BAYESIAN ESTIMATION OF ALTIMETER ECHO PARAMETERS

J. Severini[(1)], C. Mailhes[(1)], P. Thibaut[(2)] and J.-Y. Tourneret[(1)]

[(1)] University of Toulouse, IRIT-ENSEEIHT-TESA , Toulouse, France
[(2)] Collecte Localisation Satellite (CLS), Ramonville Saint-Agne, France
{jerome.severini,corinne.mailhes,jean-yves.tourneret}@enseeiht.fr, pierre.thibaut@cls.fr

ABSTRACT

This paper studies a Bayesian algorithm for estimating the parameters associated to Brown's model. The joint posterior distribution of the unknown parameter vector (amplitude, epoch and significant wave height) associated to this model is derived. This posterior is too complex to obtain closed form expressions of the minimum mean square error and the maximum a Posteriori estimators. We propose to sample according to this distribution using an hybrid Metropolis within Gibbs algorithm. The simulated samples are then used to estimate the unknown parameters of Brown's model. The proposed strategy provides better estimations than the standard maximum likelihood estimator at the price of an increased computational cost.

Index Terms— Altimetry, Bayesian estimation, Gibbs sampler, Metropolis-Hastings algorithm.

1. INTRODUCTION

The retracking algorithm is a key algorithm in altimetric ground processing chains. This algorithm estimates ocean parameters from the measured waveforms. It has a major impact on the determination of the so-called sea state bias. Several teams have been working intensively to characterize and refine retracking. Recent papers have shown the very good agreement between retracking results obtained with Topex, Poseidon-1, Poseidon-2 and Envisat altimeters. A common idea in most actual retracking strategies is to estimate the unknown ocean parameters by comparing the measured altimetric waveform with a return power model according to least square estimators derived from the maximum likelihood principle. Because consecutive altimetric waveforms are representative of continuous ocean processes, it seems interesting to introduce prior knowledge when estimating the ocean parameters (whereas waveforms are processed independently from the previous ones in actual retracking procedures). This paper investigates a Bayesian approach which combines information coming from the data (contained in the likelihood) and prior knowledge regarding the unknown parameters (contained in the parameter prior distribution). The resulting posterior distribution for the unknown parameters is too complex to derive the standard Bayesian estimators. Monte Carlo Markov chains (MCMCs) are then used to generate samples distributed according to this posterior. The unknown parameters of Brown's model are finally estimated using these generated samples.

The paper is organized as follows: Section 2 recalls the conventional Brown's model. The posterior distribution of the corresponding unknown parameters is derived in Section 3. The MCMC method proposed to generate samples distributed according to this posterior is presented in Section 4. Simulations results depicted in Section 5 illustrate the performance of the proposed estimation strategy. Conclusions and perspectives are reported in Section 6.

2. DATA MODEL AND PROBLEM FORMULATION

According to Brown's model [1], altimeter waveforms are characterized by four parameters: the amplitude P_u (or equivalently the backscatter coefficient σ_0), the epoch τ, the significant wave height SWH and the off-nadir angle ξ. The resulting altimeter waveform denoted s_k can be written

$$s_k = \frac{P}{2}\left[1+\operatorname{erf}\left(\frac{k-\tau-\beta\sigma_c^2}{\sqrt{2}\sigma_c}\right)\right]\exp\left[-\beta\left(k-\tau-\frac{\beta\sigma_c^2}{2}\right)\right], \tag{1}$$

where $\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-z^2}\,dz$ is the Gaussian error function, c denotes the light speed,

$$\begin{cases} P = P_u \, \exp\left(-\frac{4}{\gamma}\sin^2\xi\right), \\ \beta = \alpha\left[\cos(2\xi) - \frac{\sin^2(2\xi)}{\gamma}\right], \\ \sigma_c^2 = \frac{\text{SWH}^2}{4c^2} + \sigma_p^2, \end{cases}$$

and α, γ, σ_p^2 are three known parameters depending on the satellite. In practice, there are two kinds of noises corrupting the altimeter waveforms: an additive instrumental thermal noise and a multiplicative speckle noise. The effect of the additive noise can be mitigated by subtracting from the observations an average of the first waveform samples. This paper assumes that this operation does not affect the estimation of the other parameters. In this case, the observed altimeter waveform can be written

$$y_k = s_k n_k, \qquad k = 1, ..., K, \tag{2}$$

where K is the number of observed samples, n_k is a multiplicative speckle noise distributed according to a gamma distribution $\mathcal{G}(L, L)$ (using the notation of [2, p. 451]) and L is the so-called number of looks. Actual retracking procedures estimate the unknown parameter vector $\boldsymbol{\theta} = (P_u, \tau, \text{SWH}, \xi)$ of (1) by using the maximum likelihood (ML) principle [3]. The properties of the maximum likelihood estimator (MLE) for estimating the Brown's model parameters have been studied in [4]. A comparison between the mean square errors (MSEs) of the MLE with the corresponding Cramér-Rao lower bounds has shown there is some space for improving estimation of $\boldsymbol{\theta}$. This is the purpose of the Bayesian estimation strategy explored in this paper.

3. BAYESIAN ESTIMATION

This paper studies a new Bayesian strategy for estimating the unknown parameter vector $\boldsymbol{\theta}$ from altimeter waveform samples. Bayesian estimators are based on the posterior distribution of $\boldsymbol{\theta}$

denoted as $f(\boldsymbol{\theta}|\boldsymbol{y})$ where $\boldsymbol{y} = (y_1, ..., y_K)$ denotes the altimetric waveform vector. This posterior distribution is related to the likelihood of the observations $f(\boldsymbol{y}|\boldsymbol{\theta})$ and the parameter prior density $f(\boldsymbol{\theta})$ via Bayes' theorem:

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{y})} \propto f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}), \tag{3}$$

where $\propto$ stands for "proportional to". As a consequence, Bayes estimators use prior information regarding the unknown parameter vector (summarized in $f(\boldsymbol{\theta})$) in addition to the likelihood which is used for most common estimation methods (including the ML method). This paper shows that significant gain in estimation performance can be obtained when using appropriate prior information regarding the unknown parameters. The likelihood function and parameter priors used for the proposed analysis are defined below.

3.1. Likelihood

In the case of independent noise samples n_k, the likelihood of the observation vector $\boldsymbol{y}$ can be expressed as the product of K gamma probability density functions (pdfs):

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \left[\frac{L^L}{\Gamma(L)}\right]^K \exp\left(-L\sum_{k=1}^{K}\frac{y_k}{s_k}\right)\prod_{k=1}^{K}\frac{y_k^{L-1}}{s_k^L}. \tag{4}$$

3.2. Parameter priors

This paper proposes to use three kinds of parameter priors:

- *uniforms priors*: for a given parameter, we have computed the histogram of estimates computed on a satellite cycle and determined an interval containing most estimates. The parameter prior is supposed to be uniform on this interval. For instance, according to Fig. 1, the prior for SWH (in meters) is uniform on $[0, 11]$, the prior for τ is uniform on $[30, 32.5]$ (τT_s is expressed in seconds where T_s is the sampling rate) and the prior for P_u (in FFT power units) is uniform on $[9.5, 25]$ (note that parameter P_u is related to σ_0 via an appropriate scaling factor).
- *priors constructed from histograms associated to satellite cycles*: these histograms (also presented in [5]) have been interpolated by using linear combination of splines [6]. These approximations provide smooth PDFs for the unknown parameters as illustrated in Fig. 1.
- *dynamic priors*: priors constructed at a given time instant t from parameter estimates resulting from the altimeter waveforms at time instant $t-1$. In this paper, we have used Gaussian priors with appropriate constant variances and means resulting from estimates at time $t-1$.

3.3. Posterior distribution

The posterior distribution of the unknown parameter vector $\boldsymbol{\theta}$ can be computed from Bayes' theorem according to (3). Unfortunately, this posterior distribution is too complex to derive closed-form expressions of standard Bayesian estimators such as the minimum mean square error (MMSE) or maximum a posteriori (MAP) estimators [7]. Instead, this paper proposes to use MCMC methods to generate samples according to the posterior distribution of $\boldsymbol{\theta}$. These simulated samples are then used to approximate the MMSE or MAP estimators of $\boldsymbol{\theta}$. The sampler used to generate samples distributed according to the posterior (3) is detailed in the next section.

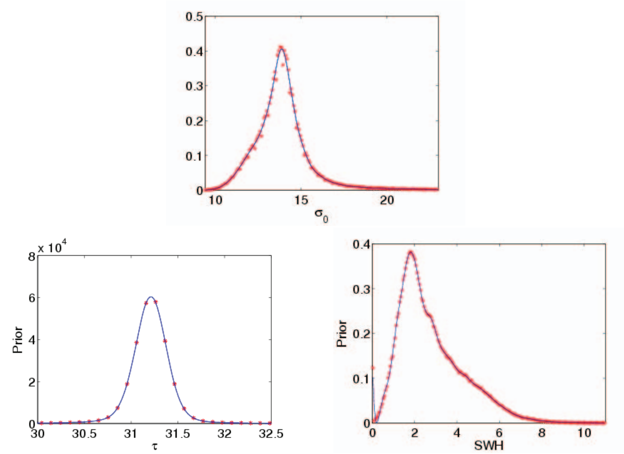


Fig. 1: Histograms of σ_0, τ, SWH ($*$ raw data, – extrapolated data).

4. A METROPOLIS WITHIN GIBBS SAMPLER

4.1. Gibbs sampler

MCMC methods are simulation algorithms which draw samples according to a posterior distribution known up to a multiplicative constant [2]. A famous MCMC method is the Gibbs sampler which generates iteratively samples distributed according to the full conditional distributions associated to the posterior distribution of interest. To generate samples according to $f(\boldsymbol{\theta}|\boldsymbol{y})$, the Gibbs sampler uses iteratively the sampling procedures reported in ALGO. 1. After a so-called *burn-in* period (i.e. when t is sufficiently large), the samples generated according to the algorithm below are known to be distributed according to the target distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$. However, some of the below generations are not easy to handle because the conditional distributions do not belong to known families of distributions. In such cases, hybrid Metropolis-within-Gibbs moves can be used instead. These moves consist of generating samples distributed according to an appropriate proposal distribution and accept or reject these samples with a given probability. This strategy is detailed in the next section.

---

for each t = 1,...,T do
· generate P_u^{t+1} according to $f(P_u|\tau^t, \mathrm{SWH}^t, \xi^t, \boldsymbol{y})$
· generate τ^{t+1} according to $f(\tau|P_u^{t+1}, \mathrm{SWH}^t, \xi^t, \boldsymbol{y})$
· generate SWH^{t+1} according to $f(\mathrm{SWH}|\tau^{t+1}, P_u^{t+1}, \xi^t, \boldsymbol{y})$
· generate ξ^{t+1} according to $f(\xi|\mathrm{SWH}^{t+1}, \tau^{t+1}, P_u^{t+1}, \boldsymbol{y})$

ALGO. 1: Gibbs Sampler Algorithm.

4.2. Metropolis-Hastings

The Metropolis-Hastings (MH) algorithm is an extremely flexible method for generating samples distributed according to a given pdf. Denote as f the target pdf, here one of the conditional pdfs appearing in ALGO. 1 and as q a proposal density used to generate "candidates" (also referred to as *instrumental distribution*). The MH algorithm constructs a Markov chain whose stationary distribution is our distribution of interest f using the proposal q. After initialization, the MH algorithm generates a candidate ϕ according to the proposal

($\phi \sim q$). This candidate is accepted or rejected according to the following rule

$$\boldsymbol{\theta}^{t+1} = \begin{cases} \phi \text{ with probability } \rho = \min\left\{1, \frac{f(\phi|\boldsymbol{y})}{f(\boldsymbol{\theta}^t|\boldsymbol{y})} \frac{q(\boldsymbol{\theta}^t|\phi)}{q(\phi|\boldsymbol{\theta}^t)}\right\} \\ \boldsymbol{\theta}^t \text{ otherwise.} \end{cases}$$

Note that we only need to know f and q up to proportionally constants since both constants cancel in the calculation of ρ. The instrumental distributions used in this paper are the *a priori* distributions. In this case, the MH acceptation probability ρ reduces to the likelihood ratio $\rho = \min\left\{1, f(\boldsymbol{y}|\phi)/f(\boldsymbol{y}|\boldsymbol{\theta}^t)\right\}$.
The hybrid Metropolis-within-Gibbs algorithm includes the accept/reject procedure in ALGO. 1, when the generation according to a conditional distribution is not possible. The Metropolis-within-Gibbs algorithm generating samples distributed according to the posterior (3) is summarized in ALGO. 2. After these simulations have been conducted, the MMSE estimator of $\boldsymbol{\theta}$ is computed as

$$\widehat{\boldsymbol{\theta}}_{\text{MMSE}} = \frac{1}{n_c} \sum_{t=n_b+1}^{T=n_b+n_c} \boldsymbol{\theta}^t, \tag{5}$$

where n_b is the number of *burn-in* iterations, n_c is the number of iterations used for computing the estimator and $\boldsymbol{\theta}^t$ is the parameter vector generated at iteration t.

---

for each t = 1,...,T do

· generate ϕ_{P_u} according to $f(P_u)$

$$P_u^{t+1} = \begin{cases} \phi_{P_u} \text{ with probability} \\ \qquad \rho = \min\left\{1, \frac{f(\boldsymbol{y}|\phi_{P_u},\tau^t,\text{SWH}^t,\xi^t)}{f(\boldsymbol{y}|P_u^t,\tau^t,\text{SWH}^t,\xi^t)}\right\} \\ P_u^t \text{ otherwise} \end{cases}$$

· generate ϕ_τ according to $f(\tau)$

$$\tau^{t+1} = \begin{cases} \phi_\tau \text{ with probability} \\ \qquad \rho = \min\left\{1, \frac{f(\boldsymbol{y}|P_u^{t+1},\phi_\tau,\text{SWH}^t,\xi^t)}{f(\boldsymbol{y}|P_u^{t+1},\tau^t,\text{SWH}^t,\xi^t)}\right\} \\ \tau^t \text{ otherwise} \end{cases}$$

· generate ϕ_{SWH} according to $f(\text{SWH})$

$$\text{SWH}^{t+1} = \begin{cases} \phi_{\text{SWH}} \text{ with probability} \\ \qquad \rho = \min\left\{1, \frac{f(\boldsymbol{y}|P_u^{t+1},\tau^{t+1},\phi_{\text{SWH}},\xi^t)}{f(\boldsymbol{y}|P_u^{t+1},\tau^{t+1},\text{SWH}^t,\xi^t)}\right\} \\ \text{SWH}^t \text{ otherwise} \end{cases}$$

· generate ϕ_ξ according to $f(\xi)$

$$\xi^{t+1} = \begin{cases} \phi_\xi \text{ with probability} \\ \qquad \rho = \min\left\{1, \frac{f(\boldsymbol{y}|P_u^{t+1},\tau^{t+1},\text{SWH}^{t+1},\phi_\xi)}{f(\boldsymbol{y}|P_u^{t+1},\tau^{t+1},\text{SWH}^{t+1},\xi^t)}\right\} \\ \xi^t \text{ otherwise} \end{cases}$$

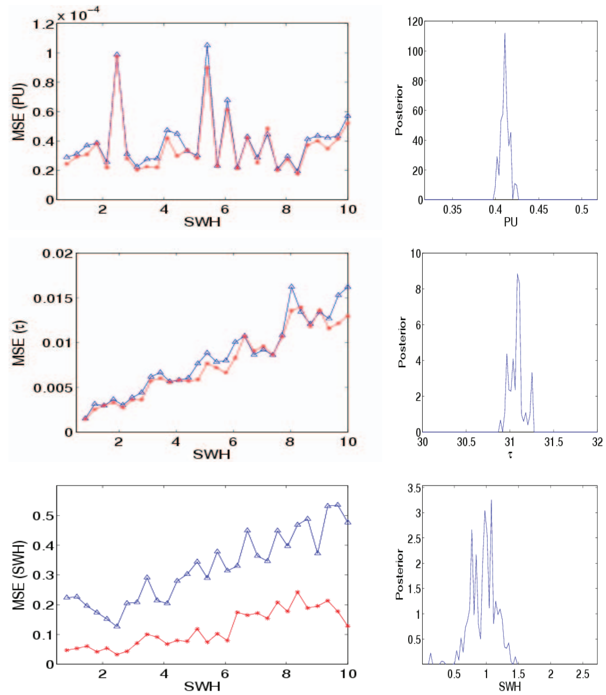
ALGO. 2: Metropolis within Gibbs Algorithm.

5. SIMULATIONS

This section shows some simulation results obtained with the proposed Bayesian estimation method. Synthetic signals have been generated with the following parameters: $L = 100$ (number of looks), $K = 104$ (number of samples) and $\xi = 0$ (off-nadir angle). The other parameters σ_p^2 and γ have been computed according to the Envisat configuration [8].

5.1. Uniform priors

The first set of simulations has been obtained using uniform priors for P_u, τ and SWH. Figures 2(b) show examples of posteriors for the parameters P_u, τ and SWH estimated using the samples of the proposed MCMC method. These posteriors have been obtained for $P_u = 0.41$, $\tau = 31.2$ and SWH $= 1$. The posterior pdfs of P_u, τ and SWH are clearly centered around values which are close to the actual ones. Figures 2(a) compare the MSEs of the MLE and Bayesian estimator. These figures have been obtained with 30 different equidistantly spaced values of SWH (P_u and τ have been drawn randomly according to their prior distributions). The two algorithms perform similarly for the estimation of the amplitude P_u and the epoch τ (MSEs expressed in m^2). However, the estimation of SWH is clearly improved when using the Bayesian algorithm (there is an improvement of 30 centimeters to 60 centimeters in the square root MSE of SWH when using the proposed Bayesian method).



(a) MSEs (—✱— Bayes, —△— MLE). (b) Posterior PDFs.

Fig. 2: MSEs and Posterior distribution of P_u, τ and SWH.

5.2. Priors constructed from satellite cycles

This section investigates the use of prior distributions constructed from the histograms depicted in figures 1. These histograms (red stars) have been extrapolated with B-splines providing continuous priors (blue lines) for the different parameters (P_u, τ and SWH). The parameter posteriors are then obtained as the product between the likelihood and the priors. Figures 3 show examples of data likelihoods (red stars), posteriors (blue circles) and *a priori* distributions (green triangles) for the epoch parameter τ. The first figure (left) shows that the posterior mode is closer to the actual value of $\tau = 31.5$ when compared to the likelihood. Thus, the prior distribution has provided useful information for the estimation of this parameter. The right figure shows another example of posterior obtained for $\tau = 30.5$ (a very unlikely a priori value of τ!). In this case, the prior distribution is not in agreement with the actual value of τ. As a result, the ML estimator has to be preferred to the Bayesian estimator for the estimation of τ. These two examples show that using parameter histograms as priors does not systematically improves the estimation performance.

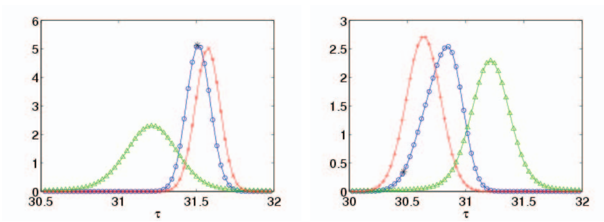


Fig. 3: likelihood, conditional posterior, prior and $*$ τ.

5.3. Dynamic Gaussian priors

This section studies dynamic Gaussian priors with time-varying means and appropriate variances. More precisely, the mean of each parameter prior at a given time instant t has been chosen as the estimated value of this parameter at time instant $t-1$. The variance of this prior has been adjusted to a large value computed using estimates from a satellite cycle. In order to illustrate the performance of this kind of prior, we have generated synthetic signals with time-varying parameters P_u, SWH and τ corresponding to typical altimeter waveforms. Figures 4 compare the MSEs of the MLE and the Bayesian estimator (computed from 100 Monte Carlo runs). Both estimators provide similar performance for the estimation of P_u. The estimation of τ is slightly improved when using the Bayesian method (a gain up to six centimeters has been observed for this example). Finally, the Bayesian estimator of SWH clearly outperforms the MLE as in the case of uniforms priors (a gain of 29 to 58 centimeters is observed for this example).

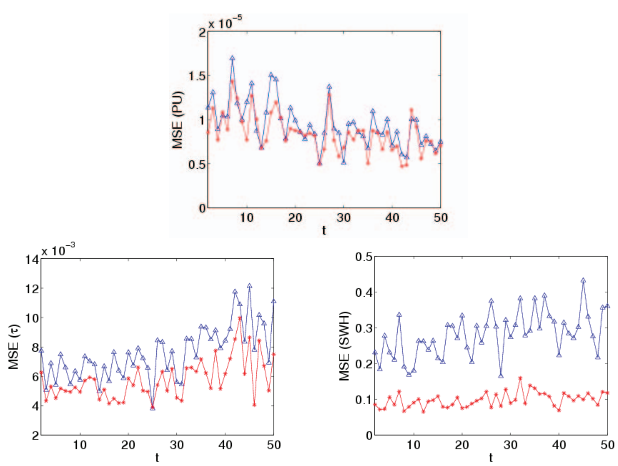


Fig. 4: MSEs (Bayes, MLE).

5.4. Sampler convergence

The Metropolis-within-Gibbs sampler allows one to draw samples asymptotically distributed according to the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$. Controlling the sampler convergence is a problem which has received considerable interest in the literature (see for instance [9] and references therein). For instance, the number of *burn-in* samples can be determined thanks to the popular potential scale reduction factor (PSRF) [10] (the reader is invited to consult [10] and [11] for definitions). For a given value of n_b and n_c in (5), a value of the PSRF below 1.2 indicates a good convergence of the sampler. The values of the PSRF denoted as $\sqrt{\widehat{\rho}}$ for the different parameters P_u, τ and SWH are reported in Table 1 for $n_b = 3000$ and $n_c = 5000$. These values have been obtained by running 100 parallel chains and by computing the associated between-class and within-class variances. They clearly indicate that the sampler has converged for $n_b = 3000$ and $n_c = 5000$.

P_u	$\sqrt{\widehat{\rho}}$	τ	$\sqrt{\widehat{\rho}}$	SWH	$\sqrt{\widehat{\rho}}$
0.4	1.0048	31	1.0224	3	1.0155

Table 1: Potential Scale Reduction Factors.

6. CONCLUSIONS

This paper studied Bayesian sampling algorithms for estimating the parameters of altimeter waveforms. The proposed estimation methodology showed that defining appropriate prior information regarding the unknown parameters can be beneficial for their estimation. It is interesting to note that the posterior distributions derived in this paper were used for parameter estimation. However, they can also provide uncertainties regarding these estimations. The Bayesian estimator presented in this paper was implemented using an hybrid Metropolis-within-Gibbs sampler whose convergence was monitored by the potential scale reduction factor. Perspectives include the optimization of the *burn-in* period which might reduce the computational cost of the proposed algorithm.

7. REFERENCES

[1] G. S. Brown, "The average impulse response of a rough surface and its applications," *IEEE Trans. Antennas and Propagation*, vol. 25, pp. 67–74, Jan. 1977.

[2] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, New York, first edition, 1999.

[3] J.-P. Dumont, *Estimation optimale des paramètres altimétriques des signaux radar Poséidon*, Ph.D. thesis, Institut National Polytechnique (INP), Toulouse, France, 1985.

[4] C. Mailhes, J.-Y. Tourneret, J. Severini, and P. Thibaut, "Cramer-rao bounds for radar altimeter waveforms," in *Proc. EUSIPCO*, Lausanne, Switzerland, Aug. 2008, vol. 4, to appear.

[5] 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*, pp. 191–193, 2004.

[6] C. De Boor, *A Practical Guide to Splines*, Springer, New York, 1978.

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

[8] A. Ollivier, *Nouvelle approche pour l'extraction de paramètres géophysiques à partir des mesures en altimétrie radar*, Ph.D. thesis, Institut National Polytechnique (INP), Grenoble, FRANCE, 2006.

[9] C. P. Robert and S. Richardson, "Markov Chain Monte Carlo methods," in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed., pp. 1–25. Springer Verlag, New York, 1998.

[10] A. Gelman and D. Rubin, "Inference from iterative simulation using multiple sequences," *Statist. Sci.*, vol. 7, no. 4, pp. 457–551, Nov. 1992.

[11] S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 434–455, Dec. 1998.