in the well known LS estimation algorithms used in [18, 20, 21, 28]. Therefore, the observation model (1) and the proposed estimation strategy are well adapted for the processing of altimetric echoes. Note that we consider $H = 3$, $M = 500$ echoes, $\boldsymbol{\Theta}_{m:} = [\text{SWH}(m), \tau(m), P_u(m)]$ and that the parameters generally belong to the following intervals of realistic values $\text{SWH} \in [0, 50]$ m, $\tau \in [5, 70]$ and $P_u > 0$. Note finally that we consider the same noise covariance for $r = 20$ successive echoes. Indeed, after averaging L echoes, the altimeter delivers 20 averaged echoes per second that will have the same noise covariance matrix.

4.2. Results on synthetic data

The proposed strategy (denoted by SBMC for smooth Bayesian MC) is first studied when considering 500 correlated altimetric echoes. This correlation is introduced by considering a smooth evolution of the altimetric parameters. More precisely, the synthetic parameters have been chosen as follows $\text{SWH}(m) = 2.5 + 2\cos(0.07m)$, $\tau(m) = 27 + 0.02m$ if $m < 250$ and $\tau(m) = 32 - 0.02m$ if $m \geq 250$, and $P_u(m) = 158 + 0.05\sin(0.1m)$, where m denotes the echo number. The synthetic echoes are corrupted by a speckle noise resulting from the averaging of $L = 90$ echoes. The SBMC is compared to the state of the art ULS algorithm described in [17, 18]. Fig. 1 shows the actual parameter values and the estimated ones by considering the ULS and the SBMC algorithms for 100 echoes. SBMC provides better

results than ULS due to the smoothness constraints enforced by the prior (4). Table 1 confirms this result since we obtain better standard-deviations (STDs) with the proposed approach. Note finally that the SBMC shows reduced bias (except for P_u) mainly because it exploits the fact that the noise is colored contrary to ULS.
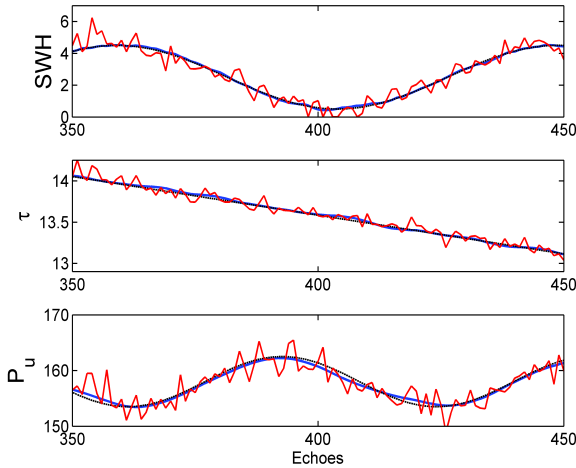
Fig. 1. Comparison of the actual parameters (black dashed line) with the estimated ones using the ULS algorithm (red line) and the proposed SBMC algorithm (blue line).

Table 1. Parameter biases and STDs on synthetic data (500 echoes).

		SWH (cm)	τ (cm)	P_u
Bias	ULS	3.18	1.11	-0.01
	SBMC	-0.02	-0.13	-0.19
STD	ULS	44.7	6.1	1.91
	SBMC	5.74	1.8	0.52

5. RESULTS ON REAL JASON-2 DATA

This section illustrates the performance of the proposed SBMC algorithm when applied to a real Jason-2 dataset. The considered data lasts 36 minutes and consists of 43000 real echoes. Fig. 2 shows 500 estimated parameters when considering the ULS and SBMC algorithms. The ULS estimates suffers from the noise corrupting the echoes while SBMC provides smoother estimates that are physically more consistent. Moreover, SBMC appears to be more robust to outliers as illustrated for the estimate #390 of P_u. Note that the estimated SWH is slightly larger for SBMC when compared to ULS. This difference is mainly due to the i.i.d. noise assumption used in ULS that is not in adequation with noise correlations as already discussed in [21, 27]. Table 2 shows a

good adequation between the means of the estimated parameters for both ULS and SBMC (except for SWH). Moreover, the estimated STDs obtained with SBMS are smaller than for ULS which is of great importance for many practical applications related to oceanography such as bathymetry.
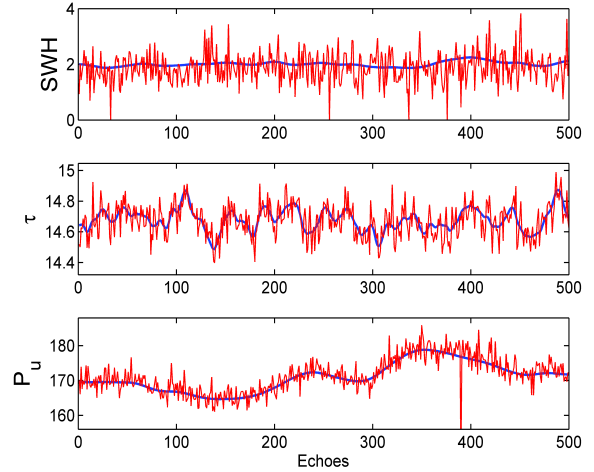
Fig. 2. Estimated parameters using the ULS (red line) and SBMC algorithms (blue line).

Table 2. Parameter means and STDs for real Jason-2 data (43000 echoes).

		SWH (cm)	τ (cm)	P_u
Mean	ULS	237	14.67	167.81
	SBMC	270	14.69	167.53
STD	ULS	53.4	10.41	5.9
	SBMC	4.3	5.22	4.82

6. CONCLUSIONS

This paper introduced a Bayesian model for smooth estimation of parameters associated with nonlinear models. The proposed model considers an appropriate prior distribution enforcing a smooth temporal evolution of the parameters of interest. Due to the complexity of the resulting joint posterior distribution, an MCMC procedure (based on a hybrid Gibbs sampler) was investigated to sample the posterior of interest and to approximate the Bayesian estimators of the unknown parameters using the generated samples. The proposed SBMC algorithm showed good performance and improved the quality of the estimated parameters when applied to both synthetic and real altimetric signals. It was also shown to be robust to parameter outliers. Future work includes the consideration of optimization algorithms for solving the proposed nonlinear regression problem with a reduced computational cost.

7. REFERENCES

[1] B. W. Hapke, "Bidirectional reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.

[2] G. Brown, "The average impulse response of a rough surface and its applications," *IEEE Trans. Antennas and Propagation*, vol. 25, no. 1, pp. 67–74, Jan. 1977.

[3] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, 2011.

[4] A. Halimi, C. Mailhes, J.-Y. Tourneret, P. Thibaut, and F. Boy, "Parameter estimation for peaky altimetric waveforms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1568–1577, March 2013.

[5] Harry L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I*, Wiley, New York, 1968.

[6] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, Massachusetts, 1995.

[7] S. Amari and S. C. Douglas, "Why natural gradient?," in *Proc. IEEE ICASSP-98*, May 1998, vol. 2, pp. 1213–1216.

[8] S. Särkkä, *Bayesian Filtering and Smoothing*, Cambridge University Press, Cambridge, UK, 2013.

[9] S. Haykin, *Kalman Filtering and Neural Networks*, John Wiley & Sons, Inc., New York, USA, 2001.

[10] F. Gustafsson and G. Hendeby, "Some relations between extended and unscented kalman filters," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 545–555, Feb 2012.

[11] P. Campisi and K. Egiazarian, *Blind image deconvolution: theory and applications*, Taylor & Francis, cop. 2007, Boca Raton, FL, 2007.

[12] R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, Dec. 2006.

[13] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Hyperspectral unmixing with l_q regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6793–6806, Nov. 2014.

[14] L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu, "Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach," *IEEE Trans. Image Process.*, vol. 32, no. 5, pp. 821–837, May 2013.

[15] V. Mazet, "Joint bayesian decomposition of a spectroscopic signal sequence," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 181–184, March 2011.

[16] S. Brooks, A. Gelman, G. L . Jones, and X.-L. Meng, *Handbook of Markov chain Monte Carlo*, ser. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2011.

[17] L. Amarouche, P. Thibaut, O. Z. Zanife, J.-P. Dumont, P. Vincent, and N. Steunou, "Improving the Jason-1 ground retracking to better account for attitude effects," *Marine Geodesy*, vol. 27, no. 1-2, pp. 171–197, Aug. 2004.

[18] A. Halimi, C. Mailhes, J.-Y. Tourneret, P. Thibaut, and F. Boy, "A semi-analytical model for delay/Doppler altimetry and its estimation algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4248–4258, July 2014.

[19] A. Halimi, C. Mailhes, J.-Y. Tourneret, T. Moreau, and F. Boy, "Including antenna mispointing in a semi-analytical model for delay/Doppler altimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 598–608, Feb. 2015.

[20] S. Maus, C. M. Green, and J. D. Fairhead, "Improved ocean-geoid resolution from retracked ERS-1 satellite altimeter waveforms," *Geophys. J. Int.*, vol. 134, no. 1, pp. 243–253, Feb. 1998.

[21] D. T. Sandwell and W. H. F. Smith, "Retracking ERS-1 altimeter waveforms for optimal gravity field recovery," *Geophys. J. Int.*, vol. 163, no. 1, pp. 79–89, Oct. 2005.

[22] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[23] C. P. Robert, *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*, Springer Texts in Statistics. Springer-Verlag, New York, 2 edition, 2007.

[24] A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," in *ArXiv e-prints*, Jun. 2014.

[25] C. Martin-Puig and G. Ruffini, "SAR altimeter retracker performance bound over water surfaces," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Cape Town, South Africa, July 12-17, 2009, pp. 449 –452.

[26] O. Germain and G. Ruffini, "A revisit to the GNSS-R code range precision," in *Proc. GNSS-R*, Noordwijk, The Netherlands, June 14-15, 2006.

[27] A. Halimi, C. Mailhes, and J.-Y. Tourneret, "Cramér-Rao bounds and estimation algorithms for delay/Doppler and conventional altimetry," in *Proc. EUSIPCO*, Marrakech-Marocco, Sept. 9-13 2013.

[28] C. Martin-Puig, P. Berry, R. Smith, C. Gommenginger, G. Ruffini, P. Cipollini, L. Stenseng, Andersen O., P.D. Cotton, J. Benveniste, and S. Dinardo, "SAR altimetry over water surfaces," in *Oceans from Space*, Venice, Italy, April 2010.