Adaptive Processing With Signal Contaminated Training Samples

Olivier Besson, *Senior Member, IEEE*, and Stéphanie Bidon, *Member, IEEE*

***Abstract*—We consider the adaptive beamforming or adaptive detection problem in the case of signal contaminated training samples, i.e., when the latter may contain a signal-like component. Since this results in a significant degradation of the signal to interference and noise ratio at the output of the adaptive filter, we investigate a scheme to jointly detect the contaminated samples and subsequently take this information into account for estimation of the disturbance covariance matrix. Towards this end, a Bayesian model is proposed, parameterized by binary variables indicating the presence/absence of signal-like components in the training samples. These variables, together with the signal amplitudes and the disturbance covariance matrix are jointly estimated using a minimum mean-square error (MMSE) approach. Two strategies are proposed to implement the MMSE estimator. First, a stochastic Markov Chain Monte Carlo method is presented based on Gibbs sampling. Then a computationally more efficient scheme based on variational Bayesian analysis is proposed. Numerical simulations attest to the improvement achieved by this method compared to conventional methods such as diagonal loading. A successful application to real radar data is also presented.**

***Index Terms*—Bayesian estimation, outliers, radar detection, robust adaptive filtering.**

I. Problem Statement

OPTIMAL multi-channel processing, either for beamforming or detection purposes, requires the knowledge of the disturbance covariance matrix $\boldsymbol{R}$, so as to cancel out this disturbance and retrieve the signal of interest (SOI) with maximum signal to interference and noise (SINR) ratio [1]. In practical situations however, this covariance matrix is unknown and needs to be estimated from a set of training samples $\boldsymbol{z}_k$, $k = 1, \ldots, K$. In an ideal situation, i.e., with independent and identically distributed samples sharing the same covariance matrix $\boldsymbol{R}$, Reed Mallet and Brennan in [2] showed that $K = 2N - 3$ samples, where N stands for the number of channels, are necessary for the average SINR of the adaptive processor to be within 3 dB of the optimal SINR. This rate of convergence is a crucial parameter since, in many cases, the disturbance characteristics might vary, and hence it is highly desirable to converge with minimal K. However the rate of convergence is sensitive to non-homogeneity among the training samples and the assumption that the training samples share a common covariance matrix is questionable in many situations [3]. Numerous factors can contribute to non-homogeneity of the training samples, including clutter heterogeneity (e.g., dense scattering environments, land-sea clutter interfaces, power level fluctuations among the various patches of clutter) or contamination of the training samples by signal-like components (e.g., in the case of multiple closely-spaced targets with approximately the same velocity) [4]. In this paper, we focus on the latter problem, i.e., the presence in the training samples of outliers with a signature close to that of the SOI. This phenomenon has a detrimental effect on the performance of the adaptive processor, even when a limited number of samples are contaminated [4]. In the latter reference, the probability of detection and probability of ghosting of Kelly's generalized likelihood ratio test and the mean level adaptive detector are studied, in the case where only one sample out of K is contaminated. Even with this small proportion of outliers, some degradation is observed. Note that if all training samples contain a signal component, the number of samples required for convergence of the adaptive processor is highly increased and is typically of the order $(N-1) \times SINR_{\text{opt}}$[5], [6]. In fact, since the adaptive weight vector lies in a space orthogonal to that of the interferences, if a signal with the same signature as the SOI is present in the training samples, the adaptive filter is likely to place nulls towards the SOI, resulting in a significant SINR loss. The usual approach to cope with this problem is to select the most homogeneous set among the training samples and to censor the other samples in the covariance matrix estimation procedure, see e.g., [7] for derivation of outliers resistant adaptive schemes. Very often, a test statistic is formed for each sample and compared to a threshold: only the samples whose test statistics do not exceed the threshold are retained. Several test statistics for detecting non-homogeneous samples have been proposed in the literature, e.g., power selection criteria [8], the generalized inner-product [9], the adaptive power residue [10] or the non-homogeneity detector [11]. Refinements of these methods include a re-iterative censoring of the samples, see e.g., [10], [12], [13]. These methods perform well but may need a fairly large amount of initial data. In [14] an outlier-resistant adaptive beamforming scheme is proposed, which does not censor any sample. The author considers a compound-Gaussian model for the noise and the usual iterative scheme [15] which is known to converge to the maximum likelihood estimate of the noise covariance matrix [16]. The idea

The authors are with the Department of Electronics, Optronics and Signal of Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), University of Toulouse, ISAE, Toulouse, 31055 France (e-mail: olivier.besson@isae.fr; stephanie.bidon@isae.fr).

is to introduce diagonal loading within the iterations in order to obtain an improved robustness in the case where signal-like components contaminate the training samples.

In this paper, we also investigate an approach where all samples are used but, at the same time, we try to detect whether a sample contains a signal-like component: if this is the case, the contribution of the latter to the training sample is somehow removed, but the concerned snapshot is not censored. Therefore, our approach consists of a joint signal detection-covariance matrix estimation scheme. Its principle relies on introducing binary variables which indicate the presence or the absence of a signal-like component. These variables are assumed random and therefore a Bayesian framework is favored here, where both the presence indicators, the corresponding amplitudes and the covariance matrix are jointly estimated. All variables are assigned some (non-informative) priors and the main contribution of the paper lies in deriving the minimum mean-square error (MMSE) estimator of $\boldsymbol{R}$ under this framework. A Markov Chain Monte Carlo (MCMC) simulation method, based on Gibbs sampling, is first proposed. Then a computationally more efficient scheme, based on variational Bayesian (VB) analysis, is presented.

II. Model Assumptions

Let us assume that we have K training samples $\boldsymbol{z}_k$ which consist of noise and possibly a signal-like component, so that they can be modeled as

$$\boldsymbol{z}_k = \boldsymbol{n}_k + \tilde{\alpha}_k^* \boldsymbol{v} \tag{1}$$

where $\boldsymbol{v}$ stands for the signature of the signal of interest and the following hypotheses are made.

The noise vectors $\boldsymbol{n}_k$ are assumed to be independent, complex-valued Gaussian distributed, with zero-mean and covariance matrix $\boldsymbol{R}$, i.e.,

$$p\left(\boldsymbol{n}_k|\boldsymbol{R}\right) = \pi^{-N}|\boldsymbol{R}|^{-1} e^{-\boldsymbol{n}_k^H \boldsymbol{R}^{-1} \boldsymbol{n}_k}. \tag{2}$$

As said previously, we adhere to a Bayesian framework and hence, we consider that $\boldsymbol{R}$ is a random variable, with a conjugate prior distribution relative to (2), namely an inverse Wishart distribution with mean $\bar{\boldsymbol{R}}$ and ν degrees of freedom, viz

$$\pi\left(\boldsymbol{R}|\nu, \bar{\boldsymbol{R}}\right) \propto |\boldsymbol{R}|^{-(\nu+N)} \mathrm{etr}\left\{-(\nu-N)\bar{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\} \tag{3}$$

where $\propto$ means proportional to and $\mathrm{etr}\{.\}$ stands for the exponential of the trace of the matrix between braces.We consider a situation where no accurate information about $\boldsymbol{R}$ is available that would help to choose a specific $\bar{\boldsymbol{R}}$. In fact, since it will amount to diagonal loading (see below), a method which is known to be robust against various mismatches, we will mostly consider the case $\bar{\boldsymbol{R}} \propto \boldsymbol{I}$ in the sequel.

The complex amplitudes $\tilde{\alpha}_k$ of the signal-like components in $\boldsymbol{z}_k$ are assumed to be independent, and distributed according to a Bernouilli-Gaussian distribution:

$$\pi\left(\tilde{\alpha}_k|p, \sigma_\alpha^2\right) = (1-p)\delta(\tilde{\alpha}_k) + p\pi^{-1}\sigma_\alpha^{-2} e^{-\sigma_\alpha^{-2}|\tilde{\alpha}_k|^2}. \tag{4}$$

As it will be more convenient, especially in the variational Bayesian analysis, we will use a statistically equivalent model for $\tilde{\alpha}_k$, namely $\tilde{\alpha}_k = i_k \alpha_k$ where i_k and α_k are independent random variables. The binary variables $i_k \in \{0,1\}$ indicate the presence or the absence of a signal-like component in the training samples. They follow a Bernoulli distribution, denoted as $\mathsf{Ber}\,(p)$, which is given by $\pi(i_k|p) = p^{i_k}(1-p)^{1-i_k}$. The amplitudes α_k are independent complex-valued Gaussian distributed random variables with zero mean and (known) variance σ_α^2, i.e., $\pi\left(\alpha_k|\sigma_\alpha^2\right) = \pi^{-1}\sigma_\alpha^{-2} e^{-\sigma_\alpha^{-2}|\alpha_k|^2}$. The distribution of vector $\boldsymbol{\alpha} = [\,\alpha_1 \quad \cdots \quad \alpha_K\,]^T$ is denoted as $\boldsymbol{\alpha} \sim \mathsf{CN}\left(\boldsymbol{0}, \sigma_\alpha^2 \boldsymbol{I}\right)$. In the sequel, we assume that the probability of contaminated samples p is known. Extension to the case where p is itself an unknown random variable with non informative uniform prior will be dealt with in the Appendix where we proceed to the generalization of the estimators derived below. Also, a sensitivity analysis will be conducted in Section V in order to assess the robustness of the estimators to a non perfect knowledge of p. Accordingly, the extension to the case where σ_α^2 is a random variable with non-informative prior will be investigated in the Appendix.

As stated in the introduction, our main objective is to obtain an estimate $\hat{\boldsymbol{R}}$ of $\boldsymbol{R}$ –or directly an estimate of $\boldsymbol{R}^{-1}$– so as to compute a filter $\boldsymbol{w} \propto \hat{\boldsymbol{R}}^{-1}\boldsymbol{v}$. Towards this end, a systematic approach consists in deriving the minimum mean-square error (MMSE) estimator of $\boldsymbol{R}$, which entails computing the posterior mean of $\boldsymbol{R}$. Therefore, as a preliminary step, one needs to obtain an expression for $p\left(\boldsymbol{R}|\boldsymbol{Z}\right)$ where $\boldsymbol{Z} = [\,\boldsymbol{z}_1 \quad \cdots \quad \boldsymbol{z}_K\,]$. Under the hypotheses made, the likelihood function is given by

$$\begin{aligned} p\left(\boldsymbol{Z}|\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}\right) &= \prod_{k=1}^{K} \pi^{-N}|\boldsymbol{R}|^{-1} e^{-(\boldsymbol{z}_k - i_k\alpha_k^*\boldsymbol{v})^H \boldsymbol{R}^{-1}(\boldsymbol{z}_k - i_k\alpha_k^*\boldsymbol{v})} \\ &= \pi^{-NK}|\boldsymbol{R}|^{-K} \mathrm{etr}\Bigg\{-\boldsymbol{R}^{-1}\left(\boldsymbol{Z} - \boldsymbol{v}(\boldsymbol{i}\odot\boldsymbol{\alpha})^H\right) \\ &\qquad \times \left(\boldsymbol{Z} - \boldsymbol{v}(\boldsymbol{i}\odot\boldsymbol{\alpha})^H\right)^H\Bigg\} \end{aligned} \tag{5}$$

where $\boldsymbol{i} = [\,i_1 \quad \cdots \quad i_K\,]^T$ and $i_k = 1$ if $\boldsymbol{z}_k$ is contaminated and 0 otherwise. In the previous equation $\odot$ stands for the Hadamard (element-wise) product. Note that $\boldsymbol{i}\odot\boldsymbol{\alpha} = \boldsymbol{D_i}\boldsymbol{\alpha}$ where $\boldsymbol{D_i} = \mathrm{diag}\,(i_1, \ldots, i_K)$. In order to obtain $p\left(\boldsymbol{R}|\boldsymbol{Z}\right)$, we use Bayes rule, viz $p\left(\boldsymbol{R}|\boldsymbol{Z}\right) \propto p\left(\boldsymbol{Z}|\boldsymbol{R}\right)\pi\left(\boldsymbol{R}\right)$ and hence first derive $p\left(\boldsymbol{Z}|\boldsymbol{R}\right)$. Towards this end, we need to marginalize $p\left(\boldsymbol{Z}|\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}\right)$ with respect to (w.r.t.) $\boldsymbol{i}$ and $\boldsymbol{\alpha}$. From (5) and using the fact that $i_k \sim \mathsf{Ber}\,(p)$, it follows that

$$\begin{aligned} p\left(\boldsymbol{Z}|\boldsymbol{\alpha}, \boldsymbol{R}\right) = \prod_{k=1}^{K} \Big[&(1-p)\pi^{-N}|\boldsymbol{R}|^{-1} e^{-\boldsymbol{z}_k^H \boldsymbol{R}^{-1}\boldsymbol{z}_k} \\ &+ p\pi^{-N}|\boldsymbol{R}|^{-1} e^{-(\boldsymbol{z}_k - \alpha_k^*\boldsymbol{v})^H \boldsymbol{R}^{-1}(\boldsymbol{z}_k - \alpha_k^*\boldsymbol{v})}\Big]. \end{aligned} \tag{6}$$

Next, observing that

$$\begin{aligned} &\sigma_\alpha^{-2}|\alpha_k|^2 + (\boldsymbol{z}_k - \alpha_k^*\boldsymbol{v})^H \boldsymbol{R}^{-1}(\boldsymbol{z}_k - \alpha_k^*\boldsymbol{v}) \\ &\quad = \left(\sigma_\alpha^{-2} + \boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}\right)\left|\alpha_k - \frac{\boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{v}}{\sigma_\alpha^{-2} + \boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}}\right|^2 \end{aligned}$$

$$+\boldsymbol{z}_k^H\left(\boldsymbol{R}+\sigma_\alpha^2\boldsymbol{v}\boldsymbol{v}^H\right)^{-1}\boldsymbol{z}_k \tag{7}$$

we can write

$$\begin{aligned}
p(\boldsymbol{Z}|\boldsymbol{R}) &= \int p\left(\boldsymbol{Z}|\boldsymbol{\alpha},\boldsymbol{R}\right)\pi\left(\boldsymbol{\alpha}\right)d\boldsymbol{\alpha} \\
&= \prod_{k=1}^{K}\int \pi^{-1}\sigma_\alpha^{-2}e^{-\sigma_\alpha^{-2}|\alpha_k|^2}\Big[(1-p)\pi^{-N}|\boldsymbol{R}|^{-1}e^{-\boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k} \\
&\qquad + p\pi^{-N}|\boldsymbol{R}|^{-1}e^{-(\boldsymbol{z}_k-\alpha_k^*\boldsymbol{v})^H\boldsymbol{R}^{-1}(\boldsymbol{z}_k-\alpha_k^*\boldsymbol{v})}\Big]\,d\alpha_k \\
&= \prod_{k=1}^{K}\Big[(1-p)\pi^{-N}|\boldsymbol{R}|^{-1}e^{-\boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k} \\
&\qquad + p\pi^{-N}|\boldsymbol{R}+\sigma_\alpha^2\boldsymbol{v}\boldsymbol{v}^H|^{-1}e^{-\boldsymbol{z}_k^H(\boldsymbol{R}+\sigma_\alpha^2\boldsymbol{v}\boldsymbol{v}^H)^{-1}\boldsymbol{z}_k}\Big]
\end{aligned} \tag{8}$$

where we used the fact that $|\boldsymbol{R}+\sigma_\alpha^2\boldsymbol{v}\boldsymbol{v}^H| = \sigma_\alpha^2|\boldsymbol{R}|\left(\sigma_\alpha^{-2}+\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}\right)$. The posterior distribution of $\boldsymbol{R}$ only is finally obtained as

$$\begin{aligned}
p\left(\boldsymbol{R}|\boldsymbol{Z}\right) &\propto p\left(\boldsymbol{Z}|\boldsymbol{R}\right)\pi\left(\boldsymbol{R}\right) \\
&\propto |\boldsymbol{R}|^{-(\nu+N)}\mathrm{etr}\left\{-(\nu-N)\bar{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\} \\
&\times \prod_{k=1}^{K}\Big[(1-p)\pi^{-N}|\boldsymbol{R}|^{-1}e^{-\boldsymbol{z}_k^H\boldsymbol{R}^{-1}\boldsymbol{z}_k} \\
&\qquad + p\pi^{-N}|\boldsymbol{R}+\sigma_\alpha^2\boldsymbol{v}\boldsymbol{v}^H|^{-1}\,e^{-\boldsymbol{z}_k^H\left(\boldsymbol{R}+\sigma_\alpha^2\boldsymbol{v}\boldsymbol{v}^H\right)^{-1}\boldsymbol{z}_k}\Big].
\end{aligned} \tag{9}$$

The MMSE estimator is given by the posterior mean, i.e., $\int \boldsymbol{R}p\left(\boldsymbol{R}|\boldsymbol{Z}\right)d\boldsymbol{R}$. From inspection of (9), it does not seem possible to obtain a closed-form expression for the posterior mean. Moreover, generating samples drawn from $p\left(\boldsymbol{R}|\boldsymbol{Z}\right)$ and approximating the integral by an arithmetic mean is not feasible as the posterior distribution in (9) does not belong to any known class of distributions. Therefore, we investigate two different approaches namely Gibbs sampling and variational Bayesian analysis which work on the joint posterior distributions of $\boldsymbol{i}$, $\boldsymbol{\alpha}$ and $\boldsymbol{R}$. In addition to leading to tractable distributions, these approaches enable one to obtain estimates of $\boldsymbol{i}$ and $\boldsymbol{\alpha}$ and hence to detect the contaminated samples together with their amplitudes. Thus they provide additional information compared to estimating $\boldsymbol{R}$ only.

III. Gibbs Sampler

As evidenced in (9), $p\left(\boldsymbol{R}|\boldsymbol{Z}\right)$ does not lead to a closed-form expression for $\int \boldsymbol{R}p\left(\boldsymbol{R}|\boldsymbol{Z}\right)d\boldsymbol{R}$ and it does not seem to be trivial to generate samples according to this distribution. Therefore, resorting to a Gibbs sampler appears to be a judicious way to proceed [17], [18]. The principle of the Gibbs sampler is to successively generate samples distributed according to the posterior distribution of one variable, conditioned on all other variables, as usually the conditional posterior distributions of each variable are easier to simulate. Doing so, it is known that the samples so generated will be asymptotically distributed according to the posterior distribution of each variable, so that, for instance, the MMSE estimator can be approximated by the arithmetic mean of the set of samples drawn. In our case, this amounts to derive $p\left(\boldsymbol{\alpha}|\boldsymbol{i},\boldsymbol{R},\boldsymbol{Z}\right)$, $p\left(\boldsymbol{i}|\boldsymbol{\alpha},\boldsymbol{R},\boldsymbol{Z}\right)$ and $p\left(\boldsymbol{R}|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{Z}\right)$. The starting point of the derivation is the joint posterior distribution $p\left(\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R}|\boldsymbol{Z}\right)$ which, under the stated assumptions, is given by

$$\begin{aligned}
p\left(\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R}|\boldsymbol{Z}\right) &\propto p\left(\boldsymbol{Z}|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R}\right)\pi\left(\boldsymbol{i}\right)\pi\left(\boldsymbol{\alpha}\right)\pi\left(\boldsymbol{R}\right) \\
&\propto |\boldsymbol{R}|^{-(\nu+K+N)}\mathrm{etr}\left\{-(\nu-N)\bar{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\} \\
&\quad\times\prod_{k=1}^{K}p^{i_k}(1-p)^{1-i_k}e^{-\sigma_\alpha^{-2}|\alpha_k|^2} \\
&\quad\times e^{-(\boldsymbol{z}_k-i_k\alpha_k^*\boldsymbol{v})^H\boldsymbol{R}^{-1}(\boldsymbol{z}_k-i_k\alpha_k^*\boldsymbol{v})}.
\end{aligned} \tag{10}$$

It is straightforward to infer that

$$\begin{aligned}
p(\boldsymbol{i}|\boldsymbol{\alpha},\boldsymbol{R},\boldsymbol{Z}) &\propto \prod_{k=1}^{K}p^{i_k}(1-p)^{1-i_k}e^{-(\boldsymbol{z}_k-i_k\alpha_k^*\boldsymbol{v})^H\boldsymbol{R}^{-1}(\boldsymbol{z}_k-i_k\alpha_k^*\boldsymbol{v})} \\
&\propto \prod_{k=1}^{K}p^{i_k}(1-p)^{1-i_k}e^{-i_k\left[|\alpha_k|^2\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}-2\mathrm{Re}\left(\alpha_k\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right)\right]}
\end{aligned} \tag{11}$$

where we used the fact that $i_k^2 = i_k$. It follows that

$$i_k|\boldsymbol{\alpha},\boldsymbol{R},\boldsymbol{Z} \sim \mathsf{Ber}\left(\Pr\left[i_k=1|\boldsymbol{\alpha},\boldsymbol{R},\boldsymbol{Z}\right]\right) \tag{12}$$

with

$$\begin{aligned}
&\Pr\left[i_k=1|\boldsymbol{\alpha},\boldsymbol{R},\boldsymbol{Z}\right] \\
&\qquad = \frac{pe^{-|\alpha_k|^2\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}+2\mathrm{Re}\left(\alpha_k\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right)}}{(1-p)+pe^{-|\alpha_k|^2\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}+2\mathrm{Re}\left(\alpha_k\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{z}_k\right)}}.
\end{aligned} \tag{13}$$

Accordingly, (10) implies that

$$\begin{aligned}
p\left(\boldsymbol{\alpha}|\boldsymbol{i},\boldsymbol{R},\boldsymbol{Z}\right) &\propto e^{-\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}} \\
&\times \mathrm{etr}\left\{-\boldsymbol{R}^{-1}\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)^H\right\} \\
&\propto e^{-\boldsymbol{\alpha}^H\left[\sigma_\alpha^{-2}\boldsymbol{I}+\left(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}\right)\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\right]\boldsymbol{\alpha}+\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{Z}^H\boldsymbol{R}^{-1}\boldsymbol{v}+\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{D}_{\boldsymbol{i}}\boldsymbol{\alpha}}
\end{aligned} \tag{14}$$

and hence $\boldsymbol{\alpha}|\boldsymbol{i},\boldsymbol{R},\boldsymbol{Z}$ is Gaussian distributed

$$\boldsymbol{\alpha}|\boldsymbol{i},\boldsymbol{R},\boldsymbol{Z} \sim \mathsf{CN}\left(\boldsymbol{\mu}_\alpha\left(\boldsymbol{i},\boldsymbol{R}^{-1},\boldsymbol{Z}\right),\boldsymbol{\Sigma}_\alpha\left(\boldsymbol{i},\boldsymbol{R}^{-1}\right)\right) \tag{15}$$

with mean $\boldsymbol{\mu}_\alpha\left(\boldsymbol{i},\boldsymbol{R}^{-1},\boldsymbol{Z}\right)$ and covariance matrix $\boldsymbol{\Sigma}_\alpha\left(\boldsymbol{i},\boldsymbol{R}^{-1}\right)$ given by

$$\boldsymbol{\Sigma}_\alpha\left(\boldsymbol{i},\boldsymbol{R}^{-1}\right) = \left[\sigma_\alpha^{-2}\boldsymbol{I}+\left(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}\right)\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\right]^{-1} \tag{16a}$$

$$\boldsymbol{\mu}_\alpha\left(\boldsymbol{i},\boldsymbol{R}^{-1},\boldsymbol{Z}\right) = \boldsymbol{\Sigma}_\alpha\left(\boldsymbol{i},\boldsymbol{R}^{-1}\right)\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{Z}^H\boldsymbol{R}^{-1}\boldsymbol{v}. \tag{16b}$$

Finally, let us consider the conditional posterior distribution of $\boldsymbol{R}$:

$$\begin{aligned}
p\left(\boldsymbol{R}|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{Z}\right) &\propto |\boldsymbol{R}|^{-(\nu+K+N)}\mathrm{etr}\left\{-(\nu-N)\bar{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\} \\
&\quad\times\mathrm{etr}\left\{-\boldsymbol{R}^{-1}\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)^H\right\}
\end{aligned} \tag{17}$$

which is recognized as an inverse Wishart distribution [19] with $\nu+K$ degrees of freedom and parameter matrix

$$\boldsymbol{M}(\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{Z}) = (\nu-N)\bar{\boldsymbol{R}}+\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)^H. \tag{18}$$

TABLE I
GIBBS SAMPLER FOR ESTIMATION OF $\boldsymbol{i}$, $\boldsymbol{\alpha}$ AND $\boldsymbol{R}^{-1}$

Input: initial values $\boldsymbol{R}^{-1}(0)$, $\boldsymbol{i}(0)$
1: **for** $n = 1, \cdots, N_{\text{bi}} + N_r$ **do**
2: sample $\boldsymbol{\alpha}(n)$ from $p\left(\boldsymbol{\alpha}|\boldsymbol{i}(n-1), \boldsymbol{R}(n-1), \boldsymbol{Z}\right)$.
3: sample $\boldsymbol{R}^{-1}(n)$ from $p\left(\boldsymbol{R}^{-1}|\boldsymbol{i}(n-1), \boldsymbol{\alpha}(n), \boldsymbol{Z}\right)$.
4: for $k = 1, \cdots, K$, sample $i_k(n)$ from $\mathsf{Ber}\left(\Pr\left[i_k = 1|\boldsymbol{\alpha}(n), \boldsymbol{R}^{-1}(n), \boldsymbol{Z}\right]\right)$.
5: **end for**
Output: sequence of random variables $\boldsymbol{i}(n)$, $\boldsymbol{\alpha}(n)$, $\boldsymbol{R}^{-1}(n)$.

At this stage, it should be noted that both $p\left(\boldsymbol{i}|\boldsymbol{\alpha}, \boldsymbol{R}, \boldsymbol{Z}\right)$ and $p\left(\boldsymbol{\alpha}|\boldsymbol{i}, \boldsymbol{R}, \boldsymbol{Z}\right)$ depend on $\boldsymbol{R}$ through its inverse $\boldsymbol{R}^{-1}$. Note also that the optimal filter depends directly on $\boldsymbol{R}^{-1}$. Therefore, it is more natural (and convenient) to work with the precision matrix rather than with the covariance matrix. Towards this end we need to derive the posterior distribution of $\boldsymbol{R}^{-1}$, conditioned on $\boldsymbol{i}$, $\boldsymbol{\alpha}$ and $\boldsymbol{Z}$: from (17), this posterior distribution is given by

$$\boldsymbol{R}^{-1}|\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{Z} \sim \mathsf{CW}\left(\nu + K, \left[\boldsymbol{M}\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{Z}\right)\right]^{-1}\right). \tag{19}$$

It is important to note that all three distributions in (12), (15) and (19) are well-known distributions and hence it will be rather straightforward to draw samples from them. Our Gibbs sampling scheme is summarized in Table I. Once the matrices $\boldsymbol{R}^{-1}(n)$ are computed, the MMSE estimator of $\boldsymbol{R}^{-1}$ can be approximated as

$$\hat{\boldsymbol{R}}^{-1}_{\text{mmse}} = \frac{1}{N_r} \sum_{n=N_{\text{bi}}+1}^{N_{\text{bi}}+N_r} \boldsymbol{R}^{-1}(n) \tag{20}$$

where N_{bi} stands for the number of burn-in iterations and N_r is the effective number of iterations. The Gibbs sampler also allows to obtain an approximate maximum a posteriori (MAP) estimator of $\boldsymbol{i}$, $\boldsymbol{\alpha}$ and $\boldsymbol{R}$ by selecting the triplet $\left\{\boldsymbol{i}(n), \boldsymbol{\alpha}(n), \boldsymbol{R}^{-1}(n)\right\}$ which results in the maximum value of the joint posterior distribution.

IV. VARIATIONAL BAYESIAN ANALYSIS

Despite its usually good performance, the Gibbs sampler may suffer from slow convergence, i.e., a large number of samples may be required for the Markov chain to converge and the sampler to provide accurate estimates. This results in an increased computational cost, a non desirable feature. In order to improve convergence speed, variational Bayesian methods have recently emerged [20]–[22]. Briefly stated, the essence of variational Bayesian analysis is to approximate the posterior distribution by a product of distributions, each one of the latter involving only some variables. More precisely, in our case the goal is to approximate $p\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right)$ as

$$p\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right) \simeq q\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right) = q_i(\boldsymbol{i}|\boldsymbol{Z}) q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z}) q_R(\boldsymbol{R}|\boldsymbol{Z}). \tag{21}$$

One then seeks the individual distributions $q_i(.)$, $q_\alpha(.)$ and $q_R(.)$ of (virtually) independent random variables $\boldsymbol{i}$, $\boldsymbol{\alpha}$ and $\boldsymbol{R}$ such that the resulting joint distribution of $\boldsymbol{i}$, $\boldsymbol{\alpha}$ and $\boldsymbol{R}$ is as close as possible to $p\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right)$. Observe, that if this approximation is accurate, the MMSE estimator of say parameter vector $\boldsymbol{i}$ is given by $\int \boldsymbol{i} p\left(\boldsymbol{i}|\boldsymbol{Z}\right) d\boldsymbol{i} \simeq \int \boldsymbol{i} q_i(\boldsymbol{i}) d\boldsymbol{i}$. The approximation is obtained by minimizing the Kullback-Leibler divergence between the two distributions, and it can be shown that the solution satisfies the following set of equations [20]

$$\ln q_i(\boldsymbol{i}|\boldsymbol{Z}) = \text{const.} + \left\langle \ln p\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right)\right\rangle_{\boldsymbol{\alpha},\boldsymbol{R}} \tag{22a}$$

$$\ln q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z}) = \text{const.} + \left\langle \ln p\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right)\right\rangle_{\boldsymbol{i},\boldsymbol{R}} \tag{22b}$$

$$\ln q_R(\boldsymbol{R}|\boldsymbol{Z}) = \text{const.} + \left\langle \ln p\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right)\right\rangle_{\boldsymbol{i},\boldsymbol{\alpha}} \tag{22c}$$

where the expectations $\langle . \rangle_{.}$ should be understood w.r.t. $q\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right)$, e.g.,

$$\left\langle \ln p\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right)\right\rangle_{\boldsymbol{\alpha},\boldsymbol{R}} = \int\int \ln p\left(\boldsymbol{i}, \boldsymbol{\alpha}, \boldsymbol{R}|\boldsymbol{Z}\right) q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z}) q_R(\boldsymbol{R}|\boldsymbol{Z}) d\boldsymbol{\alpha} d\boldsymbol{R}.$$

However, the set of equations in (22) do not provide an explicit solution since, for instance, $q_i(\boldsymbol{i}|\boldsymbol{Z})$ depends on expectations computed from $q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z})$ and $q_R(\boldsymbol{R}|\boldsymbol{Z})$. A natural way to compute the distributions which satisfy (22) is thus to iterate between (22a), (22b), and (22c): each factor is updated with the current value of the other factors. More precisely, as will become clear shortly, the procedure amounts to update a few moments of each distribution. To be specific, let us start with the joint posterior distribution of $\boldsymbol{i}$, $\boldsymbol{\alpha}$ and $\boldsymbol{R}$ in (10). Using(22a), we have that (dropping the subscript in the expectation when only one variable is concerned and hence there is no ambiguity.)

$$\begin{aligned}
\ln q_i(\boldsymbol{i}|\boldsymbol{Z}) &= \text{const.} + \sum_{k=1}^{K} i_k \ln p + (1 - i_k) \ln(1-p) \\
&\quad - \left\langle \text{Tr}\left\{\left(\boldsymbol{Z} - \boldsymbol{v}\boldsymbol{\alpha}^H \boldsymbol{D}_{\boldsymbol{i}}^H\right)^H \boldsymbol{R}^{-1} \left(\boldsymbol{Z} - \boldsymbol{v}\boldsymbol{\alpha}^H \boldsymbol{D}_{\boldsymbol{i}}^H\right)\right\}\right\rangle_{\boldsymbol{\alpha},\boldsymbol{R}} \\
&= \text{const.} + \sum_{k=1}^{K} i_k \ln p + (1 - i_k) \ln(1-p) \\
&\quad - \left(\boldsymbol{v}^H \left\langle \boldsymbol{R}^{-1}\right\rangle \boldsymbol{v}\right) \left\langle \boldsymbol{\alpha}^H \boldsymbol{D}_{\boldsymbol{i}}^H \boldsymbol{D}_{\boldsymbol{i}} \boldsymbol{\alpha}\right\rangle \\
&\quad + 2\text{Re}\left(\boldsymbol{v}^H \left\langle \boldsymbol{R}^{-1}\right\rangle \boldsymbol{Z} \boldsymbol{D}_{\boldsymbol{i}} \left\langle \boldsymbol{\alpha}\right\rangle\right) \\
&= \text{const.} + \sum_{k=1}^{K} i_k \ln p + (1 - i_k) \ln(1-p) \\
&\quad - \left(\boldsymbol{v}^H \left\langle \boldsymbol{R}^{-1}\right\rangle \boldsymbol{v}\right) \sum_{k=1}^{K} i_k^2 \left\langle |\alpha_k|^2\right\rangle \\
&\quad + 2\sum_{k=1}^{K} i_k \text{Re}\left(\left\langle \alpha_k\right\rangle \boldsymbol{v}^H \left\langle \boldsymbol{R}^{-1}\right\rangle \boldsymbol{z}_k\right) \\
&= \sum_{k=1}^{K} \ln q_{i_k}(i_k|\boldsymbol{Z}).
\end{aligned} \tag{23}$$

Therefore, it appears that the variables i_k are independent Bernouilli distributed variables with

$$\begin{aligned}
\left\langle i_k\right\rangle &= \Pr\left[i_k = 1|\boldsymbol{Z}\right] \\
&= \frac{p e^{-\left\langle |\alpha_k|^2\right\rangle \left(\boldsymbol{v}^H \left\langle \boldsymbol{R}^{-1}\right\rangle \boldsymbol{v}\right) + 2\text{Re}\left(\left\langle \alpha_k\right\rangle \boldsymbol{v}^H \left\langle \boldsymbol{R}^{-1}\right\rangle \boldsymbol{z}_k\right)}}{(1-p) + p e^{-\left\langle |\alpha_k|^2\right\rangle \left(\boldsymbol{v}^H \left\langle \boldsymbol{R}^{-1}\right\rangle \boldsymbol{v}\right) + 2\text{Re}\left(\left\langle \alpha_k\right\rangle \boldsymbol{v}^H \left\langle \boldsymbol{R}^{-1}\right\rangle \boldsymbol{z}_k\right)}}.
\end{aligned} \tag{24}$$

Note that $q_{i_k}(i_k|\boldsymbol{Z})$ depends on the expected values of $\boldsymbol{R}^{-1}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}\boldsymbol{\alpha}^H$ which are not known but will be updated in the iter-

ative procedure. We next proceed to the derivation of $q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z})$ from (22b)

$$\begin{aligned}\ln q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z}) &= \text{const.} - \sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha} \\ &\quad - \left\langle \left(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}\right)\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\boldsymbol{\alpha}\right\rangle_{\boldsymbol{i},\boldsymbol{R}} \\ &\quad + \left\langle \boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{Z}^H\boldsymbol{R}^{-1}\boldsymbol{v} + \boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{D}_{\boldsymbol{i}}\boldsymbol{\alpha}\right\rangle_{\boldsymbol{i},\boldsymbol{R}} \\ &= \text{const.} - \boldsymbol{\alpha}^H\left[\sigma_\alpha^{-2}\boldsymbol{I} + \left(\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v}\right)\left\langle\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\right\rangle\right]\boldsymbol{\alpha} \\ &\quad + \boldsymbol{\alpha}^H\left\langle\boldsymbol{D}_{\boldsymbol{i}}\right\rangle^H\boldsymbol{Z}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v} + \boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{Z}\left\langle\boldsymbol{D}_{\boldsymbol{i}}\right\rangle\boldsymbol{\alpha}.\end{aligned} \tag{25}$$

The distribution in (25) is recognized as a Gaussian distribution with mean and covariance matrix given by

$$\langle\boldsymbol{\alpha}\rangle = \left[\sigma_\alpha^{-2}\boldsymbol{I} + \left(\boldsymbol{v}^H\langle\boldsymbol{R}^{-1}\rangle\boldsymbol{v}\right)\left\langle\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\right\rangle\right]^{-1}\left\langle\boldsymbol{D}_{\boldsymbol{i}}\right\rangle^H\boldsymbol{Z}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v} \tag{26a}$$

$$\left\langle\boldsymbol{\alpha}\boldsymbol{\alpha}^H - \langle\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}\rangle^H\right\rangle = \left[\sigma_\alpha^{-2}\boldsymbol{I} + \left(\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v}\right)\left\langle\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\right\rangle\right]^{-1}. \tag{26b}$$

Again the parameters of this distribution depend on the expectations of $\boldsymbol{i}$ and $\boldsymbol{R}^{-1}$. Note also, that $q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z}) = \prod_{k=1}^K q_{\alpha_k}(\alpha_k|\boldsymbol{Z})$, i.e., the α_k are independent since the covariance matrix of $\boldsymbol{\alpha}$ is diagonal. In order to complete the analysis, we now derive the distribution of $\boldsymbol{R}$:

$$\begin{aligned}&\ln q_R(\boldsymbol{R}|\boldsymbol{Z}) = \text{const.} - (\nu+K+N)\ln|\boldsymbol{R}| - \text{Tr}\left\{(\nu-N)\bar{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\} \\ &- \left\langle \text{Tr}\left\{\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)^H\boldsymbol{R}^{-1}\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)\right\}\right\rangle_{\boldsymbol{i},\boldsymbol{\alpha}} \\ &= \text{const.} - (\nu+K+N)\ln|\boldsymbol{R}| - \text{Tr}\left\{(\nu-N)\bar{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\} \\ &- \text{Tr}\left\{\boldsymbol{Z}^H\boldsymbol{R}^{-1}\boldsymbol{Z}\right\} + \text{Tr}\left\{\boldsymbol{Z}^H\boldsymbol{R}^{-1}\boldsymbol{v}\langle\boldsymbol{\alpha}\rangle^H\langle\boldsymbol{D}_{\boldsymbol{i}}\rangle^H\right\} \\ &+ \text{Tr}\left\{\langle\boldsymbol{D}_{\boldsymbol{i}}\rangle\langle\boldsymbol{\alpha}\rangle\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{Z}\right\} - \left(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}\right)\text{Tr}\left\{\left\langle\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\right\rangle\left\langle\boldsymbol{\alpha}\boldsymbol{\alpha}^H\right\rangle\right\} \\ &= \text{const.} - (\nu+K+N)\ln|\boldsymbol{R}| - \text{Tr}\left\{(\nu-N)\bar{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\} \\ &- \text{Tr}\left\{\left(\boldsymbol{Z}-\boldsymbol{v}\langle\boldsymbol{\alpha}\rangle^H\langle\boldsymbol{D}_{\boldsymbol{i}}\rangle^H\right)^H\boldsymbol{R}^{-1}\left(\boldsymbol{Z}-\boldsymbol{v}\langle\boldsymbol{\alpha}\rangle^H\langle\boldsymbol{D}_{\boldsymbol{i}}\rangle^H\right)\right\} \\ &- \left(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}\right)\left[\text{Tr}\left\{\left\langle\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\right\rangle\left\langle\boldsymbol{\alpha}\boldsymbol{\alpha}^H\right\rangle\right\} - \langle\boldsymbol{\alpha}\rangle^H\langle\boldsymbol{D}_{\boldsymbol{i}}\rangle^H\langle\boldsymbol{D}_{\boldsymbol{i}}\rangle\langle\boldsymbol{\alpha}\rangle\right].\end{aligned} \tag{27}$$

The above distribution is recognized as an inverse Wishart distribution with $\nu+K$ degrees of freedom. The average value of $\boldsymbol{R}^{-1}$ is thus, provided that $\nu+K>N$,

$$\begin{aligned}\langle\boldsymbol{R}^{-1}\rangle = (\nu+K)\Bigg[&\left(\boldsymbol{Z}-\boldsymbol{v}\langle\boldsymbol{\alpha}\rangle^H\langle\boldsymbol{D}_{\boldsymbol{i}}\rangle^H\right)\left(\boldsymbol{Z}-\boldsymbol{v}\langle\boldsymbol{\alpha}\rangle^H\langle\boldsymbol{D}_{\boldsymbol{i}}\rangle^H\right)^H \\ &+ \left(\sum_{k=1}^K\left\langle i_k^2\right\rangle\left\langle|\alpha_k|^2\right\rangle - \langle i_k\rangle^2\left|\langle\alpha_k\rangle\right|^2\right)\boldsymbol{v}\boldsymbol{v}^H + (\nu-N)\bar{\boldsymbol{R}}\Bigg]^{-1}.\end{aligned} \tag{28}$$

The overall procedure using this variational approach is described in Table II.

TABLE II
VARIATIONAL BAYESIAN METHOD FOR ESTIMATION OF $\boldsymbol{i}$, $\boldsymbol{\alpha}$ AND $\boldsymbol{R}^{-1}$

Input: initial values $\left\langle\boldsymbol{R}^{-1}\right\rangle$, $\langle\boldsymbol{i}\rangle$
1: **for** $n=1,\cdots$ **do**
2: update the mean and covariance matrix of $\boldsymbol{\alpha}$ from (26a)-(26b).
3: update the mean of $\boldsymbol{R}^{-1}$ from (28).
4: update the mean of $\boldsymbol{i}$ from (24).
5: **end for**
Output: VB estimators $\langle\boldsymbol{i}\rangle$, $\langle\boldsymbol{\alpha}\rangle$ and $\left\langle\boldsymbol{R}^{-1}\right\rangle$ of $\boldsymbol{i}$, $\boldsymbol{\alpha}$ and $\boldsymbol{R}^{-1}$.

V. NUMERICAL ILLUSTRATIONS

Performance of the Bayesian estimators is assessed first on synthetic data (application to real radar data will be presented in next section). We consider a radar scenario where $N=16$ pulses, generated according to (1) and (2), are coherently processed. The vector $\boldsymbol{n}_k$ encompasses both thermal noise and clutter components. These two components are assumed to be mutually uncorrelated so that the covariance matrix $\boldsymbol{R}$ is

$$\boldsymbol{R} = \boldsymbol{R}_n + \boldsymbol{R}_c$$

where the subscript identifies the particular component. The thermal noise is assumed to be temporally white with power $\sigma^2 = 0$ dB so that $\boldsymbol{R}_n = \sigma^2\boldsymbol{I}_N$ while the clutter signal is correlated from pulse to pulse. Its spectrum is assumed located around the zero-Doppler with a Gaussian shape such that the (n_1,n_2)th element of the covariance matrix $\boldsymbol{R}_c$ is given by

$$[\boldsymbol{R}_c]_{n_1,n_2} \propto e^{-2\pi^2\sigma_f^2(n_1-n_2)^2} \tag{29}$$

where $\sigma_f^2 = 0.01$ is the spectrum variance. The factor of proportionality in (29) is chosen to obtain a desired value for the clutter-to-noise ratio (CNR) defined as

$$CNR = \frac{\text{Tr}\{\boldsymbol{R}_c\}}{\text{Tr}\{\boldsymbol{R}_n\}}.$$

In all simulations we set $CNR = 20$ dB. We would like to emphasize the fact that *the covariance matrix $\boldsymbol{R}$ is not drawn from its prior distribution* in (3). K range gates are used to estimate the covariance matrix $\boldsymbol{R}$ while K_{cont} of them are contaminated in the direction under test $\boldsymbol{v} = \left[1 \quad e^{j2\pi f_{D-\text{cont}}} \quad \cdots \quad e^{j2\pi(N-1)f_{D-\text{cont}}}\right]^T$ where $f_{D-\text{cont}}$ is the Doppler frequency of the contamination. For each contaminated cell $k=1,\ldots,K_{\text{cont}}$, the contamination level is monitored by the signal-to-noise-ratio defined as

$$SNR = \frac{\sigma_\alpha^2\left(\boldsymbol{v}^H\boldsymbol{v}\right)}{\text{Tr}\{\boldsymbol{R}_n\}}.$$

Unless otherwise stated, $SNR = 15$ dB in the simulations.

As stated earlier in Section II, the prior matrix $\bar{\boldsymbol{R}}$ is chosen so as to regularize the covariance matrix estimation problem, i.e., $\bar{\boldsymbol{R}} = \sigma_{DL}^2\boldsymbol{I}$. The numerical value of the factor σ_{DL}^2 is chosen as for conventional diagonal loading [23], i.e., $\sigma_{DL}^2 \approx 5\sigma^2$. Note that this choice may not be optimal in the present case but, since the optimal loading level is scenario dependent and hard to find, we will resort to this usual rule of thumb. On the other hand, to choose the degree of a priori ν, observe from (17) that the mean of $\boldsymbol{R}|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{Z}$ when $\bar{\boldsymbol{R}} = \sigma_{DL}^2\boldsymbol{I}$ is

$$\begin{aligned}\mathsf{E}\{\boldsymbol{R}|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{Z}\} = \frac{1}{\nu+K-N}\Big[&(\nu-N)\sigma_{DL}^2\boldsymbol{I} \\ &+ \left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)^H\Big].\end{aligned} \tag{30}$$

The degree of a priori is then fixed to $\nu = K+N$ to ensure that the weight given respectively to the observations and to $\bar{\boldsymbol{R}}$

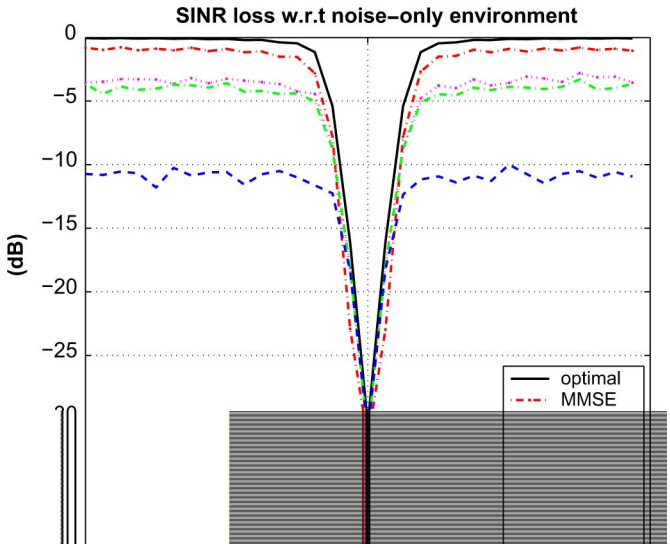


Fig. 1. SINR-loss versus Doppler frequency of the contaminated cells. $N = 16$, $K = N + 1$, $K_{\text{cont}} = 4$, $SNR = 15$ dB.

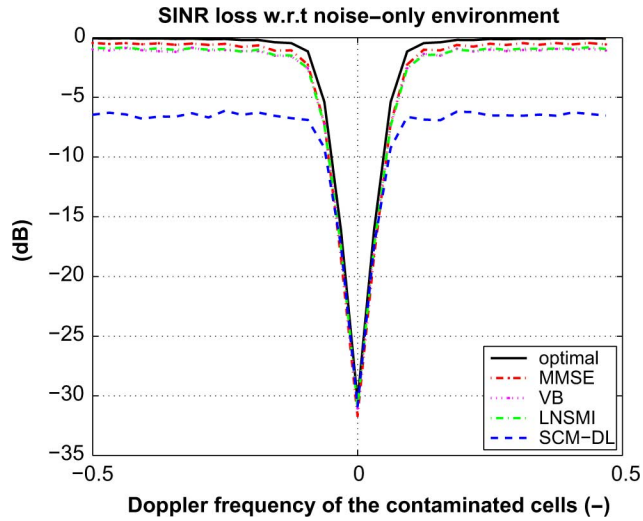


Fig. 2. SINR-loss versus Doppler frequency of the contaminated cells. $N = 16$, $K = 2N$, $K_{\text{cont}} = 4$, $SNR = 15$ dB.

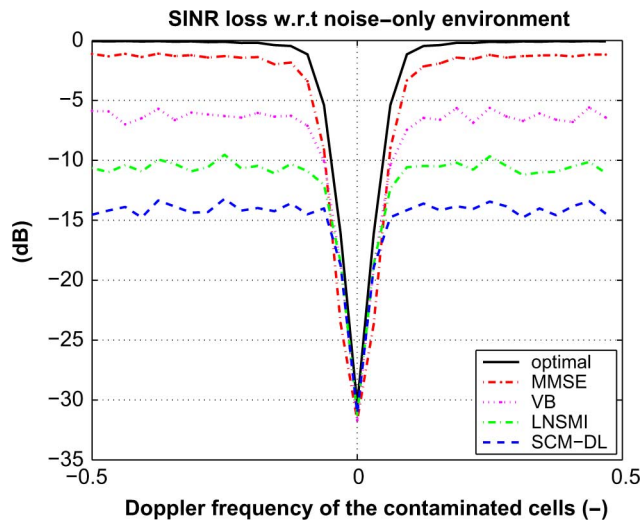


Fig. 3. SINR-loss versus Doppler frequency of the contaminated cells. $N = 16$, $K = N + 1$, $K_{\text{cont}} = 8$, $SNR = 15$ dB.

in (30) is tantamount to a conventional diagonal loading. Intensive numerical simulations have demonstrated that many iterations are necessary to obtain an appropriate MMSE-estimation ($N_{\text{bi}} = 500$ and $N_r = 5000$) whereas only a few iterations ($N_{VB} = 50$) are required for the VB algorithm to converge.

The performance of the beamformers based on the Bayesian estimators (20) and(28) is assessed through the SINR-loss with respect to the noise-only-environment defined by [24]

$$\mathcal{L} = \frac{\left|\boldsymbol{w}^H \boldsymbol{v}\right|^2}{\boldsymbol{w}^H \boldsymbol{R} \boldsymbol{w}} \frac{1}{\boldsymbol{v}^H \boldsymbol{R}_n^{-1} \boldsymbol{v}}$$

where $\boldsymbol{w} \propto \hat{\boldsymbol{R}}^{-1} \boldsymbol{v}$ and $\hat{\boldsymbol{R}}$ denotes one of the estimates derived above. In addition to the MMSE- and VB-based beamformers, performance are also shown for the optimal beamformer ($\boldsymbol{w}_{\text{opt}} \propto \boldsymbol{R}^{-1} \boldsymbol{v}$), the conventional DL beamformer [23], [25] and the LNSMI beamformer [14]. For the latter the number of iterations used was $N_{LNSMI} = 15$ and the loading level was chosen as in [14], i.e., 10^{-3} times the largest eigenvalue of the sample covariance matrix of the normalized snapshots. Observe that, unlike conventional methods, the Bayesian algorithms of Table I and Table II also bring important information about the level of contamination for each range cell of the training interval.

We first consider an ideal situation where both the assumed level of contamination $p = K_{\text{cont}}/K$ and the target power σ_α^2 match perfectly the values used to generate the data. Fig. 1 displays the result in the case $K = N + 1$, $K_{\text{cont}} = 4$ and $SNR = 15$ dB. The influence of the number of secondary data K is assessed in Fig. 2 where K is increased to $K = 2N$ while the number of contaminated cells is kept to $K_{\text{cont}} = 4$. In contrast, we study in Fig. 3 the influence of K_{cont} which is increased to $K_{\text{cont}} = 8$ while $K = N + 1$. Finally, Fig. 4 investigates the influence of the power of the signal in the contaminated cells. Inspection of these 4 figures enables one to draw the following prominent properties of the various beamformers:

- The MMSE beamformer clearly outperforms the VB as well as the LNSMI, especially in low number of samples or high number of contaminated cells. For example, the improvement compared to VB is, in the thermal noise domain, about $2-3$ dB for $K = N+1$, $K_{\text{cont}} = 4$ and about 5 dB for $K = N+1$, $K_{\text{cont}} = 8$. Only in the case $K = 2N$ (Fig. 2) do the VB and LNSMI approximately perform as well as the MMSE: in this case the three methods are really close to optimal.
- The MMSE beamformer is nearly insensitive to variations in the number of contaminated cells in terms of SINR loss: it incurs a 0.5 dB loss from $K_{\text{cont}} = 4$ in Fig. 1 to $K_{\text{cont}} = 8$ in Fig. 3. Additionally, a variation of the signal power in the contaminated cells induces negligible differences for the SINR loss.
- VB and LNSMI yield close SINR losses at least when $K_{\text{cont}} = 4$. When the number of contaminated cells increases (see Fig. 3), VB is seen to incur a less severe loss, about 2.5 dB for VB against 6 dB for LNSMI. In contrast, VB and LNSMI behave similarly when the signal power in the contaminated cells increases: however, they are less

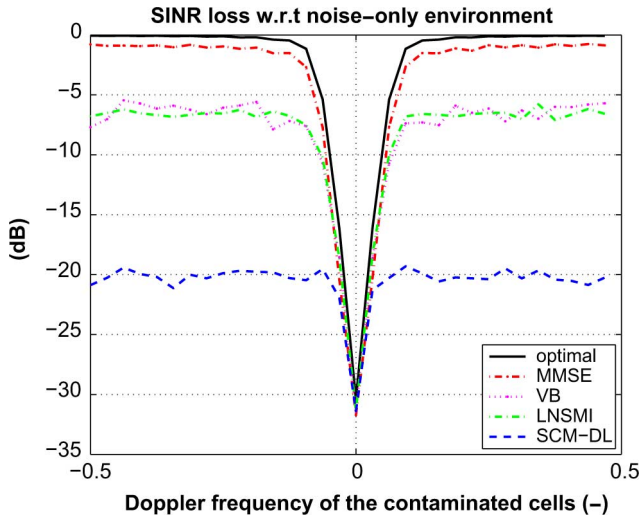


Fig. 4. SINR-loss versus Doppler frequency of the contaminated cells. $N = 16$, $K = N + 1$, $K_{\text{cont}} = 4$, $SNR = 25$ dB.

robust than MMSE as a $2 - 3$ dB loss is observed when SNR goes from 15 dB to 25 dB.

- All methods clearly outperform diagonal loading: the latter undergoes severe performance degradation when the number or the signal level of contaminated cells increases. This poor performance is due to the nature of the data itself, i.e., the presence of outliers, and not to the choice of the loading level. Indeed, through many simulations (not reported in this paper) with different loading levels we experimented that diagonal loading is not effective in the clutter region whatever the diagonal loading level. Hence, looking for an optimal loading level for the particular model considered herein is questionable and, moreover, a delicate issue.

Our second series of simulations deals with **robustness analysis** of the beamformers. In the previous simulations, performance of our Bayesian estimators has been studied when the processing parameters have exactly the same values as those used for signal generation. We now introduce mismatches between the parameters used for generation and processing. The analysis will show that, to a certain extent, our estimators are quite robust so that a non-perfect knowledge of the scenario parameters does not appear to be critical for a practical implementation.

Figs. 5–6 study the influence of a mismatch in the assumed contamination level $p = K_{\text{cont}}/K$ and in the assumed target power σ_α^2, respectively. In these figures, $K = N+1$, $K_{\text{cont}} = 4$, $SNR = 15$ dB and we plot the SINR-loss resulting from this error with respect to the SINR-loss obtained without mismatch. The following observations can be made:

- Clearly the MMSE beamformer is almost not affected by a wrong assumption on p or on σ_α^2.
- The VB appears to be more sensitive to both parameters. Interestingly enough, overestimating the number of contaminated cells results in an improved SINR of the VB-estimator, which then comes close to the MMSE estimator based on Gibbs sampling.

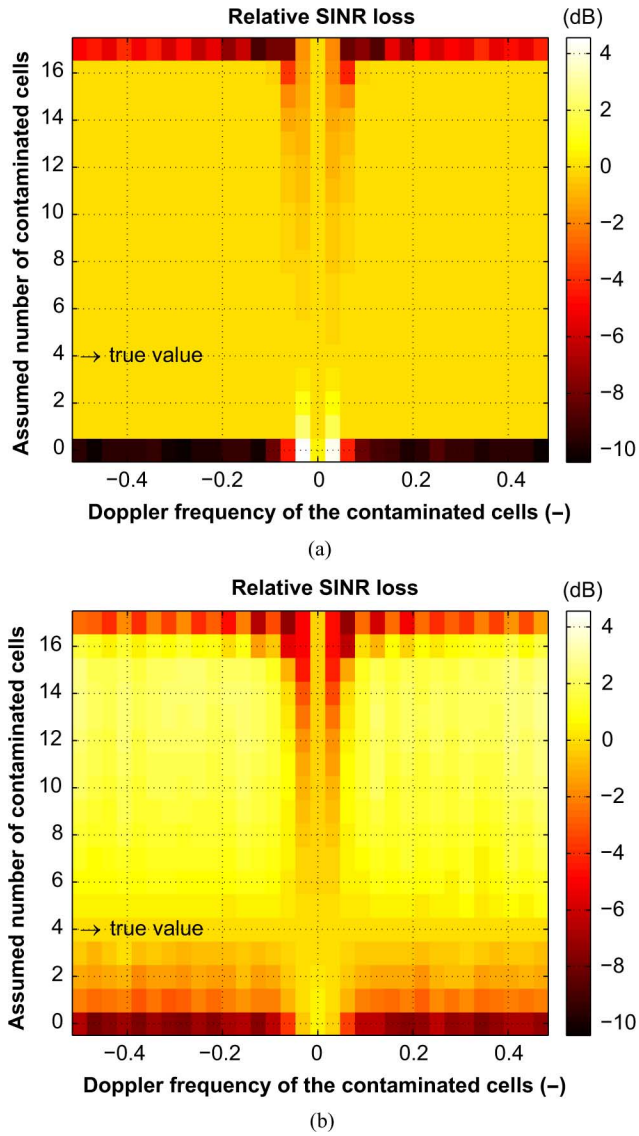
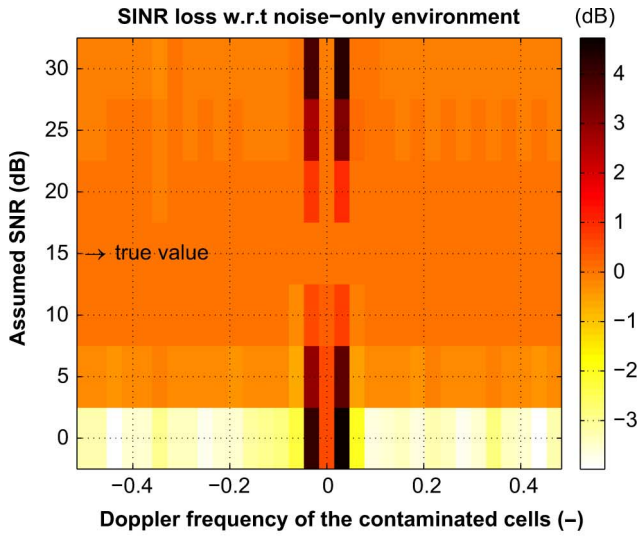
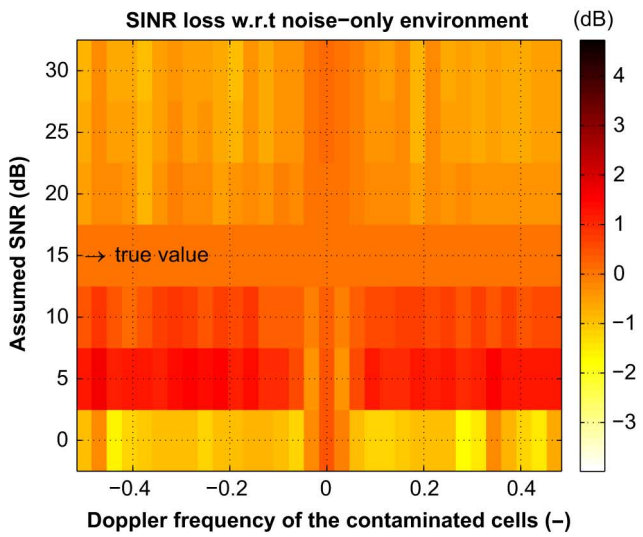


Fig. 5. Robustness towards a wrong assumption about the density of contamination p. SINR-loss with respect to the Doppler frequency of the contaminated cells and the assumed number of contaminated cells. (a) MMSE estimation. (b) VB estimation.

Finally, to complete our robustness analysis, the effect of mismatch between the true direction of contamination and the direction under test is investigated. To do so, data are generated with a single direction of contamination $f_{D-\text{cont}}$ and all the directions f_D are tested. The SINR-loss so obtained is depicted in Fig. 7(a) when the contamination is near the clutter edge for $f_{D-\text{cont}} = 0.09$ and in Fig. 7(b) when the contamination occurs in the thermal noise domain for $f_{D-\text{cont}} = 0.25$. For each beamformer presented, performance is affected only locally around the direction of contamination. The MMSE-beamformer endures only small losses in this region. Losses are more pronounced for both the VB- and LNSMI beamformer as they have naturally a lower SINR-loss than that of the MMSE in the true direction of contamination. Finally, the SCM-DL beamformer undergoes important losses in this region which may create a blind zone especially for a contamination near the clutter edge.

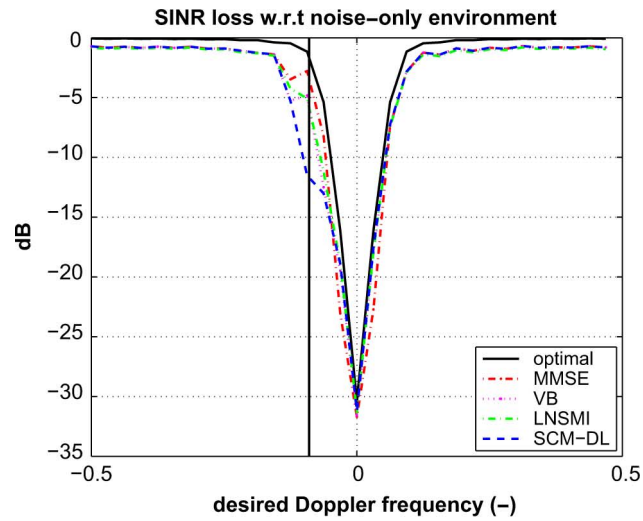


(a)

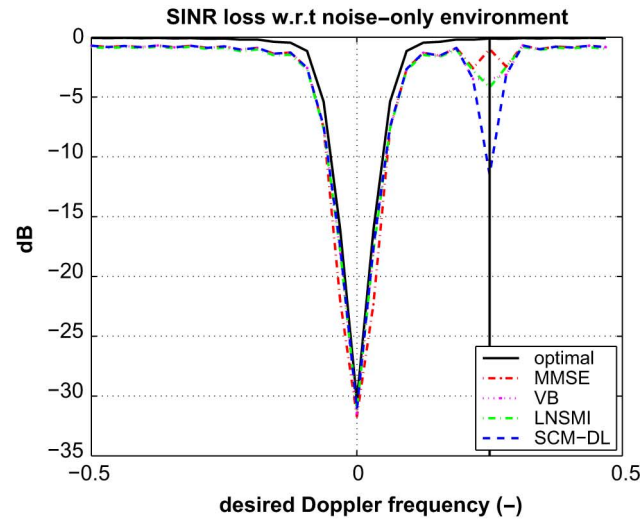


(b)

Fig. 6. Robustness towards a wrong assumption about the contaminating signal power. SINR-loss with respect to the Doppler frequency of the contaminated cells and the assumed contaminating signal power. (a) MMSE estimation. (b) VB estimation.



(a)



(b)

Fig. 7. Robustness to steering vector mismatch. SINR-loss with respect to the desired Doppler frequency. (a) Doppler frequency of contaminated cells $f_{D-\text{cont}} = -0.09$. (b) Doppler frequency of contaminated cells $f_{D-\text{cont}} = 0.25$.

VI. Experimental PARSAX Radar Data

A. Experimental Setup

Performance of the Bayesian methods are now assessed on experimental data collected at the Delft University of Technology on November 2010. Indeed, the International Research Centre for Telecommunications and Radar (IRCTR) has been developing these last years the polarimetric agile radar in S- and X-band (PARSAX) [27]. The system is situated on the rooftop of a 100m-high building and entails two reflector antennas, one to to transmit and one to receive. These two antennas can be considered as colocated. The radar system has a flexible architecture with respect to the generated waveform and the pre-processing algorithms. For the experiment, a frequency modulated continuous waveform has been chosen as well as a deramping technique for range compression. To assess the performance of our detectors, synthetic targets have been injected in a scene consisting mostly of a background noise and ground clutter. Location and SNR[1] of these targets are depicted on Fig. 6; they have been chosen to reproduce a possible slow and heavy traffic. Note that given the range resolution, i.e., $\delta_R = c/(2B) = 1.5$ m, a target may be extended in range. Other parameters describing the scenario are given in Table III. Note that we had at our disposal a data set containing only thermal noise (the antenna was pointed in the upward direction during a sunny day) so that the data are scaled to set the noise floor to 0 dB.

To give more insight into the scenario, the estimated power spectrum of the data to be processed is depicted in Fig. 8 before and after injecting the targets. As can be already observed, there is likely to be a true target around the range bin 855.

[1]To interpret more conveniently the next results, target amplitudes have not been drawn randomly here but have been taken deterministically as $\alpha_k^2 = \sigma^2 SNR_k$.

TABLE III
PARSAX SCENARIO PARAMETERS

carrier frequency	$f_0 = 3.315$ GHz
bandwidth	$B = 100$ MHz
pulse repetition interval	$T_r = 1$ ms
# pulse	$N = 32$
# range gate	$K = N + 2$
thermal noise power	$\sigma^2 = 0$ dB

Given our robustness analysis we choose a high probability of contamination equal to $p = 0.5$ and a contamination power equal to $\sigma_\alpha^2 = 15$ dB which is a compromise between the different target powers. As the true covariance matrix $\boldsymbol{R}$ is unknown, the SINR-loss can only be estimated, thus we prefer study hereafter the detection map obtained via the ACE [28] test statistic, i.e.,

$$t_{\text{ACE}} = \frac{\left|\boldsymbol{v}^H \hat{\boldsymbol{R}}^{-1} \boldsymbol{z}\right|^2}{\left(\boldsymbol{v}^H \hat{\boldsymbol{R}}^{-1} \boldsymbol{v}\right)\left(\boldsymbol{z}^H \hat{\boldsymbol{R}}^{-1} \boldsymbol{z}\right)} \tag{31}$$

where $\boldsymbol{z}$ is the cell under test. For each technique, the covariance matrix $\hat{\boldsymbol{R}}$ is estimated via a symmetric K range-gate interval around the CUT with two guard cells. The ACE test statistic represents the angle between the target under test and the primary data $\boldsymbol{z}$ in the quasi-whitened space. It seems here to be a pertinent detector as it emerges as the solution to different detection problems [28]–[30], especially in non-homogeneous environment.

B. Results

Detection maps obtained with the ACE detector (31) are depicted in Fig. 9. The following comments can be made in accordance.

- Each target scatterer can be easily identified with the MMSE-ACE test statistic; even when the scatterer is near the clutter edge (targets 6 and 8) or when its power is low (targets 5 and 6). A few sidelobes are seen especially for target 8 near the clutter edge. Also, one can observe a slight clutter undernulling for a few range gates (e.g., around range bin 834 and 866).
- Same remarks can be made for the VB-ACE test statistic though a few more sidelobes can be seen (e.g., at range bins 831 and 833) and the scatterer peaks are somehow less pronounced. However, one has to keep in mind that the computational load of the VB algorithm is dramatically less than that of the MMSE-algorithm.
- The LSNMI-ACE misses some scatterers (e.g., within targets 2 and 3 and target 5). However, these misses are compensated in a way by the absence of sidelobes (except for target 8 near the clutter edge) and a lower floor in the thermal noise domain. Note that the LNSMI-ACE is inclined to a great clutter undernulling.
- The SCM-DL-ACE test statistic has the poorest performance which is in accordance with the prior study based on synthetic data. A lot of scatterers are missed (e.g., within targets 2, 3, 7 and the two low power targets 5 and 6).

Before concluding our study, a last remark can be made about the possible targets that might be naturally in the data. Indeed,

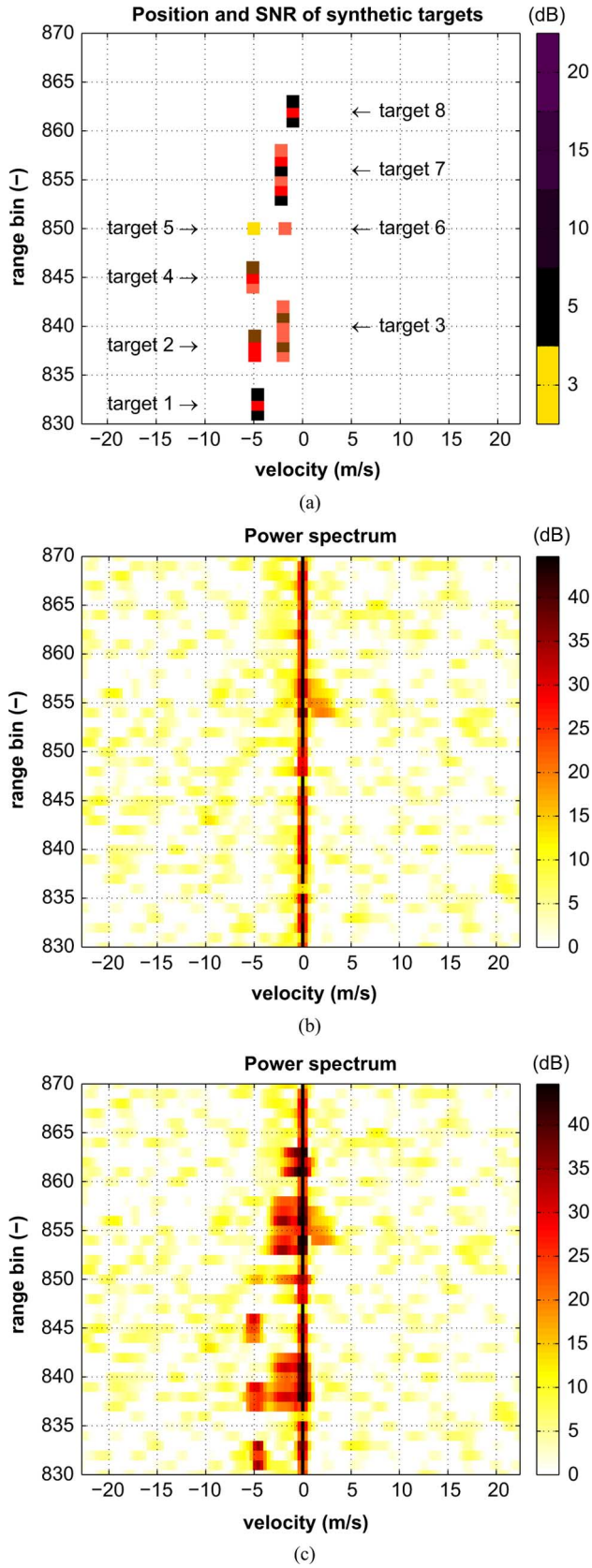


Fig. 8. Range-velocity maps describing the data. (a) Location and SNR of the injected synthetic targets. Power spectrum via a 1D-APES filter [26] (b) prior to target injection (c) after target injection.

the previously mentioned true target at range bin 855 has been clearly identified by both the MMSE-, VB- and SCM-DL-ACE

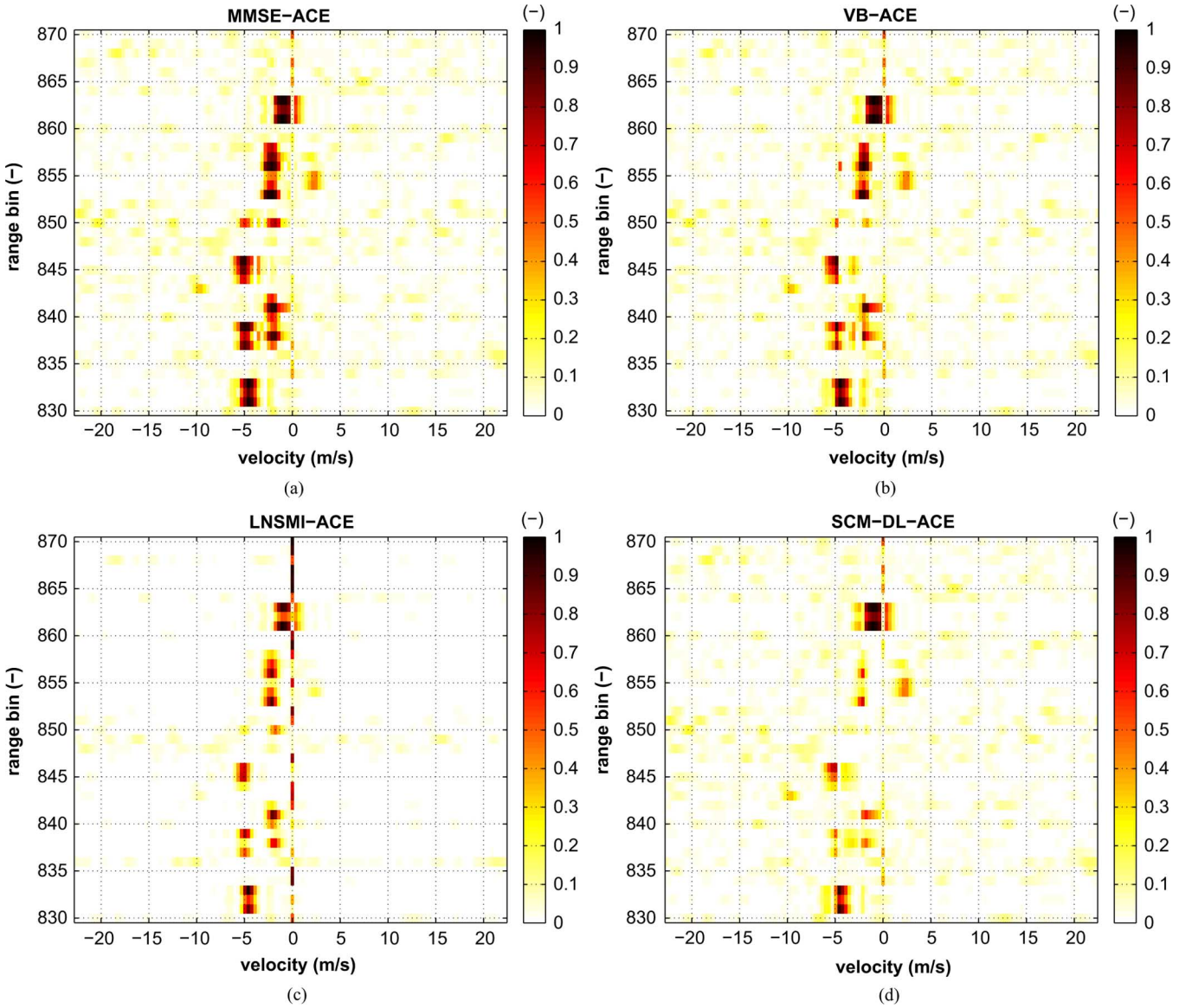


Fig. 9. Range-velocity map. ACE-like detectors.

detectors. Also, the same detectors identify a target at range 843 which could not clearly be seen on the power spectrum map of Fig. 8. Note that the LNSMI-ACE detector misses both of them.

VII. Conclusions

In this paper, we considered the adaptive filtering problem in the case where some training samples are possibly contaminated by signal-like components. This practically relevant situation, which is very detrimental to conventional adaptive filters, was addressed using a Bayesian approach where the variables indicating the presence of contaminated cells were assumed to follow a Bernoulli-Gaussian distribution. A non-informative prior was assumed for the noise covariance matrix, which helped regularize the covariance matrix estimation problem. Within this framework, the MMSE estimator was derived and implemented using a Gibbs sampler. A computationally simpler scheme, based on variational Bayesian (VB) analysis was also proposed. The MMSE estimator was shown to come very close to the optimal adaptive filter, whatever the proportion of contaminated samples or the power of the signal component in the training samples. Moreover, its robustness to non-perfect knowledge of these parameters was evidenced. The VB results in slightly increased SNR losses compared to the MMSE, yet at a much lower computational cost. Its implementation, as well as its performances, are comparable to those of the LNSMI. These algorithms were assessed against real radar data: in particular, the MMSE showed its ability to detect small targets close to the clutter ridge.

Appendix
Extension to Random Hyper-Parameters

In the main body of this paper, we assumed that p and σ_α^2 were known quantities and we carried out a sensitivity analysis to assess the robustness of the various estimators to a non precise knowledge of these variables. An alternative option is to treat them as random variables with non-informative priors. In this appendix, we briefly indicate how the Gibbs sampling scheme and the variational Bayesian method can be extended to

the case of random hyper-parameters. More precisely, we now assume that p is a random variable with uniform distribution on $[0,1]$. Doing so we do not make any hypothesis about the probability of signal contamination and we estimate it jointly with the other parameters. Regarding σ_α^2 we choose a conjugate inverse-Gamma prior distribution, denoted as $\sigma_\alpha^2 \sim \mathsf{IG}(a,b)$, whose expression is

$$\pi(\sigma_\alpha^2) \propto \left(\sigma_\alpha^2\right)^{-(a+1)} e^{-b\sigma_\alpha^{-2}}. \tag{32}$$

Depending on the choice of a and b this prior can be made rather non informative. We now derive all necessary distributions to extend the estimators of the previous sections. For the sake of conciseness, we just provide the main steps since derivations are similar to those in the past sections. To begin with, the joint posterior distribution of all variables is now given by

$$\begin{aligned}
&p\left(\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R},p,\sigma_\alpha^2|\boldsymbol{Z}\right)\\
&\propto p\left(\boldsymbol{Z}|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R}\right)\pi\left(\boldsymbol{i}|p\right)\pi(p)\pi\left(\boldsymbol{\alpha}|\sigma_\alpha^2\right)\pi\left(\sigma_\alpha^2\right)\pi\left(\boldsymbol{R}\right)\\
&\propto |\boldsymbol{R}|^{-K}\mathrm{etr}\left\{-\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)^H\boldsymbol{R}^{-1}\left(\boldsymbol{Z}-\boldsymbol{v}\boldsymbol{\alpha}^H\boldsymbol{D}_{\boldsymbol{i}}^H\right)\right\}\\
&\times\left(\prod_{k=1}^K p^{i_k}(1-p)^{1-i_k}\right)\left(\sigma_\alpha^2\right)^{-(a+K+1)}e^{-\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}}e^{-b\sigma_\alpha^{-2}}\\
&\times|\boldsymbol{R}|^{-(\nu+N)}\mathrm{etr}\left\{-(\nu-N)\bar{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\}.
\end{aligned}\tag{33}$$

A. Gibbs Sampler

It is straightforward to show that $p\left(\boldsymbol{i}|\boldsymbol{\alpha},\boldsymbol{R},p,\sigma_\alpha^2,\boldsymbol{Z}\right)$ is still given by (11) and hence $\boldsymbol{i}$, conditioned on $\boldsymbol{\alpha}$, $\boldsymbol{R}$, p, σ_α^2 and $\boldsymbol{Z}$, is still Bernoulli distributed with the probability that it equals one still given by (13). Accordingly, the posterior distributions $p\left(\boldsymbol{\alpha}|\boldsymbol{i},\boldsymbol{R},p,\sigma_\alpha^2,\boldsymbol{Z}\right)$ and $p\left(\boldsymbol{R}|\boldsymbol{i},\boldsymbol{\alpha},p,\sigma_\alpha^2,\boldsymbol{Z}\right)$ are the same as in (14) and(17). The difference lies in the posterior distributions of p and σ_α^2. As for p, we have from (33) that

$$\begin{aligned}
p\left(p|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R},\sigma_\alpha^2,\boldsymbol{Z}\right) &\propto \prod_{k=1}^K p^{i_k}(1-p)^{1-i_k}\\
&= p^{\sum_{k=1}^K i_k}(1-p)^{\sum_{k=1}^K(1-i_k)}
\end{aligned}\tag{34}$$

which is a Beta distribution, i.e.,

$$p|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R},\sigma_\alpha^2,\boldsymbol{Z} \sim \mathsf{Beta}\left(1+\sum_{k=1}^K i_k, 1+\sum_{k=1}^K(1-i_k)\right). \tag{35}$$

Finally, the conditional posterior of σ_α^2 can be written as

$$p\left(\sigma_\alpha^2|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R},p,\boldsymbol{Z}\right) \propto \left(\sigma_\alpha^2\right)^{-(a+K+1)} e^{-\sigma_\alpha^{-2}\left[b+\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right]} \tag{36}$$

and hence $\sigma_\alpha^2|\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R},p,\boldsymbol{Z} \sim \mathsf{IG}\left(a+K, b+\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right)$. The Gibbs sampling scheme of Table I needs to be modified in order to account for the new variables which need to be generated.

B. Variational Bayesian Method

We now seek an approximation of the form

$$\begin{aligned}
p\left(\boldsymbol{i},\boldsymbol{\alpha},\boldsymbol{R},p,\sigma_\alpha^2|\boldsymbol{Z}\right) \simeq\ & q_i(\boldsymbol{i}|\boldsymbol{Z})q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z})q_R(\boldsymbol{R}|\boldsymbol{Z})q_p(p|\boldsymbol{Z})\\
& q_{\sigma_\alpha^2}(\sigma_\alpha^2|\boldsymbol{Z}).
\end{aligned}\tag{37}$$

Using (33) along with derivations that led to(23), one can easily show that

$$\begin{aligned}
\ln q_i(\boldsymbol{i}|\boldsymbol{Z}) = \text{const.} &+ \sum_{k=1}^K i_k\left\langle\ln p\right\rangle + (1-i_k)\left\langle\ln(1-p)\right\rangle\\
&- \left(\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v}\right)\sum_{k=1}^K i_k^2\left\langle|\alpha_k|^2\right\rangle\\
&+ 2\sum_{k=1}^K i_k\mathrm{Re}\left(\left\langle\alpha_k\right\rangle\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{z}_k\right)
\end{aligned}\tag{38}$$

which implies that the variables i_k are still independent Bernoulli distributed variables with mean value

$$\begin{aligned}
&\left\langle i_k\right\rangle\\
&= \frac{e^{\left\langle\ln p\right\rangle}e^{-\left\langle|\alpha_k|^2\right\rangle\left(\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v}\right)+2\mathrm{Re}\left(\left\langle\alpha_k\right\rangle\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{z}_k\right)}}{e^{\left\langle\ln(1-p)\right\rangle}+e^{\left\langle\ln p\right\rangle}e^{-\left\langle|\alpha_k|^2\right\rangle\left(\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v}\right)+2\mathrm{Re}\left(\left\langle\alpha_k\right\rangle\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{z}_k\right)}}.
\end{aligned}\tag{39}$$

The derivation of $q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z})$ leads to

$$\begin{aligned}
&\ln q_\alpha(\boldsymbol{\alpha}|\boldsymbol{Z})\\
&= \text{const.} - \boldsymbol{\alpha}^H\left[\left\langle\sigma_\alpha^{-2}\right\rangle\boldsymbol{I} + \left(\boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v}\right)\left\langle\boldsymbol{D}_{\boldsymbol{i}}^H\boldsymbol{D}_{\boldsymbol{i}}\right\rangle\right]\boldsymbol{\alpha}\\
&\quad + \boldsymbol{\alpha}^H\left\langle\boldsymbol{D}_{\boldsymbol{i}}\right\rangle^H\boldsymbol{Z}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{v} + \boldsymbol{v}^H\left\langle\boldsymbol{R}^{-1}\right\rangle\boldsymbol{Z}\left\langle\boldsymbol{D}_{\boldsymbol{i}}\right\rangle\boldsymbol{\alpha}.
\end{aligned}\tag{40}$$

Hence $\boldsymbol{\alpha}|\boldsymbol{Z}$ is Gaussian distributed with mean and covariance matrix given by (26a)–(26b) except that $\left\langle\sigma_\alpha^{-2}\right\rangle$ should be substituted for σ_α^{-2}. Regarding $\boldsymbol{R}$ it turns out that there is no modification compared to (27), and that $\boldsymbol{R}$ has the same inverse Wishart distribution. Let us now consider p:

$$\begin{aligned}
\ln q_p(p|\boldsymbol{Z}) &= \text{const.} + \left\langle\sum_{k=1}^K i_k\ln p + (1-i_k)\ln(1-p)\right\rangle_{\boldsymbol{i}}\\
&= \text{const.} + \sum_{k=1}^K\left\langle i_k\right\rangle\ln p + \left\langle 1-i_k\right\rangle\ln(1-p).
\end{aligned}\tag{41}$$

Therefore, $p|\boldsymbol{Z} \sim \mathsf{Beta}\left(1+\sum_{k=1}^K\left\langle i_k\right\rangle, 1+\sum_{k=1}^K\left\langle 1-i_k\right\rangle\right)$. As evidenced by (39), one needs to obtain the mean of $\ln p$ and $\ln(1-p)$ which are given by [31, 4.253]

$$\left\langle\ln p\right\rangle = \psi\left(1+\sum_{k=1}^K\left\langle i_k\right\rangle\right) - \psi(K+2) \tag{42a}$$

$$\left\langle\ln(1-p)\right\rangle = \psi\left(1+\sum_{k=1}^K\left\langle 1-i_k\right\rangle\right) - \psi(K+2) \tag{42b}$$

where $\psi(.)$ is Euler's psi function. Let us finally consider $q_{\sigma_\alpha^2}(\sigma_\alpha^2|\boldsymbol{Z})$:

$$\begin{aligned}
\ln q_{\sigma_\alpha^2}(\sigma_\alpha^2|\boldsymbol{Z}) = \text{const.} &- (a+K+1)\ln\sigma_\alpha^2\\
&-\sigma_\alpha^{-2}\left[b+\left\langle\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right\rangle\right]
\end{aligned}\tag{43}$$

which implies that $\sigma_\alpha^2|\boldsymbol{Z} \sim \mathsf{IG}\left(a+K, b+\left\langle \boldsymbol{\alpha}^H\boldsymbol{\alpha}\right\rangle\right)$. The mean value of σ_α^{-2} is thus

$$\left\langle \sigma_\alpha^{-2}\right\rangle = \frac{a+K}{b+\left\langle \boldsymbol{\alpha}^H\boldsymbol{\alpha}\right\rangle}. \tag{44}$$

Again, the algorithm of Table II should be modified to update accordingly all new variables.

ACKNOWLEDGMENT

The authors would like to thank the IRCTR at TU-Delft for kindly providing the PARSAX experimental data.

REFERENCES

[1] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Anal*. Reading, MA, USA: Addison-Wesley, 1991.

[2] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace Electron. Syst.*, vol. 10, no. 6, pp. 853–863, Nov. 1974.

[3] W. L. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 621–633, Apr. 2000.

[4] K. Gerlach, "The effects of signal contamination on two adaptive detectors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 297–309, Jan. 1995.

[5] D. M. Boroson, "Sample size considerations for adaptive arrays," *IEEE Trans. Aerospace Electron. Syst.*, vol. 16, no. 4, pp. 446–451, Jul. 1980.

[6] H. L. Van Trees, *Optimum Array Processing*. New York, NY, USA: Wiley, 2002.

[7] K. Gerlach, "Outlier resistant adaptive matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 885–901, Jul. 2002.

[8] D. J. Rabideau and A. O. Steinhardt, "Improved adaptive clutter cancellation through data-adaptive training," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 3, pp. 879–891, Jul. 1999.

[9] P. Chen, W. L. Melvin, and M. C. Wicks, "Screening among multivariate normal data," *J. Multivar. Anal.*, vol. 69, no. 1, pp. 10–29, Apr. 1999.

[10] K. Gerlach, S. D. Blunt, and M. L. Picciolo, "Robust adaptive matched filtering using the Fracta algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 929–945, Jul. 2004.

[11] M. Rangaswamy, "Statistical Anal. of the nonhomogeneity detector for non-Gaussian interference backgrounds," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2101–2111, Jun. 2005.

[12] S. D. Blunt and K. Gerlach, "Efficient robust AMF using the Fracta algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, p. 537548, Apr. 2005.

[13] F. Lin, M. Rangaswamy, C. Wolfe, J. Chaves, and A. Krishnamurthy, "Three variants of an outlier removal algorithm for radar stap," in *Proc. IEEE 4th SAM Workshop*, Waltham, MA, Jul. 12–14, 2006, pp. 621–625.

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

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

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

[17] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York, NY, USA: Springer Verlag, 2004.

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

[19] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York, NY, USA: Wiley, 1982.

[20] V.Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*. New York, NY, USA: Springer-Verlag, 2006.

[21] H. Attias, "A variational Bayesian framework for graphical models," in *Advances in Neural Inform. Processing Syst*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds. Cambridge, MA, USA: MIT Press, 2000, pp. 209–215.

[22] M. J. Beal, "Variational algorithm for approximate Bayesian inference," Ph.D. dissertation, Univ. College London, London, U.K., 2003.

[23] Y. I. Abramovich and A. I. Nevrev, "An Anal. of effectiveness of adaptive maximization of the signal to noise ratio which utilizes the inversion of the estimated covariance matrix," *Radio Eng. Electron. Phys.*, vol. 26, pp. 67–74, Dec. 1981.

[24] J. Ward, "Space-time adaptive processing for airborne radar,". Lexington, MA, USA, Lincoln Lab., MIT, Tech. Rep. 1015, Dec. 1994.

[25] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, Jul. 1988.

[26] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1469–1484, Jun. 1996.

[27] O. A. Krasnov, G. P. Babur, Z. Wang, L. P. Ligthart, and F. van der Zwan, "Basics and first experiments demonstrating isolation improvements in the agile polarimetric FM-CW radar—PARSAX," *Int. J. Microw. Wireless Technol. (Special Issue)*, vol. 2, no. 3–4, pp. 419–428, Aug. 2010.

[28] L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 1996, pp. 1114–1117.

[29] O. Besson, "Detection in the presence of surprise or undernulled interference," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 352–354, May 2007.

[30] S. Bidon, O. Besson, and J.-Y. Tourneret, "The adaptive coherence estimator is the generalized likelihood ratio test for a class of heterogeneous environments," *IEEE Signal Process. Lett.*, vol. 15, pp. 281–284, 2008.

[31] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, A. Jeffrey, Ed., 5 ed. New York, NY, USA: Academic, 1994.

Olivier Besson (SM'04) received the M.Sc. and Ph.D. degrees in signal processing in 1988 and 1992, respectively, from the Institut National Polytechnique, Toulouse, France.

He is currently a Professor with the Department of Electronics, Optronics and Signal of ISAE (Institut Supérieur de l'Aéronautique et de l'Espace), Toulouse. His research interests are in the area of robust adaptive array processing, mainly for radar applications. Dr. Besson is a former Associate Editor of the IEEE Transactions Signal Processing and the IEEE Signal Processing Letters. He is a member of the Sensor Array and Multichannel technical committee (SAM TC) of the IEEE Signal Processing Society.

Stéphanie Bidon (M'08) received the engineer and master degrees from ENSICA, Toulouse, in 2004 and 2005 respectively and the Ph.D. degree from INP, Toulouse, in 2008.

She is now with the Department of Electronics, Optronics and Signal of ISAE (Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse) as an assistant professor. Her research interests include digital signal processing particularly with application to airborne radar.