MULTI-PARAMETER ESTIMATION FOR COGNITIVE RADAR IN COMPOUND GAUSSIAN CLUTTER

Anish Turlapaty and Yuanwei Jin

Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, MD 21853

ABSTRACT

In this paper, we consider the problem of multi-parameter estimation in the presence of compound Gaussian clutter for cognitive radar using the variational Bayesian approach. The main advantage of variational Bayesian is that the estimation of multi-variate parameters is decomposed to multiple estimation of univariate parameters, thus enabling analytically tractable approximations. Numerical tests demonstrate that the proposed approach leads to improved estimation accuracy than the expectation maximization (EM) method, particularly in the case of non-Gaussian nonlinear models and a small sample size.

Index Terms— Variational Bayesian, compound Gaussian clutter, cognitive radar

1. INTRODUCTION

Cognitive radar has been proposed as a fully adaptive radar transmission and reception system in [1]. In cognitive radar, both the transmitter and the receiver parameters are estimated and updated by learning from the unknown environment, forming a belief on what is learned, and propagating this belief by Bayesian inference. From the parameter estimation perspective, the Bayesian approach enables inclusion of prior information (knowledge) of radar target and clutter by estimating the posterior density of the unknown parameters. The estimation is optimal in the sense of minimizing the Bayesian mean squared error (MSE). Typically, the full joint probability density function (pdf) of all the parameters of interest including the nuisance parameters is considered. However, in the case of high-dimensional multi-variate integration of Bayesian posterior density, the calculation of the posterior pdf and its marginal can be computationally prohibitive and tractable analytical solutions are often not available. Furthermore, the estimation accuracy is directly related to the number of data samples in the Bayesian estimator. However, in many radar applications, the number of available data samples is limited. These computational challenges and limitations must be addressed to develop cognitive radar.

In [2], we proposed a variational Bayesian (VB) based method for parameter estimation and waveform design where a single parameter estimation problem is considered. Variational Bayesian aims to minimize free energy (FE) [3, 2], which is equivalent to minimizing the Kullback-Liebler divergence between the true density and an approximation density. As a result of this functional optimization for density estimation, the marginal VB posterior density has an explicit functional structure, thus leading to closed form solutions [4]. This paper extends our prior work in [2] to multiple parameter estimation in the context of cognitive radar. In adaptive radar detection, estimating the clutter covariance matrix is a very important task since the detection performance depends directly on the accuracy of the estimate. For example, in high-resolution and low-grazing-angle radar, only a small sea surface area is illuminated by a narrow radar beam. The sea clutter due to reflection from the small patch of sea surface is random and non-stationary [5], which is commonly modeled as a compound-Gaussian distribution to characterize its heavy-tailed clutter distributions [6, 7]. Hence, the Bayesian estimator must consider a multi-parameter estimation problem by which the parameters in the compound-Gaussian model, the radar target response, as well as other nuisance parameters are estimated. In this paper, we compare the performance of the proposed VB method with the expectation-maximization (EM) algorithm. In the EM method, expectations of sufficient statistics are computed with respect to the posterior density of hidden variables and then used to iteratively estimate the unknown parameters by the maximum likelihood principle [8, 9]. We show in this paper that the variational algorithms outperform the EM method particularly when estimating parameters that follow non-Gaussian nonlinear models in Bayesian inference. Hence, the proposed variational algorithms provide appealing computational advantages for cognitive radar.

2. PROBLEM FORMULATION

2.1. Radar Signal Model

The compound clutter model is a product of two random processes [9, 6, 7],

$$\psi_t = \sqrt{u_t} w_t \tag{1}$$

where the speckle w_t characterizing the local scattering and is modeled as a zero mean complex Gaussian (ZMCG) process $w_t \sim \mathcal{CN}(0, \sigma^2)$. The component u_t is a slow changing

This work was supported in part by the U.S. Army Research Laboratory, the Office of Naval Research and the Army Research Office under grant no. W911NF-11-1-0160 and the National Science Foundation under grant no. CMMI-1126008.

process termed texture that follows an inverse Gamma distribution $u_t \sim \Gamma^{-1}(\alpha)$, where the pdf of u_t is

$$p(u_t;\alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)} u_t^{-\alpha-1} \exp\left(-\frac{\alpha}{u_t}\right) \quad (2)$$

The model (1) is referred to as K-clutter [10, 5] and the parameter α is known as the Nakagami parameter [11]. Next, we assume the radar transmits a waveform Φ_t and the electromagnetic (EM) energy hits a target with a complex amplitude response x. The reflected EM energy is intercepted by the radar receiver. The radar signal model that includes the clutter from (1) is given by

$$y_t = \Phi_t x + \psi_t, \ t = 1, 2, \cdots, N \quad (3)$$

The conditional probabilistic model of the measurements is a complex Gaussian distribution given by

$$y_t | u_t, x, \sigma^2 \sim \mathcal{CN}(\Phi_t x, u_t \sigma^2) \quad (4)$$

2.2. Problem of Parameter Estimation

The complete hierarchical stochastic model, i.e., the joint probability density function of the measurements, hidden variables, and the unknown parameters at time t is given by

$$\begin{aligned} p(y_t, u_t, x, \sigma^2, \alpha; \Phi_t) &= p(y_t|u_t, x, \sigma^2; \Phi_t) \quad (5) \\ & \quad p_0(x, \sigma^2) p(u_t|\alpha) p_0(\alpha) \end{aligned}$$

where $p(y_t|u_t, x, \sigma^2; \Phi_t)$ is the conditional density given in (4). The probabilistic model of texture $p(u_t|a)$ is given by (2). $p_0(x, \sigma^2)$ and $p_0(\alpha)$ are prior densities of (x, σ^2) and α, respectively. Initially, the unknown parameter vector is given by $[x, \sigma^2, \alpha]$. When we use a variational estimation method, the covariance estimate depends on the current estimate of α. However, initially this mutual dependence of estimates introduces a multiplicative error in the estimates of these two parameters. To correct this error, we use the technique of covariance adjustment and introduce an additional parameter λ in the texture model (2) by redefining the covariance in terms of an adjusted covariance σ_a^2, [9], i.e., the relation between the actual covariance and the adjusted covariance is $\sigma^2 = \sigma_a^2/\lambda$. Hence, the adjusted texture model becomes

$$p(u_t|\alpha, \lambda) = (\alpha\lambda)^\alpha / \Gamma(\alpha)\, u_t^{-\alpha-1} e^{-\alpha/(\lambda u_t)} \quad (6)$$

The new augmented parameter vector to be estimated is

$$\boldsymbol{\theta} = [x, \sigma^2, \alpha, \lambda] \quad (7)$$

3. VARIATIONAL BAYESIAN ESTIMATOR

3.1. Background of variational Bayesian inference

Consider a hierarchical probabilistic model

$$p(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) = p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) p(\mathbf{X}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (8)$$

where $\mathbf{Y}$ are the measurements, $\mathbf{X}$ are hidden variables, and $\boldsymbol{\theta}$ are unknown parameters. In the exact Bayesian approach, the unknown parameters are determined by evaluating the joint posterior density $p(\mathbf{X}, \boldsymbol{\theta}|\mathbf{Y})$ using the Bayes rule

$$p(\mathbf{X}, \boldsymbol{\theta}|\mathbf{Y}) = p(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) / p(\mathbf{Y}) \quad (9)$$

while the marginal posterior $p(\boldsymbol{\theta}|\mathbf{Y})$ is evaluated by integrating $\mathbf{X}$ out from the joint posterior. However, in practice the denominator in (9) is usually theoretically intractable except in some special cases. Moreover, in the case of multiple parameters in $\boldsymbol{\theta}$ even the numerical integration is computationally expensive and time consuming [4]. To address this issue, approximation methods are needed to determine alternative density functions. Variational Bayesian estimation aims to find approximations $q(\mathbf{X})$ and $q(\boldsymbol{\theta})$ to the marginal posterior densities of parameters that minimize the variational free energy of the approximate density and the joint pdf in (8) [12]. So the key idea is factorization of $q(\mathbf{X}, \boldsymbol{\theta})$ into $q(\mathbf{X})q(\boldsymbol{\theta})$, thus separating the densities of $\mathbf{X}$ and $\boldsymbol{\theta}$. The variational free energy (FE) for $\mathbf{Y}, \mathbf{X}$ and $\boldsymbol{\theta}$ is given by [2, 4, 3]

$$F(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) = -\int q(\mathbf{X}) q(\boldsymbol{\theta}) \log \frac{p(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta})}{q(\mathbf{X}) q(\boldsymbol{\theta})} d\mathbf{X} d\boldsymbol{\theta} \quad (10)$$

The approximate density $q(X)$ of hidden variables is determined by variational Bayesian expectation (VBE)

$$q(\mathbf{X}) = \arg\min_{q(\mathbf{X})} F(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) \quad (11)$$

where the optimal solution to (11) is given by

$$\log q(\mathbf{X}) = \langle \log p(\mathbf{X}|\boldsymbol{\theta}) \rangle_{q(\boldsymbol{\theta})} + \langle \log p(Y|X, \boldsymbol{\theta}) \rangle_{q(\boldsymbol{\theta})} + \text{const.} \quad (12)$$

where $\langle \cdot \rangle$ is an inner product. The approximate density $q(\boldsymbol{\theta})$ of the parameters is determined by variational Bayesian minimization (VBM)

$$q(\boldsymbol{\theta}) = \arg\min_{q(\boldsymbol{\theta})} F(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) \quad (13)$$

where the optimal solution to (13) is given by [13, 14]

$$\log q(\boldsymbol{\theta}) = \log p_0(\boldsymbol{\theta}) + \langle \log p(Y|X, \boldsymbol{\theta}) \rangle_{q(\mathbf{X})} + \text{const.} \quad (14)$$

3.2. Variational Bayesian estimator

The new joint probability density function depends on the measurement vector $\mathbf{Y} = [y_1, \cdots, y_N]^T$, the vector of independent hidden variables $\mathbf{U} = [u_1, \cdots, u_N]^T$, and the adjusted parameter vector $\boldsymbol{\theta}$, which takes the form of

$$\begin{aligned} p(y_t, u_t, x, \sigma^2, \alpha, \lambda; \Phi_t) &= p(y_t|u_t, x, \sigma^2; \Phi_t) \quad (15) \\ & \quad p_0(x, \sigma^2) p(u_t|\alpha, \lambda) p_0(\alpha, \lambda) \end{aligned}$$

The full data log-likelihood function is given by

$$\begin{aligned} \Lambda(\mathbf{Y}, \mathbf{U}, \boldsymbol{\theta}) = & \sum_t \log p(y_t|u_t, x, \sigma^2; \Phi_t) + \sum_t \log p(u_t|\alpha, \lambda) \\ & + \log p_0(x, \sigma^2) + \log p_0(\alpha, \lambda) \quad (16) \end{aligned}$$

Eqn. (16) involves six parameters, which is very complicated. Next, we present a variation Bayesian approach to approximate it in the sense of minimization of the free-energy (i.e.,

KL divergence, [2]), so that analytically trackable closed form expressions of pdfs can be obtained. The evaluation of (16) involves the expectation and minimization steps (i.e., E-step and M-step) while using the VB approach. Due to space limitations, we will omit some details of the mathematical derivation.

3.2.1. VB estimation step (VBE)

The first step in variational estimation is to determine the approximate density of the hidden variable u_t given the measurements y_t by using the minimum free energy principle. Since the hidden variables u_t are independent across time, the approximate density of the vector $\mathbf{U}$ can be written as $q(\mathbf{U}) = \prod_{t=1}^{N} q(u_t)$. Its approximate density will be obtained by minimizing the free-energy defined in (10), which can be re-written as

$$F(\mathbf{Y},\mathbf{U},\boldsymbol{\theta})=-\Big\langle \Lambda(\mathbf{Y},\mathbf{U},\boldsymbol{\theta})\Big\rangle_{q(\mathbf{U})}+\Big\langle \sum_{t=1}^{N}\log(q(u_t))\Big\rangle_{q(\mathbf{U})}$$

based upon (16). Hence, the optimization problem is given by

$$q(u_t) = \arg\min_{q(u_t)} F(\mathbf{Y},\mathbf{U},\boldsymbol{\theta}) \tag{17}$$

Following the optimal solution in (12), we obtain

$$\begin{aligned}\log q(u_t) \;\; &= \;\; \langle \log p(u_t|\alpha,\lambda)\rangle_{q(\alpha,\lambda)} + \\ &\quad \langle \log p(y_t|u_t,x,\sigma^2)\rangle_{q(x,\sigma^2)} + \text{const.}\end{aligned} \tag{18}$$

Next, inserting the pdf (4) in (18), we have

$$\begin{aligned}q(u_t) \;\propto\; & u_t^{-\left(\langle\alpha\rangle_{q(\alpha,\lambda)}+1\right)-1} \\ & e^{-\left[\langle\alpha/\lambda\rangle_{q(\alpha,\lambda)}+\langle|y_t-\Phi_t x|^2\frac{1}{\sigma^2}\rangle_{q(x,\sigma^2)}\right]/u_t}\end{aligned} \tag{19}$$

which is consistent with the definition of an inverse Gamma density function $\Gamma^{-1}(\cdot)$, leading to a closed form $q(u_t) = \Gamma^{-1}\Big(u_t; c_U, d_U\Big)$ with its parameters being $c_U = \langle\alpha\rangle_{q(\alpha,\lambda)} + 1$ and $d_U = \langle\alpha/\lambda\rangle_{q(\alpha,\lambda)} + \langle|y_t-\Phi_t x|^2/\sigma^2\rangle_{q(x,\sigma^2)}$.

3.2.2. VB minimization step (VBM)

The optimization problem for estimating the pdf of the parameter vector $\boldsymbol{\theta}$ is given by

$$q(x,\sigma^2,\alpha,\lambda) = \arg\min_{q(x,\sigma^2,\alpha,\lambda)} F(\mathbf{Y},\mathbf{U},\boldsymbol{\theta}) \tag{20}$$

We write the variational posterior density of the parameter vector in (14) as follows

$$\begin{aligned}\log q(x,\sigma_a^2,\alpha,\lambda) \;=\; & \log p_0(\alpha,\lambda) + \log p_0(x,\sigma_a^2) \\ & + \sum_{t=1}^{N}\Big\langle \log p(y_t|u_t,x,\sigma_a^2;\Phi_t)\Big\rangle_{q(u_t)} + \\ & \sum_{t=1}^{N}\Big\langle \log p(u_t|\alpha,\lambda)\Big\rangle_{q(u_t)} + \text{const.}\end{aligned} \tag{21}$$

Note that the terms for the joint pdf of (x,σ_a^2) are separable from the pdf of (α,λ), i.e.,

$$q(x,\sigma_a^2,\alpha,\lambda) = q(x,\sigma_a^2)\; q(\alpha,\lambda) \tag{22}$$

Next, we evaluate $\log q(x,\sigma_a^2)$ and $\log q(\alpha,\lambda)$ from (21) and (22) to derive approximate closed forms. We start by writing the log-function of pdf of (x,σ_a^2) as

$$\begin{aligned}\log q(x,\sigma_a^2) = \log p_0(x,\sigma_a^2) + & \sum_{t=1}^{N}\Big\langle \log 1/(\pi u_t\sigma_a^2)\Big\rangle_{q(u_t)} \\ & - \sum_{t=1}^{N}\Big\langle |y_t-\Phi_t x|^2/(u_t\sigma_a^2)\Big\rangle_{q(u_t)} + \text{const.}\end{aligned} \tag{23}$$

which leads to the joint pdf $q(x,\sigma_a^2)$, which is a complex Gaussian inverse Gamma ($\mathcal{CGIG}$) distribution

$$\begin{aligned}q(x,\sigma_a^2) \;\; \propto \;\; & p_0(x,\sigma_a^2)(1/\sigma_a^2)^N \\ & e^{-\sum_{t=1}^{N}|y_t-\Phi_t x|^2/\sigma_a^2\langle 1/u_t\rangle_{q(u_t)}}\end{aligned} \tag{24}$$

Note that the functional form is nicely preserved if the prior pdf is assumed to be $p_0(x,\sigma_a^2) \sim \mathcal{CGIG}(\rho,\eta,\beta,\mu_x)$, which leads to an explicit single $\mathcal{CGIG}$ distribution expression for $q(x,\sigma_a^2)$.

Next, we examine the logarithm of the joint pdf of (α,λ), which is given by (where C is a constant)

$$\log q(\alpha,\lambda)=\log p_0(\alpha,\lambda)+\sum_{t=1}^{N}\langle\log p(u_t|\alpha,\lambda)\rangle_{q(u_t)}+ C \tag{25}$$

By (6) and $T_l \triangleq \sum_{t=1}^{N}\langle\log u_t\rangle_{q(u_t)}$, $T_i \triangleq \sum_{t=1}^{N}\langle 1/u_t\rangle_{q(u_t)}$ we obtain the joint pdf $q(\alpha,\lambda)$ as follows

$$q(\alpha,\lambda) \propto p_0(\alpha,\lambda)\frac{(\frac{\alpha}{\lambda})^{N\alpha}}{(\Gamma(\alpha))^N}e^{-\alpha(T_l+T_i/\lambda)} \tag{26}$$

Since this is not a known probability distribution, it can not provide closed form solutions. We can rely on numerical means for calculation, or alternatively, we propose to induce another layer of factorization on the joint pdf $q(\alpha,\lambda)$ as

$$q(\alpha,\lambda) = q(\alpha)q(\lambda) \tag{27}$$

via the optimization problem to obtain closed form expressions of $q(\alpha)$ and $q(\lambda)$, i.e.,

$$\arg\min_{q(\alpha),q(\lambda)} F(\mathbf{Y},\mathbf{U},x,\sigma^2,\alpha,\lambda) \tag{28}$$

where the new free-energy quantity is defined by

$$\begin{aligned}& F(\mathbf{Y},\mathbf{U},x,\sigma^2,\alpha,\lambda) = \\ & -\Big\langle \Lambda(\mathbf{Y},\mathbf{U},x,\sigma_a^2,\alpha,\lambda)\Big\rangle_{q(\alpha)q(\lambda)q(\mathbf{x},\sigma^2)q(\mathbf{U})} \\ & +\Big\langle q(\alpha)q(\lambda)q(\mathbf{x},\sigma^2)q(\mathbf{U})\Big\rangle_{q(\alpha)q(\lambda)q(\mathbf{x},\sigma^2)q(\mathbf{U})}\end{aligned} \tag{29}$$

As a result, we obtain the approximate inverse Gamma distribution for $q(\lambda) = \Gamma^{-1}(\lambda; c_\lambda, d_\lambda)$ and the approximate Gamma distribution for $q(\alpha) = \Gamma(c_\alpha(0), d_\alpha(0))$ by Lindley's approximation [15]. Due to space limitations, we will omit the detailed expressions of $q(\alpha)$ and $q(\lambda)$ in this paper.

4. NUMERICAL SIMULATIONS

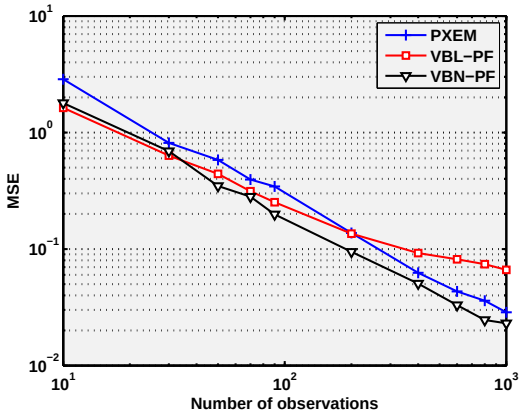
We present estimation performance of the three estimators by numerical simulations. The three estimators are: **(1) VBN-PF:** Variational Bayesian approach with induced factorization on $q(\alpha, \lambda)$ and a numerical integration for evaluation of the moments of α. **(2) VBL-PF:** Variational Bayesian approach with induced factorization on $q(\alpha, \lambda)$ and the Lindley's approximation for evaluation of moments of α [15]. **(3) PXEM:** parameter expanded expectation maximization approach, where $q(u_t)$ is updated based on Bayesian approach, while each of the unknown parameters is determined by using the maximum likelihood principle [9]. The key differences between the EM and VB algorithms are twofold. First, in the EM method, unknown parameters are considered as deterministic values and estimated using the maximum likelihood (ML) method whereas in the VB method the unknown parameters are modeled as random variables and the Bayesian approach is used to determine their approximate posterior densities. Second, in the EM method, each M-step only receives the updated sufficient statistics from the E-step, where as in VB, as a result of the Bayesian principle, the VBM step utilizes the updated priors of the randomized parameters.

The parameters of the clutter model and target model are chosen as follows: Nakagami parameter $\alpha = 3$, clutter power $\sigma_a^2 = 5$ dB, the complex target response $x = 1.6 + 1.0i$, the spectral density of the waveform is normalized $|\Phi_t| = 1$, the number of observations N varies from 10 to 1000.

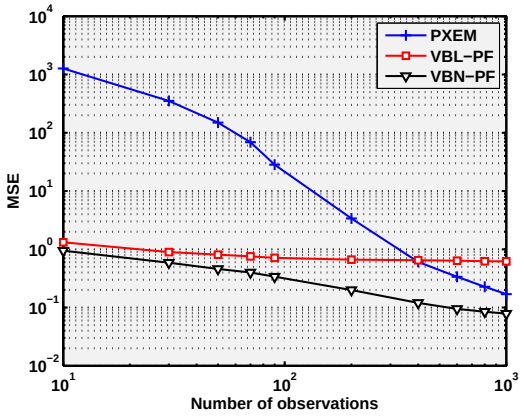
Figs. 1(a) shows that, for the parameters σ^2, the variational algorithm VBL-PF outperforms the PXEM algorithm when the number of observations $N < 200$ while the VBN-PF method outperforms the PXEM for all values of N. For the estimation of Nakagami parameter α, the performance of VB methods is much better than PXEM in terms of MSE, as shown in Fig. 1(b). This is because in the PXEM method, the ML solution is obtained by solving a nonlinear equation due to the non-Gaussian clutter model. For the estimation of radar target response x, the three methods have very similar performance for all values of N as depicted in Fig. 1(c). The explanation for the similarity is that the unknown parameter is a linear function of the observations and follows a Gaussian distribution. Note that ML and Bayesian algorithms usually have similar performance for Gaussian linear models. The advantage of VB approach can be observed for non-Gaussian nonlinear models such as the ones that involve σ^2 and α. This observation is consistent with existing literature on variational Bayesian studies [4, 14]. Finally, The estimation performance of the adjustment parameter λ is not discussed as it is only a nuisance parameter for correcting the multiplicative error.

5. CONCLUSION

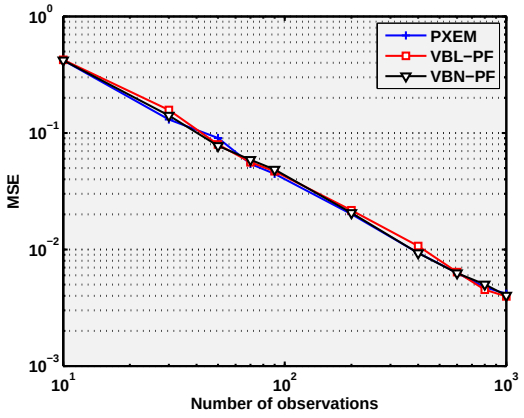
We develop variational Bayesian algorithms for estimating multi-parameters of a compound Gaussian clutter model and target response. The VB method yields closed form expressions for the posterior probability density functions and results in improved estimation performance for the clutter model, especially for parameters of non-Gaussian nonlinear models and when the number of measurements is small.



(a)



(b)



(c)

Fig. 1. Comparison of mean squared error (MSE) between PXEM, VBN-PF and VBL-PF. (a) clutter noise variance σ^2, (b) Nakagami parameter α, and (c) target response x.

6. REFERENCES

[1] S. Haykin, "Cognitive radar: a way of the future," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 30–40, January 2006.

[2] A. Turlapaty and Y. Jin, "Parameter estimation and waveform design for cognitive radar by minimal free-energy principle," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6244–6248.

[3] K. Friston, "The free-energy principle: a unified brain theory," *Nature Reviews Neuroscience*, vol. 11, pp. 127–138, Feburary 2010.

[4] V. Smidl and A. Quinn, "Variational Bayesian filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5020–5030, October 2008.

[5] K. J. Sangston and K. R. Gerlach, "Coherent detection of radar targets in a non-Gaussian background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 2, pp. 330–340, April 1994.

[6] J. W. Wright, "A new model for sea clutter," *IEEE Transactions on Antennas and Propagation*, vol. AP-16, pp. 217–223, 1968.

[7] M. Greco, F. Bordoni, and F. Gini, "X-band sea-clutter nonstationarity: Influence of long waves," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 2, pp. 269–283, April 2004.

[8] J. V. Candy, *Bayesian Signal Processing*, Wiley-IEEE Press, Hoboken, NJ, 2009.

[9] J. Wang, A. Dogandzic, and A. Nehorai, "Maximum likelihood estimation of compound-Gaussian clutter and target parameters," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3884–3898, October 2006.

[10] F. Gini, M. V. Greco, M. Diani, and L. Verrazzani, "Performance analysis of two adaptive radar detectors against non-Gaussian real sea clutter data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 4, pp. 1429–1439, October 2000.

[11] S. Haykin, R. Bakker, and B. W. Currie, "Uncovering nonlinear dynamics - The case study of sea clutter," *Proceedings of the IEEE*, vol. 90, no. 5, pp. 860–881, May 2002.

[12] M. J. Beal and Z. Ghahramani, "The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures," in *Bayesian Statistics 7*, J. M. Bernardo and M. J. Bayarri, Eds., pp. 453 – 463. University of Oxford Press, 2003.

[13] J. Paisley, D. Blei, and M. Jordan, "Variational Bayesian inference with stochastic search," in *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, Edinburgh, Scotland, June 2012, pp. 1–8.

[14] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, November 2008.

[15] D.V. Lindley, "Approximate Bayesian methods," *Trabajos de Estadistica Y de Investigacion Operativa*, vol. 31, no. 1, pp. 223–245, 1980.