KNOWLEDGE-AIDED BAYESIAN COVARIANCE MATRIX ESTIMATION IN COMPOUND-GAUSSIAN CLUTTER

Francesco Bandiera[*]*, Olivier Besson*[†] *and Giuseppe Ricci*[*]

[*]Università del Salento, Dipartimento di Ingegneria dell'Innovazione, Lecce, Italy.
[†]Université de Toulouse, ISAE, Départment Electronique Optronique Signal, Toulouse, France.
francesco.bandiera@unisalento.it, olivier.besson@isae.fr, giuseppe.ricci@unisalento.it

ABSTRACT

We address the problem of estimating a covariance matrix $\boldsymbol{R}$ using K samples $\boldsymbol{z}_k$ whose covariance matrices are $\tau_k\boldsymbol{R}$, where τ_k are random variables. This problem naturally arises in radar applications in the case of compound-Gaussian clutter. In contrast to the conventional approach which consists in considering $\boldsymbol{R}$ as a deterministic quantity, a knowledge-aided (KA) approach is advocated here, where $\boldsymbol{R}$ is assumed to be a random matrix with some prior distribution. The posterior distribution of $\boldsymbol{R}$ is derived. Since it does not lead to a closed-form expression for the minimum mean-square error (MMSE) estimate of $\boldsymbol{R}$, both $\boldsymbol{R}$ and τ_k are estimated using a Gibbs-sampling strategy. The maximum a posteriori (MAP) estimator of $\boldsymbol{R}$ is also derived. It is shown that it obeys an implicit equation which can be solved through an iterative procedure, similarly to the case of deterministic τ_ks, except that KA is now introduced in the iterative scheme. The new estimators are shown to improve over conventional estimators, especially in small sample support.

Index Terms— Covariance matrix, estimation, radar.

1. INTRODUCTION AND PROBLEM STATEMENT

An ubiquitous task of most radar systems is to detect the presence of a target, in a given range cell, in the presence of clutter and thermal noise, whose underlying statistics are generally unknown. In the Gaussian case and when the noise covariance matrix is known, the optimal detector depends directly on the latter, and therefore, when it is unknown, most of the adaptive detection schemes proposed in the literature depend either explicitly or implicitly on an estimate of the covariance matrix. Indeed, this is how the adaptive matched filter [1] or the adaptive normalized matched filter [2, 3] proceed, replacing the true covariance matrix by an estimate in the optimal detector. The estimate of the noise covariance matrix is usually constructed from a set of secondary data samples $\boldsymbol{z}_k$, $k = 1, \cdots, K$ -obtained from range cells close to the cell under test- that, ideally, would share the same statistical properties (or at least the same covariance matrix) as the data in the cell under test. Unfortunately, most real environments are heterogeneous, i.e. there exists some mismatch between the noise statistics in the primary data and in the secondary data. While modeling of heterogeneous environments is a highly debatable topic, a widely used and physically motivated model for non-homogeneous clutter is the compound-Gaussian model [4], whose validity has been assessed on real data, see e.g. [5]. It enables one to model local clutter power fluctuations along the range cells and is deemed reasonable, especially for high range resolution radars. The clutter returns are modeled as $\boldsymbol{z}_k = \sqrt{\tau_k}\boldsymbol{g}_k$ where $\boldsymbol{g}_k$ is a Gaussian vector with covariance matrix $\boldsymbol{R}$ and τ_k is a positive random variable, independent of $\boldsymbol{g}_k$. Various distributions for τ_k have been proposed in the literature but, for most of them, they do not result in a closed-form expression for the optimal detector. Accordingly, covariance matrix estimation is generally intractable. In order to get round this difficulty, use of a normalized sample covariance matrix has been advocated in [6]. Also, considering τ_k as an unknown deterministic quantity, it was shown that the maximum likelihood estimate of $\boldsymbol{R}$ obeys an implicit equation, that can be solved through an iterative procedure [7, 8] and the latter is guaranteed to converge [9].

In this paper, we consider a knowledge-aided scheme for estimating the covariance matrix $\boldsymbol{R}$, see e.g. [10] for a good review of the principles and of the rationale of such approaches. Briefly stated, we assume that some rough knowledge of $\boldsymbol{R}$, viz $\bar{\boldsymbol{R}}$, is available: it can be obtained, e.g. from the general clutter covariance matrix model of [11]. However, $\bar{\boldsymbol{R}}$ cannot predict the local behavior of the clutter and hence only samples $\boldsymbol{z}_k|\tau_k \sim \mathcal{CN}(\boldsymbol{0}, \tau_k\boldsymbol{R})$ are available, τ_ks being random variables. Additionally, similarly to the framework introduced in [12, 13], we assume that $\boldsymbol{R}$ is random, with some prior distribution that depends on $\bar{\boldsymbol{R}}$. Within this framework, we consider optimal i.e. minimum mean-square error (MMSE) estimation of both $\boldsymbol{\tau} = \begin{bmatrix}\tau_1 & \tau_2 & \cdots & \tau_K\end{bmatrix}^T$ and $\boldsymbol{R}$, and maximum a posteriori (MAP) estimation of $\boldsymbol{R}$.

2. DATA MODEL

In this section, we provide the assumptions about our model. As stated previously, we assume that the K vectors $\boldsymbol{z}_k \in \mathbb{C}^N$ are, conditionally to τ_k and $\boldsymbol{R}$, independent, zero-mean Gaussian distributed with covariance matrix $\tau_k\boldsymbol{R}$, i.e.

$$f(\boldsymbol{z}_k|\tau_k, \boldsymbol{R}) = \pi^{-N} |\boldsymbol{R}|^{-1} \tau_k^{-N} \exp\left\{-\tau_k^{-1}\boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right\}. \quad (1)$$

Now, since τ_k and $\boldsymbol{R}$ are random variables, which we assume to be independent, one needs to assign prior distributions for them. On one hand, if one wishes to make the least possible assumptions on the random variables, a non-informative, e.g. Jeffreys, prior can be the solution of choice [14]. However, this approach more or less corresponds to maximum likelihood estimation as almost no prior information about the unknown variables is available. When prior information is to be included in the model, a tradeoff must be made between plausibility, relevance of the prior distributions and mathematical tractability [15]. In our case, we propose to choose for $\boldsymbol{\tau}$ and $\boldsymbol{R}$ conjugate priors. More precisely, we assume that $\boldsymbol{R}$ is drawn from a complex inverse Wishart distribution, with mean $\bar{\boldsymbol{R}}$ and ν

The work of O. Besson was supported by the Mission pour la Recherche et l'Innovation Scientifique (MRIS) of the DGA.

($> N$) degrees of freedom, i.e.

$$f(\boldsymbol{R}) = \frac{|(\nu-N)\bar{\boldsymbol{R}}|^{\nu}\,|\boldsymbol{R}|^{-(\nu+N)}}{\tilde{\Gamma}_N(\nu)}\text{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\bar{\boldsymbol{R}}\right\} \quad (2)$$

where $\tilde{\Gamma}_N(\nu) = \pi^{N(N-1)/2}\prod_{n=1}^{N}\Gamma(\nu-n+1)$ and $\text{etr}\{.\}$ stands for the exponential of the trace of the matrix between braces. We denote as $\boldsymbol{R} \sim \mathcal{CW}^{-1}\left((\nu-N)\bar{\boldsymbol{R}},\nu\right)$ this distribution. On average, $\mathcal{E}\{\boldsymbol{R}\} = \bar{\boldsymbol{R}}$ and ν sets the "distance" between $\boldsymbol{R}$ and $\bar{\boldsymbol{R}}$: as ν increases, $\boldsymbol{R}$ is closer to $\bar{\boldsymbol{R}}$. However, for small ν, $\boldsymbol{R}$ can be quite different from $\bar{\boldsymbol{R}}$. In any case, the two matrices are different with probability one. In order to have a conjugate prior for τ_k, we assume that the τ_ks are independent and distributed according to an inverse Gamma distribution with parameters q_k and β_k, i.e.

$$f(\tau_k) = \frac{\beta_k^{q_k}}{\Gamma(q_k)}\tau_k^{-(q_k+1)}\exp\left\{-\beta_k\tau_k^{-1}\right\} \quad (3)$$

which we denote as $\tau_k \sim \mathcal{IG}(q_k,\beta_k)$. In the sequel, we assume that $\mathcal{E}\{\tau_k\} = 1$ which implies that $\beta_k = q_k - 1$. Hence, the distribution in (3) depends only on q_k: the smaller q_k, the larger the variance of τ_k. Equations (1)-(3) form the model upon which we aim at estimating $\boldsymbol{R}$ and possibly $\boldsymbol{\tau}$. In what follows, we let $\boldsymbol{Z} = \begin{bmatrix}\boldsymbol{z}_1 & \boldsymbol{z}_2 & \cdots & \boldsymbol{z}_K\end{bmatrix}^T$.

3. COVARIANCE MATRIX ESTIMATION

3.1. Posterior distributions

In order to derive the MMSE estimator of $\boldsymbol{R}$, the first step is to find the posterior distribution $f(\boldsymbol{R}|\boldsymbol{Z})$. The joint posterior distribution of $\boldsymbol{\tau}$ and $\boldsymbol{R}$ can be written as

$$\begin{aligned} f(\boldsymbol{\tau},\boldsymbol{R}|\boldsymbol{Z}) &\propto f(\boldsymbol{Z}|\boldsymbol{\tau},\boldsymbol{R})f(\boldsymbol{\tau})f(\boldsymbol{R}) \\ &\propto |\boldsymbol{R}|^{-(\nu+K+N)}\,\text{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\bar{\boldsymbol{R}}\right\} \\ &\times \prod_{k=1}^{K}\tau_k^{-(q_k+N+1)}\exp\left\{-\tau_k^{-1}\left[\beta_k + \boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right]\right\} \end{aligned} \quad (4)$$

where $\propto$ means proportional to. Integrating with respect to $\boldsymbol{\tau}$, one obtains

$$\begin{aligned} f(\boldsymbol{R}|\boldsymbol{Z}) &= \int f(\boldsymbol{\tau},\boldsymbol{R}|\boldsymbol{Z})d\boldsymbol{\tau} \\ &\propto |\boldsymbol{R}|^{-(\nu+K+N)}\,\text{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\bar{\boldsymbol{R}}\right\} \\ &\times \prod_{k=1}^{K}\int \tau_k^{-(q_k+N+1)}\exp\left\{-\tau_k^{-1}\left[\beta_k + \boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right]\right\}d\tau_k \\ &\propto |\boldsymbol{R}|^{-(\nu+K+N)}\,\text{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\bar{\boldsymbol{R}}\right\} \\ &\times \prod_{k=1}^{K}\left[\beta_k + \boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right]^{-(q_k+N)}. \end{aligned} \quad (5)$$

3.2. MMSE estimation

The MMSE estimate of $\boldsymbol{R}$ is given by the posterior mean

$$\mathcal{E}\{\boldsymbol{R}|\boldsymbol{Z}\} = \int \boldsymbol{R}\,f(\boldsymbol{R}|\boldsymbol{Z})d\boldsymbol{R}. \quad (6)$$

Unfortunately, one does not know how to obtain this integral in analytical form. Furthermore, the distribution in (5) is not a classical distribution and one cannot either investigate generating samples drawn from (5) and averaging them to approximate the MMSE estimator. In contrast, as we shall see next, the conditional posterior distributions $f(\boldsymbol{R}|\boldsymbol{\tau},\boldsymbol{Z})$ and $f(\boldsymbol{\tau}|\boldsymbol{R},\boldsymbol{Z})$ are not only easy to obtain but also belong to familiar classes of distributions. In fact, from (4), it comes that

$$f(\boldsymbol{\tau}|\boldsymbol{R},\boldsymbol{Z}) \propto \prod_{k=1}^{K}\tau_k^{-(q_k+N+1)}\exp\left\{-\tau_k^{-1}\left[\beta_k + \boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right]\right\} \quad (7)$$

and hence

$$\tau_k|\boldsymbol{R},\boldsymbol{Z} \sim \mathcal{IG}\left(q_k+N, \beta_k + \boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right). \quad (8)$$

Accordingly,

$$\begin{aligned} f(\boldsymbol{R}|\boldsymbol{\tau},\boldsymbol{Z}) &\propto |\boldsymbol{R}|^{-(\nu+K+N)} \\ &\times \text{etr}\left\{-\boldsymbol{R}^{-1}\left[(\nu-N)\bar{\boldsymbol{R}} + \sum_{k=1}^{K}\tau_k^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^H\right]\right\} \end{aligned} \quad (9)$$

which implies that

$$\boldsymbol{R}|\boldsymbol{\tau},\boldsymbol{Z} \sim \mathcal{CW}^{-1}\left((\nu-N)\bar{\boldsymbol{R}} + \sum_{k=1}^{K}\tau_k^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^H, \nu+K\right). \quad (10)$$

Consequently, it is relatively standard to generate samples drawn from $f(\boldsymbol{R}|\boldsymbol{\tau},\boldsymbol{Z})$ and $f(\boldsymbol{\tau}|\boldsymbol{R},\boldsymbol{Z})$. This suggests the use of a Gibbs-sampler [16], which can be summarized by Table 1. Herein N_{bi} stands for the number of burn-in iterations and N_r is the number of samples which are effectively averaged. Some statistically sound criteria, such as the potential scale reduction factor [16], are available to select the values of N_{bi} and N_r that ensure convergence of the Gibbs sampler. The latter is known to generate random variables which are asymptotically distributed according to the posterior distributions $f(\boldsymbol{R}|\boldsymbol{Z})$ and $f(\boldsymbol{\tau}|\boldsymbol{Z})$, and therefore a natural way to approximate the MMSE estimator is to average the N_r last values generated by the sampler, i.e.

Table 1. Gibbs sampler

Input: $\boldsymbol{Z}$
1: generate initial value $\boldsymbol{R}^{(0)}$
2: **for** $n = 1,\cdots,N_{bi}+N_r$ **do**
3: generate $\boldsymbol{\tau}^{(n)}$ according to $f(\boldsymbol{\tau}|\boldsymbol{R}^{(n-1)},\boldsymbol{Z})$ in (8)
4: generate $\boldsymbol{R}^{(n)}$ according to $f(\boldsymbol{R}|\boldsymbol{\tau}^{(n)},\boldsymbol{Z})$ in (10)
5: **end for**
Output: sequence of random variables $\boldsymbol{\tau}^{(n)}$ and $\boldsymbol{R}^{(n)}$

$$\hat{\boldsymbol{R}}_{\text{MMSE}} = N_r^{-1}\sum_{n=N_{bi}+1}^{N_{bi}+N_r}\boldsymbol{R}^{(n)}. \quad (11)$$

3.3. MAP estimation

The (marginal) MAP estimator of $\boldsymbol{R}$ amounts to maximizing the distribution in (5), or equivalently its logarithm. Differentiating the latter with respect to $\boldsymbol{R}$ yields

$$\begin{aligned} \frac{\partial \ln f(\boldsymbol{R}|\boldsymbol{Z})}{\partial \boldsymbol{R}} &= -(\nu+K+N)\boldsymbol{R}^{-1} + (\nu-N)\boldsymbol{R}^{-1}\bar{\boldsymbol{R}}\boldsymbol{R}^{-1} \\ &+ \sum_{k=1}^{K}(q_K+N)\frac{\boldsymbol{R}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^H\boldsymbol{R}^{-1}}{\beta_k + \boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k}. \end{aligned} \quad (12)$$

It ensues that the MAP estimator of $\boldsymbol{R}$ obeys the following implicit equation

$$(\nu + K + N)\boldsymbol{R} = (\nu - N)\bar{\boldsymbol{R}} + \sum_{k=1}^{K} \frac{(q_K + N)\boldsymbol{z}_k \boldsymbol{z}_k^H}{\beta_k + \boldsymbol{z}_k^H \boldsymbol{R}^{-1} \boldsymbol{z}_k}. \quad (13)$$

It is instructive to note that this equation is very similar to that obtained via maximum likelihood estimation assuming that τ_k is unknown deterministic [7]. However, in the present scheme, we have introduced the a priori knowledge $\bar{\boldsymbol{R}}$ that counterbalances the influence of the snapshots. Indeed, the estimator is somehow a weighted combination of $\bar{\boldsymbol{R}}$ and the (properly compensated) sample covariance matrices of the snapshots. It can also be viewed as a kind of colored loading. Note that introducing diagonal loading in the iterative scheme of [7] has been proposed in [17]. Herein, this loading technique emerges naturally as the MAP estimator in a Bayesian setting. In order to solve (13), one can advocate an iterative procedure. In practice, since $\boldsymbol{R} = g(\boldsymbol{R}, \boldsymbol{Z})$, one starts with an initial value $\boldsymbol{R}^{(0)}$ and computes $\boldsymbol{R}^{(n+1)} = g(\boldsymbol{R}^{(n)}, \boldsymbol{Z})$ until convergence is achieved, see e.g. [7, 9].

Remark 1 An alternative to the above marginal MAP estimator consists in maximizing the *joint* density $f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z})$ in (4) with respect to $\boldsymbol{\tau}$ and $\boldsymbol{R}$. It is straightforward to show that, for a given $\boldsymbol{R}$, $f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z})$ can be maximized analytically, i.e. a closed-form expression for the value of $\boldsymbol{\tau}$ that maximizes $f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z})$ can be obtained. Plugging this value in $f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z})$, one is left with a maximization problem with respect to $\boldsymbol{R}$ only. Solving the latter leads to an equation which is very similar to (13), except that the factor $q_k + N$ in the right-hand side of (13) is replaced by $q_k + N + 1$. Hence, a similar iterative procedure can be employed to obtain the joint MAP estimator of $\boldsymbol{R}$.

4. NUMERICAL SIMULATIONS AND DISCUSSION

In this section, we evaluate the performance of the estimators derived previously. Towards this end, two different criteria will be used. The first measures a distance between $\hat{\boldsymbol{R}}$, a generic estimate of $\boldsymbol{R}$, and $\boldsymbol{R}$, and is given by [18]

$$d\left(\hat{\boldsymbol{R}}, \boldsymbol{R}\right) = \left[\sum_{k=1}^{N} \left(10 \log_{10} \lambda_k\right)^2\right]^{1/2} \quad (14)$$

where λ_k are the generalized eigenvalues of the matrix pencil $\hat{\boldsymbol{R}} - \lambda \boldsymbol{R}$. The metric in (14), in contrast to the usual mean-square error criterion, is the natural metric on the set of covariance matrices, see [18]. Since we are also interested in beamforming and detection applications, the second figure of merit will be the signal to noise ratio (SNR) loss at the output of the filter $\boldsymbol{w} \propto \hat{\boldsymbol{R}}^{-1}\boldsymbol{s}$, where $\boldsymbol{s}$ is the signature of the signal of interest (SOI). The SNR loss is defined as the ratio of the SNR at the output of the filter $\boldsymbol{w}$ to the optimal SNR obtained when $\boldsymbol{R}$ is known, viz

$$SNR_{\text{loss}} \triangleq \frac{|\boldsymbol{w}^H \boldsymbol{s}|^2}{\left(\boldsymbol{w}^H \boldsymbol{R} \boldsymbol{w}\right)\left(\boldsymbol{s}^H \boldsymbol{R}^{-1} \boldsymbol{s}\right)}. \quad (15)$$

It measures the capability of the filter to suppress noise and to retrieve/detect the SOI and serves as a good indicator for the detection performance. The MAP and MMSE estimators will be evaluated with respect to these two criteria. For comparison purposes, we also consider the normalized sample covariance matrix (NSCM) [6]

$$\hat{\boldsymbol{R}}_{\text{NSCM}} = \frac{N}{K} \sum_{k=1}^{K} \frac{\boldsymbol{z}_k \boldsymbol{z}_k^H}{\boldsymbol{z}_k^H \boldsymbol{z}_k} \quad (16)$$

which is commonly used in compound-Gaussian clutter. Accordingly, we consider the estimator $\hat{\boldsymbol{R}} = \bar{\boldsymbol{R}}$ in order to evaluate whether the a priori information alone can provide good performance, and if there is an interest to use the posterior information. We consider a scenario with $N = 8$ and $q = 4$. The value of ν is either $\nu = 10$ (weak a priori knowledge) and $\nu = 16$ (more precise a priori knowledge). The number of samples is varied between $K = N$ and $K = 5N$. The MAP estimator is initialized with the NSCM estimate, and the iterative scheme of (13) is run 10 times. Regarding the MMSE, $N_{bi} = 20$ and $N_r = 100$. The results are displayed in

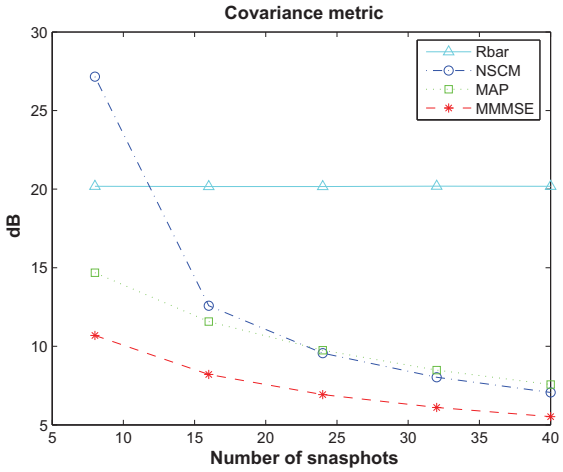

Fig. 1. Covariance metric for estimation of $\boldsymbol{R}$ versus number of snapshots. $N = 8$, $\nu = 10$ and $q = 4$.

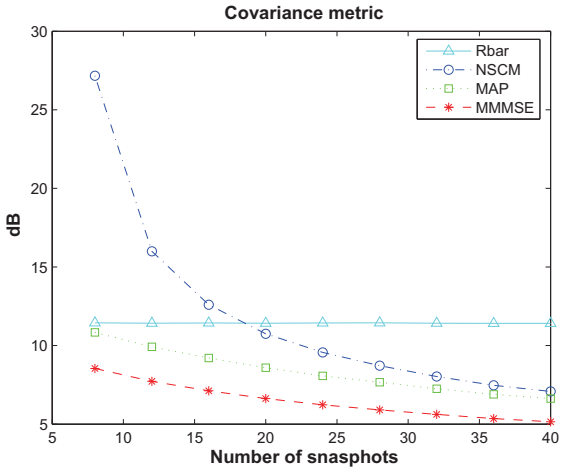

Fig. 2. Covariance metric for estimation of $\boldsymbol{R}$ versus number of snapshots. $N = 8$, $\nu = 16$ and $q = 4$.

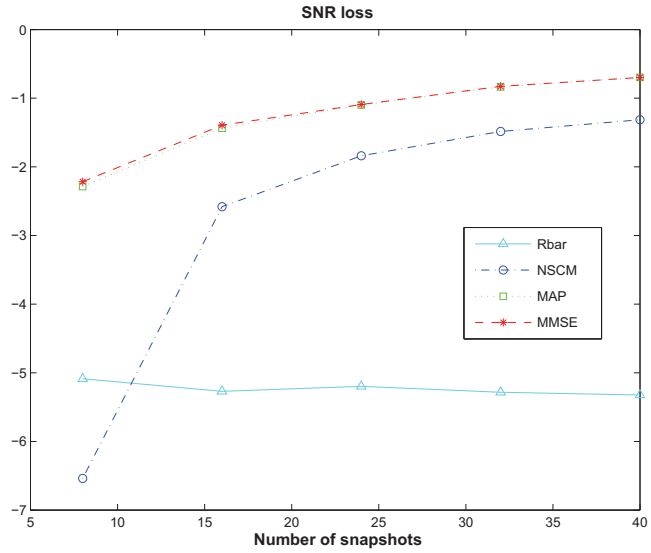

Fig. 3. SNR loss versus number of snapshots. $N = 8$, $\nu = 10$ and $q = 4$.

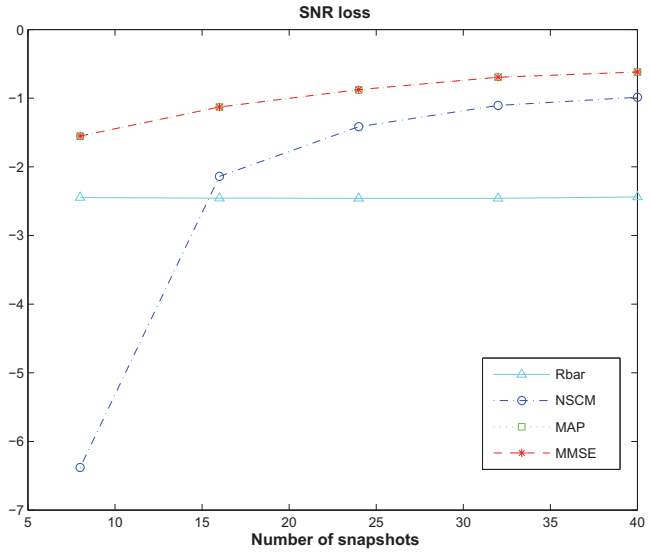

Fig. 4. SNR loss versus number of snapshots. $N = 8$, $\nu = 16$ and $q = 4$.

Figures 1 to 4. We first observe that the two criteria result approximately in the same hierarchy between the estimators. The MMSE estimator provides the best performance for both criteria. The MAP estimator performs slightly worse in terms of covariance metric but is equivalent to the MMSE estimator in terms of SNR loss, which is a very appealing feature since the computational complexity of the MAP estimator is less than that of the MMSE estimator. Compared to the NSCM, the MMSE and MAP estimators require a smaller sample support to achieve the same SNR loss: for instance even for $K = N$ their SNR loss is below 2dB, while the NSCM requires $2N$ snapshots to achieve the optimal SNR up to -3dB. Finally, when $\bar{\boldsymbol{R}}$ is used as an estimate of $\boldsymbol{R}$ it results in poor performance (especially for $\nu = 10$), indicating that the a priori knowledge by itself is not accurate enough to provide satisfactory results. It shows that $\boldsymbol{R}$ can be quite different from $\bar{\boldsymbol{R}}$ and that use of the data, via the posterior distributions, is helpful in improving MSE or SNR loss.

5. REFERENCES

[1] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions Aerospace Electronic Systems*, vol. 28, no. 1, pp. 208–216, January 1992.

[2] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-Gaussian clutter," *IEEE Transactions Aerospace Electronic Systems*, vol. 31, no. 2, pp. 617–625, April 1995.

[3] ——, "Adaptive matched filter detection in spherically invariant noise," *IEEE Signal Processing Letters*, vol. 3, no. 8, pp. 248–250, August 1996.

[4] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant process," *IEE Proceedings Radar, Sonar and Navigation*, vol. 134, no. 2, pp. 191–197, April 1987.

[5] E. Conte, A. De Maio, and C. Galdi, "Statistical analysis of real clutter at different range resolutions," *IEEE Transactions Aerospace Electronic Systems*, vol. 40, no. 3, pp. 903–918, July 2004.

[6] E. Conte, A. De Maio, and G. Ricci, "Covariance matrix estimation for adaptive CFAR detection in compound-gaussian clutter," *IEEE Transactions Aerospace Electronic Systems*, vol. 38, no. 2, pp. 415–426, April 2002.

[7] ——, "Recursive estimation of the covariance matrix of a compound-gaussian process and its application to adaptive CFAR detection," *IEEE Transactions Signal Processing*, vol. 50, no. 8, pp. 1908–1915, August 2002.

[8] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, December 2002.

[9] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis," *IEEE Transactions Signal Processing*, vol. 56, no. 1, pp. 34–48, January 2008.

[10] W. L. Melvin and J. R. Guerci, "Knowledge-aided signal processing: A new paradigm for radar and other advanced sensors," *IEEE Transactions Aerospace Electronic Systems*, vol. 42, no. 3, pp. 983–996, July 2006.

[11] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, Tech. Rep. 1015, December 1994.

[12] S. Bidon, O. Besson, and J.-Y. Tourneret, "A Bayesian approach to adaptive detection in non-homogeneous environments," *IEEE Transactions Signal Processing*, vol. 56, no. 1, pp. 205–217, January 2008.

[13] O. Besson, S. Bidon, and J.-Y. Tourneret, "Covariance matrix estimation with heterogeneous samples," *IEEE Transactions Signal Processing*, vol. 56, no. 3, pp. 909–920, March 2008.

[14] C. P. Robert, *The Bayesian Choice - From Decision-Theoretic Foundations to Computational Implementation*. New York: Springer Verlag, 2007.

[15] L. Svensson, "Bayesian inference with unknown noise covariance," Ph.D. dissertation, Chalmers University of Technology, Göteborg, November 2004.

[16] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. New York: Chapman & Hall, 2004.

[17] Y. I. Abramovich and N. K. Spencer, "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," in *Proceedings ICASSP*, Honolulu, HI, April 2007, pp. 1105–1108.

[18] S. T. Smith, "Covariance, subspace and intrinsic Cramér-Rao bounds," *IEEE Transactions Signal Processing*, vol. 53, no. 5, pp. 1610–1630, May 2005.