A Bayesian Approach to Adaptive Detection in Nonhomogeneous Environments

Stéphanie Bidon, *Student Member, IEEE*, Olivier Besson, *Senior Member, IEEE*, and Jean-Yves Tourneret, *Member, IEEE*

***Abstract*—We consider the adaptive detection of a signal of interest embedded in colored noise, when the environment is nonhomogeneous, i.e., when the training samples used for adaptation do not share the same covariance matrix as the vector under test. A Bayesian framework is proposed where the covariance matrices of the primary and the secondary data are assumed to be random, with some appropriate joint distribution. The prior distributions of these matrices require a rough knowledge about the environment. This provides a flexible, yet simple, knowledge-aided model where the degree of nonhomogeneity can be tuned through some scalar variables. Within this framework, an approximate generalized likelihood ratio test is formulated. Accordingly, two Bayesian versions of the adaptive matched filter are presented, where the conventional maximum likelihood estimate of the primary data covariance matrix is replaced either by its minimum mean-square error estimate or by its maximum *a posteriori* estimate. Two detectors require generating samples distributed according to the joint posterior distribution of primary and secondary data covariance matrices. This is achieved through the use of a Gibbs sampling strategy. Numerical simulations illustrate the performances of these detectors, and compare them with those of the conventional adaptive matched filter.**

***Index Terms*—Adaptive detection, Bayesian model, covariance matrix estimation, heterogeneous environment, maximum *a posteriori* (MAP) estimation, minimum mean square error estimation, Monte Carlo methods.**

I. Introduction

DETECTION of a signal of interest in a background of noise is a fundamental task in many applications, including radar, communications or sonar [1]. This is especially the case for radar systems whose core task is to detect a target amongst clutter, thermal noise and possibly jamming [2]–[5]. Typically, the presence of a target, with given space and/or time signature $\boldsymbol{s}$, is sought in a (range) cell under test (CUT), given an observation vector $\boldsymbol{z}$—the primary data—that corresponds to the output of an array of sensors. In the Gaussian case, when the covariance matrix $\boldsymbol{M}_p$ of the noise in the CUT is known, the optimal processor consists of a whitening step followed by matched filtering [1]. However, the noise statistics are usually unknown and, in order to estimate them, training samples $\{\boldsymbol{z}_k\}_{k=1}^{K}$, which contain noise only, are used. These training samples, also referred to as the secondary data, are usually obtained from range cells adjacent to the CUT. The principle that underlies this approach is that information about noise in the primary data can be inferred from noise in the secondary data. This implies that the two sets of data "share" some common features.

In an ideal situation, the $\boldsymbol{z}_k$ are assumed to be independent and Gaussian distributed with a covariance matrix $\boldsymbol{M}_s$ which equals $\boldsymbol{M}_p$. This scenario is often referred to as the homogeneous environment. Within this framework, the generalized likelihood ratio test (GLRT) was derived by Kelly in the landmark paper [6]. Kelly's GLRT is based on the whole set of data, including primary and secondary data. Another reference detector, namely the adaptive matched filter (AMF), was proposed by Robey *et al.* in [7]. The AMF first proceeds as if the noise covariance matrix was known, and derives the GLRT for this case. Then, the true covariance matrix is substituted for its maximum likelihood estimate (MLE), based on secondary data. The performances of these two detectors have been evaluated thoroughly on simulated data in [6] and [7] as well as on real data, see e.g., [8] and [9]. Furthermore, their robustness to steering vector mismatches has been evaluated, see e.g., [10] and [11].

However, it has been evidenced that the homogeneous assumption is an idealized situation [4], [12], and that nonhomogeneous environments are more commonly encountered. Nonhomogeneities include a covariance mismatch between the primary and secondary data (i.e., $\boldsymbol{M}_s \neq \boldsymbol{M}_p$), or the fact that the $\boldsymbol{z}_k$'s may even not share a common covariance matrix. There exist many different reasons for which the environment may not be homogeneous. This can be due for example to the terrain: dense scattering environments (e.g., urban areas), land-sea clutter interfaces, power level fluctuations among the various patches of clutter. Accordingly, the secondary data may be contaminated by signals with signatures close to that of the target, or by unintentional temporally sporadic electromagnetic interference. Nonhomogeneous environments can also be the consequence of the array geometry. For instance, it is known that, with forward-looking or bistatic configurations, the clutter covariance matrix is not stationary in range [4]. In such nonhomogenous environments, adaptive detectors based on the homogeneous assumption incur a serious loss of performance. Thorough theoretical performance analyses of a large class of detectors (including Kelly's GLRT and the AMF) in case of

Manuscript received October 5, 2006; revised May 11, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Daniel Fuhrmann.This work was supported in part by the Délégation Générale pour l'Armement (DGA) and in part by Thales Systèmes Aéroportés. This paper was presented in part at the International Conference on Acoustics, Speech, and Signal Processing, Honolulu, HI, April 15–20, 2007.

S. Bidon and O. Besson are with the Department of Avionics and Systems, ENSICA, 31056 Toulouse, France (e-mail: sbidon@ensica.fr; besson@ensica.fr).

J.-Y. Tourneret is with the IRIT/ENSEEIHT, 31071 Toulouse, France (e-mail: jean-yves.tourneret@enseeiht.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.901664

a covariance mismatch (but under the Gaussian assumption) can for example be found in the work by Richmond [13]–[15] or Blum and Mc Donald [16], [17]. These analyses provide insightful results about the performance loss incurred under nonhomogeneity. Moreover, they highlight the parameters that contribute to performance loss, and enable one to study the robustness of conventional detectors. Reference [18] focuses on the effect of secondary data signal contamination on the performance of Kelly's GLRT. Finally, the performances of these conventional detectors have also been assessed on real data, see e.g., [9], [19]. However, these analyses do not provide a means to combat nonhomogeneity, by designing a detector that can take nonhomogeneity into account. For instance, in [14], the generalized eigenrelation is advocated to analyze the performance of various adaptive detection algorithms; however, it is not clear how to design a detector under this assumption.

In order to mitigate the effects of nonhomogeneity, two main approaches can be taken. The first consists of selecting the training samples that are most homogeneous with the CUT, and to use only the retained samples to estimate the noise covariance matrix. This approach has gained a lot of interest recently. Usually, a test statistic is computed using the training samples and compared to a threshold. The training samples that do not pass the threshold are censored, and the remaining are used to estimate the primary data covariance matrix. Several test statistics have thus been proposed, including power selection criteria [20], [21], the generalized inner-product [22], [23], the adaptive power residue [24] or the nonhomogeneity detector [25]. These selection strategies result in significant performance improvement. However they may require a large number of initial samples.

In this paper, we take the second approach which consists of taking into account the nonhomogeneity at the design stage of the detector. The principle behind this approach is to formulate the hypothesis testing problem so that it includes explicitly the nonhomogeneity in terms of a statistical model for $\boldsymbol{z}$ and $\boldsymbol{z}_k$. A key issue for such an approach is thus to have a relevant model for these environments. A first attempt to take into account nonhomogeneity is to assume that $\boldsymbol{M}_s$ is only proportional to $\boldsymbol{M}_p$, while maintaining the Gaussian assumption for the data. In this case, the reference detector is the adaptive coherence estimator (ACE), originally proposed in [26] as an adaptive version of the constant false-alarm rate (CFAR) matched subspace detectors of [1]. The ACE was shown to be the GLRT [27] as well as the uniformly most powerful invariant (UMPI) test [28], in this partially homogeneous environment. For arbitrary covariance mismatch and under the Gaussian assumption, a method is proposed in [17] where a large class of detectors, parameterized by a scalar α, is analyzed under steering vector and covariance mismatches. For instance, specific values of the scalar α lead to Kelly's GLRT, the AMF or the ACE. The paper proposes a way to find an optimal α, i.e., an α that would result in the highest probability of detection for a given probability of false alarm. The optimal α depends on $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$ through some function $g(\boldsymbol{M}_p, \boldsymbol{M}_s)$, and thus cannot be selected as these matrices are not known. However, if a range of variations for $g(\boldsymbol{M}_p, \boldsymbol{M}_s)$ is known, then a close-to-optimal value of α can be found. This is one of rare attempts to design an adaptive detector that can handle arbitrary covariance mismatch, under the Gaussian assumption. A more widely used and physically motivated model for nonhomogeneous clutter (especially with high-resolution radars) is the compound-Gaussian model, see e.g., [29], [30] or [31] and references therein. This model has been assessed on experimental data, either in the X-band or the L-band and with various types of clutter, including lake, sea or agricultural terrain clutter [32]–[34]. It allows to model local clutter power fluctuations along the range cells. Adaptive detectors able to handle this type of heterogeneity have been proposed in the literature, see e.g., [30], [35]–[37] and [31] for a rather exhaustive list of references. However, the compound-Gaussian model does not cover all types of nonhomogeneity. In particular, it is not suitable for modelling a clutter that is nonstationary in range due to the array geometry. Another widely used model for the clutter covariance matrix is proposed in [12], [38]. It consists of writing the clutter covariance matrix, at each range, as the integral -over clutter patches uniformly distributed in azimuth on an iso-range curve- of the covariance matrix of each clutter patch, weighted by the radar illumination pattern and the ground reflectivity (possible refinements include covariance matrix tapering as well). This model is physically motivated and allows one to compute the covariance matrix at each range. Hence it is a useful tool to simulate clutter data. Moreover, it has a great potential use as an *a priori* information in the so-called knowledge-aided space time adaptive processing (KA-STAP) [39], which is recognized as one of the potentially most efficient way to handle heterogeneities. KA-STAP consists of improving the performance of adaptive detection schemes using additional (*a priori*) information, such as digital elevation and terrain data, synthetic aperture radar imagery [40]. Some KA schemes have already been presented, see [40] for an overview and [41]. For instance, in [42]–[44], the model of the clutter covariance matrix described in [12], [38] is used as an *a priori* value of the actual clutter covariance matrix (CCM), and a robust adaptive beamformer is designed, based on this knowledge. The beamformer uses colored loading where the loading matrix is the model-based CCM; equivalently, it can be viewed as a prewhitening step followed by adaptation. A similar information was also used in [45]. However, the drawback of this CCM model is that it does not yield a direct relation between $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$. In addition, it does not provide a statistical model relating $\boldsymbol{z}$ and $\boldsymbol{z}_k$, which could be used to design an appropriate detector.

In this paper, we consider the problem of detecting a signal of interest in a nonhomogeneous environment, i.e., when $\boldsymbol{M}_p \neq \boldsymbol{M}_s$. Clearly, a key issue for such a problem is to have a relation between these two covariance matrices, so as to infer $\boldsymbol{M}_p$ from observation of samples whose covariance matrix is $\boldsymbol{M}_s$. Our goal in this paper is to present a simple, yet flexible statistical model to handle nonhomogeneous environments. At the same time, we want to use some *a priori* information, similarly to what is done in KA-STAP. Towards this end, a Bayesian approach is naturally advocated, as it is a relevant framework to handle uncertainties and to include *a priori* information. Additionally, it provides a theoretically sound way to define the relation between $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$. More precisely, we assume that these two matrices are random, and that the distribution of $\boldsymbol{M}_s$ given $\boldsymbol{M}_p$ is known. Furthermore, we assume that $\boldsymbol{M}_p$ has an

a priori distribution with a known mean, see below for details. This framework allows us to obtain a general and flexible, yet simple, model of nonhomogeneous environments, without very restrictive assumptions. It enables us to derive robust detectors, i.e., detectors that take into account covariance mismatches. Before closing this section, we would like to point out that such a Bayesian approach with Wishart or anti-Wishart priors has already been advocated, see e.g., [46] and [47]. In particular, [47] addresses a very general problem of adaptive detection in low-rank interference using a Bayesian approach to model the interference covariance matrix. The present work is related to [48] where we proposed a similar model, with the difference that the prior knowledge is about the average value of $\boldsymbol{M}_s$ in [48], and not about the average value of $\boldsymbol{M}_p$ as in the present paper. It turns out that the model of [48] allows one to derive a simple expression of the maximum *a posteriori* (MAP) estimate of $\boldsymbol{M}_p$, leading to a detector which amounts to colored loading of the sample covariance matrix, similarly to what is proposed in [42] and [44]. As will be shown next, the MAP and minimum mean-square error (MMSE) estimates of $\boldsymbol{M}_p$ obtained in this paper result in rather different detection schemes.

The paper is organized as follows. Section II describes formally the detection problem at hand, along with the corresponding hypotheses. In Section III, we focus on the problem of estimating $\boldsymbol{M}_p$. First, we derive the MMSE estimate of $\boldsymbol{M}_p$. As the latter cannot be obtained in closed-form a Gibbs sampling strategy is advocated. Then, a closed-form expression for the MAP estimate is presented. The estimates obtained in Section III are in turn used for detection purposes in Section IV. An approximate GLRT for this detection problem is derived, as well as two Bayesian versions of the AMF where the MLE of $\boldsymbol{M}_p$ is replaced either by the MMSE estimate of $\boldsymbol{M}_p$ or by its MAP estimate. Section V presents numerical results illustrating the performance of these detectors and comparing them with those of the conventional AMF. Conclusions and perspectives are drawn in Section VI.

II. Problem Statement

A. Notations

We briefly introduce the notations used in the paper, notably those concerning the probability density functions (p.d.f.) that will be used in the sequel. A vector $\boldsymbol{x} \in \mathbb{C}^m$ is said to have a multivariate complex normal distribution with mean $\bar{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{C}$ if its p.d.f. is given by

$$f(\boldsymbol{x}) = \pi^{-m}|\boldsymbol{C}|^{-1}\exp\left\{-(\boldsymbol{x}-\bar{\boldsymbol{x}})^H\boldsymbol{C}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}})\right\} \quad (1)$$

where $|.|$ stands for the determinant of a matrix. The standard notation $\boldsymbol{x} \sim \mathcal{CN}_m(\bar{\boldsymbol{x}}, \boldsymbol{C})$ will be used in what follows. A complex matrix $\boldsymbol{X}$ of dimension $m \times m$ is said to have a central complex Wishart distribution with $\nu \geq m$ degrees of freedom and covariance matrix $\boldsymbol{A}$ if

$$\boldsymbol{X} \sim \mathcal{CW}_m(\boldsymbol{A}, \nu) \Leftrightarrow f(\boldsymbol{X}|\boldsymbol{A}, \nu) = \frac{|\boldsymbol{X}|^{\nu-m}}{c(m,\nu)|\boldsymbol{A}|^{\nu}}\text{etr}\left\{-\boldsymbol{A}^{-1}\boldsymbol{X}\right\} \quad (2)$$

with

$$c(m,\nu) = \pi^{(m(m-1)/2)}\prod_{k=1}^{m}\Gamma(\nu-m+k) \quad (3)$$

and where $\text{etr}\{.\}$ stands for the exponential of the trace of the matrix between braces. Accordingly, $\boldsymbol{Y}$ has an inverse complex Wishart distribution with $\nu > m$ degrees of freedom and positive definite parameter matrix $\boldsymbol{B}$ if its p.d.f. can be written as

$$\boldsymbol{Y} \sim \mathcal{CW}_m^{-1}(\boldsymbol{B}, \nu) \Leftrightarrow f(\boldsymbol{Y}|\boldsymbol{B}, \nu) = \frac{|\boldsymbol{B}|^{\nu}}{c(m,\nu)|\boldsymbol{Y}|^{\nu+m}}\text{etr}\left\{-\boldsymbol{Y}^{-1}\boldsymbol{B}\right\}. \quad (4)$$

The following properties will be used repeatedly [49]. First, note the equivalence:

$$\boldsymbol{X} \sim \mathcal{CW}_m(\boldsymbol{A}, \nu) \Leftrightarrow \boldsymbol{Y} = \boldsymbol{X}^{-1} \sim \mathcal{CW}_m^{-1}\left(\boldsymbol{A}^{-1}, \nu\right). \quad (5)$$

Next, if $\boldsymbol{X} \sim \mathcal{CW}_m(\boldsymbol{A}, \nu)$ and $\boldsymbol{Y} \sim \mathcal{CW}_m^{-1}(\boldsymbol{B}, \nu)$, then [49]

$$\mathcal{E}\{\boldsymbol{X}\} = \nu\boldsymbol{A} \quad (6a)$$

$$\text{cov}\{\boldsymbol{X}\} = \nu\text{Tr}\{\boldsymbol{A}\}\boldsymbol{A} \quad (6b)$$

$$\mathcal{E}\{\boldsymbol{Y}\} = \frac{\boldsymbol{B}}{\nu-m} \quad (6c)$$

$$\text{cov}\{\boldsymbol{Y}\} = \frac{(\nu-m)^{-1}\boldsymbol{B}^2 + \text{Tr}\{\boldsymbol{B}\}\boldsymbol{B}}{(\nu-m+1)(\nu-m)(\nu-m-1)} \quad (6d)$$

where $\mathcal{E}\{.\}$ stands for the statistical mean, and $\text{cov}\{\boldsymbol{X}\}$ is defined here as $\text{cov}\{\boldsymbol{X}\} \triangleq \mathcal{E}\left\{(\boldsymbol{X}-\mathcal{E}\{\boldsymbol{X}\})^2\right\}$ (i.e., it does not denote the usual covariance matrix of the vector formed by stacking the columns of $\boldsymbol{X}$).

B. Detection Problem

The detection problem considered herein is a conventional binary composite hypothesis testing problem, defined as

$$H_0: \begin{cases} \boldsymbol{z} = \boldsymbol{n}; \\ \boldsymbol{z}_k = \boldsymbol{n}_k;\ k = 1,\ldots,K \end{cases}$$
$$H_1: \begin{cases} \boldsymbol{z} = b\boldsymbol{s} + \boldsymbol{n}; \\ \boldsymbol{z}_k = \boldsymbol{n}_k;\ k = 1,\ldots,K. \end{cases} \quad (7)$$

In (7), $\boldsymbol{z}$ is the m-length space-time snapshot for the CUT, while $\boldsymbol{z}_k$ are the training samples, obtained from adjacent range cells. The $m \times 1$ vector $\boldsymbol{s}$ is the known space-time signature of the target (referred to as the steering vector) which is usually a known function of the target's direction-of-arrival and radial velocity. The scalar b stands for the amplitude of the target, which is assumed to be deterministic and unknown.

As for the noise, we assume that the vectors $\boldsymbol{n}_k$ are independent and identically distributed (i.i.d.), $\boldsymbol{n}_k|\boldsymbol{M}_s \sim \mathcal{CN}_m(\boldsymbol{0}, \boldsymbol{M}_s)$, $k = 1,\ldots,K$, where $\boldsymbol{M}_s$ is an unknown covariance matrix. The density of $\boldsymbol{z}_k$ conditionally to $\boldsymbol{M}_s$ can thus be written as

$$f(\boldsymbol{z}_k|\boldsymbol{M}_s) = \pi^{-m}|\boldsymbol{M}_s|^{-1}\text{etr}\left\{-\boldsymbol{M}_s^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^H\right\}. \quad (8)$$

Since the $\boldsymbol{z}_k$'s are independent, the joint density of $\boldsymbol{Z} = [\,\boldsymbol{z}_1 \quad \ldots \quad \boldsymbol{z}_K\,]$, conditionally to $\boldsymbol{M}_s$, is

$$f(\boldsymbol{Z}|\boldsymbol{M}_s) = \pi^{-mK}|\boldsymbol{M}_s|^{-K}\text{etr}\left\{-\boldsymbol{M}_s^{-1}\boldsymbol{S}\right\} \quad (9)$$

where

$$\boldsymbol{S} = \sum_{k=1}^{K} \boldsymbol{z}_k \boldsymbol{z}_k^H \tag{10}$$

denotes the sample covariance matrix of the secondary data. The noise vector $\boldsymbol{n}$ is assumed to have a Gaussian distribution with zero mean and covariance matrix $\boldsymbol{M}_p$. As a consequence, the distributions of the vector $\boldsymbol{z}$ under hypotheses H_0 and H_1 are given by

$$\begin{aligned} f_0(\boldsymbol{z}|\boldsymbol{M}_p) &= \pi^{-m}|\boldsymbol{M}_p|^{-1}\mathrm{etr}\left\{-\boldsymbol{M}_p^{-1}\boldsymbol{z}\boldsymbol{z}^H\right\} \\ f_1(\boldsymbol{z}|\boldsymbol{M}_p) &= \pi^{-m}|\boldsymbol{M}_p|^{-1}\mathrm{etr}\left\{-\boldsymbol{M}_p^{-1}(\boldsymbol{z}-b\boldsymbol{s})(\boldsymbol{z}-b\boldsymbol{s})^H\right\}. \end{aligned} \tag{11}$$

Moreover, $\boldsymbol{z}$ and $\boldsymbol{z}_k$ are assumed to be independent conditionally to $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$.

This paper proposes to model the heterogeneity between the primary and secondary data by assuming that $\boldsymbol{M}_p \neq \boldsymbol{M}_s$. However, the matrix $\boldsymbol{M}_s$ is supposed to be "close" to $\boldsymbol{M}_p$. Indeed, $\boldsymbol{M}_s$ must be somehow related to $\boldsymbol{M}_p$ so as to have information about $\boldsymbol{M}_p$ from the observation of secondary data whose covariance matrix is $\boldsymbol{M}_s$ (otherwise the training samples would be useless). Herein, we propose to model the relation between $\boldsymbol{M}_s$ and $\boldsymbol{M}_p$ by assuming that the conditional distribution of $\boldsymbol{M}_s|\boldsymbol{M}_p$ is an inverse complex Wishart distribution with ν degrees of freedom, whose mean is $\boldsymbol{M}_p$:

$$f(\boldsymbol{M}_s|\boldsymbol{M}_p) \propto |\boldsymbol{M}_s|^{-(\nu+m)}\mathrm{etr}\left\{-(\nu-m)\boldsymbol{M}_s^{-1}\boldsymbol{M}_p\right\}|\boldsymbol{M}_p|^{\nu} \tag{12}$$

where $\propto$ means proportional to. Using the notations introduced in the previous section, we thus have $\boldsymbol{M}_s|\boldsymbol{M}_p \sim \mathcal{CW}_m^{-1}\left((\nu-m)\boldsymbol{M}_p, \nu\right)$. Note that the inverse complex Wishart distribution is the conjugate prior for parameter $\boldsymbol{M}_s$, which will significantly simplify the analysis. Note also that (12) implies that, "on the average", the environment is homogeneous as $\mathcal{E}\left\{\boldsymbol{M}_s|\boldsymbol{M}_p\right\} = \boldsymbol{M}_p$; however, these two matrices will be different with probability one. The parameter ν allows one to adjust the degree of heterogeneity between $\boldsymbol{M}_s$ and $\boldsymbol{M}_p$. In fact, using (6d), one can show that

$$\mathrm{cov}\left\{\boldsymbol{M}_s|\boldsymbol{M}_p\right\} = \frac{\boldsymbol{M}_p^2 + (\nu-m)\mathrm{Tr}\left\{\boldsymbol{M}_p\right\}\boldsymbol{M}_p}{(\nu-m+1)(\nu-m-1)}. \tag{13}$$

Therefore, the distance between $\boldsymbol{M}_s$ and $\boldsymbol{M}_p$ decreases as ν increases (and hence, $\boldsymbol{M}_s$ is closer to $\boldsymbol{M}_p$), which corresponds to an environment which is more and more homogeneous. In contrast, for small values of ν, $\boldsymbol{M}_s$ may significantly differ from $\boldsymbol{M}_p$. In any case, the two covariance matrices will be different, and thus the scenario is that of an heterogeneous environment.

Let us now turn to the *a priori* p.d.f. of $\boldsymbol{M}_p$. The choice of this prior is of course a delicate issue. It is usually dictated by two seemingly conflicting arguments. On one hand, the prior $f(\boldsymbol{M}_p)$ should reflect our knowledge about the primary data covariance matrix, or our absence of knowledge, which can be recast through a non informative prior. On the other hand, computational complexity is an important issue. Consequently, the prior distribution of $\boldsymbol{M}_p$ is usually chosen in order to provide tractable posterior densities. We refer the reader to [50], [51] for a very comprehensive discussion about the choice of a prior for covariance matrices. In our context, we assume that we have some rough knowledge about the average value of $\boldsymbol{M}_p$ denoted as $\bar{\boldsymbol{M}}_p$. In STAP problems, $\bar{\boldsymbol{M}}_p$ can be obtained from the simplified model of the CCM, i.e., $\bar{\boldsymbol{M}}_p$ can be computed as [38]

$$\bar{\boldsymbol{M}}_p = \sum_{k=1}^{N_c} P_k \boldsymbol{v}_k \boldsymbol{v}_k^H \odot \boldsymbol{T}_k \tag{14}$$

where N_c is the number of clutter patches evenly distributed in azimuth, P_k is the power of the kth clutter patch, $\boldsymbol{v}_k$ denotes its space-time signature, and $\boldsymbol{T}_k$ is a covariance matrix taper which accounts for intrinsic clutter motion, calibration errors, etc. As already stated in the introduction, using $\bar{\boldsymbol{M}}_p$ in (14) as an *a priori* information has been advocated in e.g., [42]–[45]. In this paper, we assume that $\boldsymbol{M}_p \sim \mathcal{CW}_m\left(\mu^{-1}\bar{\boldsymbol{M}}_p, \mu\right)$, i.e.,

$$f(\boldsymbol{M}_p) \propto |\boldsymbol{M}_p|^{\mu-m}\mathrm{etr}\left\{-\mu\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\right\}. \tag{15}$$

The average value of $\boldsymbol{M}_p$ is thus $\mathcal{E}\left\{\boldsymbol{M}_p\right\} = \bar{\boldsymbol{M}}_p$ and, using (6b), the covariance matrix of $\boldsymbol{M}_p$ is given by

$$\mathrm{cov}\left\{\boldsymbol{M}_p\right\} = \frac{\mathrm{Tr}\left\{\bar{\boldsymbol{M}}_p\right\}}{\mu}\bar{\boldsymbol{M}}_p. \tag{16}$$

As μ increases, $\boldsymbol{M}_p$ is closer to $\bar{\boldsymbol{M}}_p$, and thus the prior density $f(\boldsymbol{M}_p)$ is very informative. On the other hand, for small μ, $\boldsymbol{M}_p$ may significantly depart from $\bar{\boldsymbol{M}}_p$, which results in a vague prior density $f(\boldsymbol{M}_p)$. Hence, the scalar μ enables us to tune the amount of *a priori* knowledge we have about $\boldsymbol{M}_p$. Furthermore, it should be stressed that $\boldsymbol{M}_p$ will anyway differ from $\bar{\boldsymbol{M}}_p$. Therefore, the framework we propose offers a blend of knowledge-aided processing—through the (not too restrictive) assumption about $f(\boldsymbol{M}_p)$—and robustness to covariance matrix uncertainties, as rather heterogeneous environments can be considered. It is important to note that the proposed statistical model only requires to specify the values of the parameters ν and μ. These parameters allow one to adjust the importance of heterogeneity and the amount of prior information regarding the covariance matrix of primary data, respectively.

Of course, we do not pretend that our model follows directly from a physical reasoning, as could be the case e.g., for the compound-Gaussian model. However, it is a rather realistic model. Indeed, $\bar{\boldsymbol{M}}_p$ in (14) has a meaningful physical interpretation and the p.d.f. $f(\boldsymbol{M}_p)$ in (15) simply states that $\boldsymbol{M}_p$ may be different from $\bar{\boldsymbol{M}}_p$. Additionally, even if we do not have a direct model for $\boldsymbol{M}_s$, the problem here is to model nonhomogeneous environments, i.e., environments for which $\boldsymbol{M}_s \neq \boldsymbol{M}_p$. Therefore, the issue is more to relate $\boldsymbol{M}_s$ to $\boldsymbol{M}_p$ than to have accurate models for $\boldsymbol{M}_s$ and $\boldsymbol{M}_p$. Since the difference between $\boldsymbol{M}_s$ and $\boldsymbol{M}_p$ is unpredictable and random, a logical way to express this difference is through the conditional p.d.f. $f(\boldsymbol{M}_s|\boldsymbol{M}_p)$ in (12).

III. Estimation

A preliminary and crucial step to designing a detector is to obtain estimates of the unknown parameters of the model, and more specifically in our case, estimates of the primary data covariance matrix. In this section, we focus on this issue. As will be explained below, the detectors of Section IV will require calculating integrals of the form $\int h(\boldsymbol{M}_p) f(\boldsymbol{M}_p|\boldsymbol{Z}) d\boldsymbol{M}_p$ where $f(\boldsymbol{M}_p|\boldsymbol{Z})$ is the *a posteriori* density of $\boldsymbol{M}_p$ given $\boldsymbol{Z}$, or estimating $\boldsymbol{M}_p$. In either case, it is necessary to derive the *a posteriori* distribution $f(\boldsymbol{M}_p|\boldsymbol{Z})$ and to generate matrices distributed according to this distribution. Hence, we first derive $f(\boldsymbol{M}_p|\boldsymbol{Z})$. Observe that, in a Bayesian framework, all information about $\boldsymbol{M}_p$ is embedded in its posterior distribution $f(\boldsymbol{M}_p|\boldsymbol{Z})$, whose derivation is thus very natural. Next, we derive the MMSE and MAP estimates of $\boldsymbol{M}_p$. For the sake of readability, we recast the statistical assumptions of the paper, i.e.,

$$\boldsymbol{z}|\boldsymbol{M}_p \sim \mathcal{CN}_m(\bar{\boldsymbol{z}}, \boldsymbol{M}_p) \tag{17a}$$
$$\boldsymbol{Z}|\boldsymbol{M}_s \sim \mathcal{CN}_m(\boldsymbol{0}, \boldsymbol{M}_s) \tag{17b}$$
$$\boldsymbol{M}_s|\boldsymbol{M}_p \sim \mathcal{CW}_m^{-1}((\nu-m)\boldsymbol{M}_p, \nu) \tag{17c}$$
$$\boldsymbol{M}_p \sim \mathcal{CW}_m\left(\mu^{-1}\bar{\boldsymbol{M}}_p, \mu\right) \tag{17d}$$

where $\bar{\boldsymbol{z}} = \boldsymbol{0}$ under H_0, $\bar{\boldsymbol{z}} = b\boldsymbol{s}$ under H_1, and $\boldsymbol{s}$ is a known vector. We assume that b is a deterministic and unknown amplitude, which would be tantamount to assuming that b is a random variable with a flat prior.

A. Posterior Distribution of $\boldsymbol{M}_p$

We first concentrate on obtaining an analytical expression for $f(\boldsymbol{M}_p|\boldsymbol{Z})$. Under the hypotheses stated in (17), we have

$$\begin{aligned} f(\boldsymbol{M}_p, \boldsymbol{M}_s|\boldsymbol{Z}) &\propto f(\boldsymbol{Z}|\boldsymbol{M}_p, \boldsymbol{M}_s) f(\boldsymbol{M}_s|\boldsymbol{M}_p) f(\boldsymbol{M}_p) \\ &\propto |\boldsymbol{M}_s|^{-(\nu+m+K)} |\boldsymbol{M}_p|^{\nu+\mu-m} \\ &\quad \times \operatorname{etr}\left\{-\boldsymbol{M}_s^{-1}\left[\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p\right]\right\} \\ &\quad \times \operatorname{etr}\left\{-\mu \boldsymbol{M}_p \bar{\boldsymbol{M}}_p^{-1}\right\}. \end{aligned} \tag{18}$$

Using (18) along with (12) and (15), one obtains

$$\begin{aligned} f(\boldsymbol{M}_p|\boldsymbol{Z}) &= \int f(\boldsymbol{M}_p, \boldsymbol{M}_s|\boldsymbol{Z}) d\boldsymbol{M}_s \\ &\propto |\boldsymbol{M}_p|^{\nu+\mu-m} \operatorname{etr}\left\{-\mu \boldsymbol{M}_p \bar{\boldsymbol{M}}_p^{-1}\right\} \\ &\quad \times \int |\boldsymbol{M}_s|^{-(\nu+m+K)} \\ &\quad \times \operatorname{etr}\left\{-\boldsymbol{M}_s^{-1}\left[\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p\right]\right\} d\boldsymbol{M}_s \\ &\propto \frac{|\boldsymbol{M}_p|^{\nu+\mu-m}}{|\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p|^{\nu+K}} \operatorname{etr}\left\{-\mu \boldsymbol{M}_p \bar{\boldsymbol{M}}_p^{-1}\right\} \\ &\propto \frac{|\boldsymbol{M}_p|^{\nu}}{|\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p|^{\nu+K}} f(\boldsymbol{M}_p) \end{aligned} \tag{19}$$

where, to obtain the third line, we used the fact that the integral in the second line is the integral of an inverse Wishart distribution with parameter matrix $\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p$ and $\nu + K$ degrees of freedom. Equation (19) provides, up to a multiplicative constant, a closed-form expression for $f(\boldsymbol{M}_p|\boldsymbol{Z})$, which can be now used to derive estimates of $\boldsymbol{M}_p$.

B. MMSE Estimation

The MMSE estimate of $\boldsymbol{M}_p$ is given by

$$\begin{aligned} &\int \boldsymbol{M}_p f(\boldsymbol{M}_p|\boldsymbol{Z}) d\boldsymbol{M}_p \\ &= \frac{\int |\boldsymbol{M}_p|^{\nu} |\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p|^{-(\nu+K)} \boldsymbol{M}_p f(\boldsymbol{M}_p) d\boldsymbol{M}_p}{\int |\boldsymbol{M}_p|^{\nu} |\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p|^{-(\nu+K)} f(\boldsymbol{M}_p) d\boldsymbol{M}_p}. \end{aligned} \tag{20}$$

Unfortunately, there are no analytical expressions for the integrals in (20), and one must approximate them numerically. More generally, in the sequel, it will be required to evaluate integrals of the form $\int h(\boldsymbol{M}_p) f(\boldsymbol{M}_p|\boldsymbol{Z}) d\boldsymbol{M}_p$ for some function $h(\boldsymbol{M}_p)$. Deterministic methods are not appropriate here since these integrals involve functions of high dimensions ($\boldsymbol{M}_p$ is of size $m \times m$). In such situation, it is thus usual to resort to stochastic integration methods such as Markov chain Monte Carlo (MCMC) methods. These methods consist of generating samples distributed according to the posteriors of interest -in this case $f(\boldsymbol{M}_p|\boldsymbol{Z})$- and to use these samples to approximate the integrals to be computed. Hence, in order to obtain the MMSE estimate of $\boldsymbol{M}_p$, we resort to such a method (namely a Gibbs sampler, see Section IV for details) which consists of generating matrices $\boldsymbol{M}_p^{(i)}$ (for $i = 1, \ldots, N_r$) distributed according the posterior distribution $f(\boldsymbol{M}_p|\boldsymbol{Z})$, and of averaging these matrices. More precisely, after forgetting the N_{bi} first matrices belonging to the so-called burn-in period, the MMSE estimate can be approximated by averaging the "last" matrices generated by the Gibbs sampler, yielding the following MMSE estimate:

$$\int \boldsymbol{M}_p f(\boldsymbol{M}_p|\boldsymbol{Z}) d\boldsymbol{M}_p \simeq \frac{1}{N_r} \sum_{i=N_{bi}+1}^{N_{bi}+N_r} \boldsymbol{M}_p^{(i)} \triangleq \widehat{\boldsymbol{M}}_p^{\text{mmse}}. \tag{21}$$

C. Gibbs Sampling

We now present our strategy to generate matrices $\boldsymbol{M}_p^{(i)}$ distributed according to $f(\boldsymbol{M}_p|\boldsymbol{Z})$. First, note that the generation of matrices distributed according to (19) is not straightforward, as $f(\boldsymbol{M}_p|\boldsymbol{Z})$ does not belong to any familiar class of distributions. Instead, this paper proposes to generate matrices distributed according to the joint distribution $f(\boldsymbol{M}_p, \boldsymbol{M}_s|\boldsymbol{Z})$ using a Gibbs sampling strategy. This recursive strategy has been described in several textbooks such as [52, p. 326]. Having the matrix $\boldsymbol{M}_s^{(i)}$ at the ith iteration, the generation of $\boldsymbol{M}_p^{(i+1)}$ and $\boldsymbol{M}_s^{(i+1)}$ is achieved as follows:

- generate $\boldsymbol{M}_p^{(i+1)}$ according to $f\left(\boldsymbol{M}_p|\boldsymbol{M}_s^{(i)}, \boldsymbol{Z}\right)$;
- generate $\boldsymbol{M}_s^{(i+1)}$ according to $f\left(\boldsymbol{M}_s|\boldsymbol{M}_p^{(i+1)}, \boldsymbol{Z}\right)$.

In order to obtain the distributions $f(\boldsymbol{M}_p|\boldsymbol{M}_s, \boldsymbol{Z})$ and $f(\boldsymbol{M}_s|\boldsymbol{M}_p, \boldsymbol{Z})$, we use (18). Considering $\boldsymbol{M}_s$ as a given quantity in (18), it follows that

$$\begin{aligned} f(\boldsymbol{M}_p|\boldsymbol{M}_s, \boldsymbol{Z}) &\propto |\boldsymbol{M}_p|^{\nu+\mu-m} \\ &\quad \times \operatorname{etr}\left\{-\left[\mu\bar{\boldsymbol{M}}_p^{-1} + (\nu-m)\boldsymbol{M}_s^{-1}\right]\boldsymbol{M}_p\right\}. \end{aligned} \tag{22}$$

Similarly, if $\boldsymbol{M}_p$ is fixed in (18), then

$$\begin{aligned} f(\boldsymbol{M}_s|\boldsymbol{M}_p, \boldsymbol{Z}) &\propto |\boldsymbol{M}_s|^{-(\nu+m+K)} \\ &\quad \times \operatorname{etr}\left\{-\boldsymbol{M}_s^{-1}\left[\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p\right]\right\}. \end{aligned} \tag{23}$$

Therefore, the conditional distributions of $\boldsymbol{M}_p|\boldsymbol{M}_s,\boldsymbol{Z}$ and $\boldsymbol{M}_s|\boldsymbol{M}_p,\boldsymbol{Z}$ can be expressed as

$$\boldsymbol{M}_p|\boldsymbol{M}_s,\boldsymbol{Z} \sim \mathcal{CW}_m\left(\left[\mu\bar{\boldsymbol{M}}_p^{-1}+(\nu-m)\boldsymbol{M}_s^{-1}\right]^{-1},\quad \nu+\mu\right) \tag{24}$$

$$\boldsymbol{M}_s|\boldsymbol{M}_p,\boldsymbol{Z} \sim \mathcal{CW}_m^{-1}\left(\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p,\nu+K\right). \tag{25}$$

Consequently, the Gibbs sampling strategy generates iteratively random matrices $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$ drawn from (24) and (25). In other words, an initial value of $\boldsymbol{M}_s$ is chosen. Then, a matrix $\boldsymbol{M}_p$ drawn from the Wishart distribution (24) is generated with this initial value of $\boldsymbol{M}_s$. Next, a new matrix $\boldsymbol{M}_s$ is drawn from the inverse Wishart distribution (25) with the matrix $\boldsymbol{M}_p$ generated previously. The scheme is repeated until convergence. Observe that generating matrices according to Wishart or inverse Wishart distributions is straightforward.

The convergence properties of the Gibbs sampler are well known (see, for instance, [52, p. 325] and [53, p. 181]). Indeed, the matrices $(\boldsymbol{M}_p,\boldsymbol{M}_s)$ generated with the previous algorithm are asymptotically distributed according to $f(\boldsymbol{M}_p,\boldsymbol{M}_s|\boldsymbol{Z})$. Therefore, the MMSE estimate can be approximated by averaging the "last" matrices generated by the Gibbs sampler, as was described previously. More precisely, the N_{bi} first matrices belonging to the so-called burn-in period are not used for the estimation, and the MMSE estimate of $\boldsymbol{M}_p$ can be obtained as in (21). Accordingly, if it is required to calculate integrals of the form $\int h(\boldsymbol{M}_p)f(\boldsymbol{M}_p|\boldsymbol{Z})d\boldsymbol{M}_p$, they can be approximated by averaging $h(\boldsymbol{M}_p)$ over the matrices $\left\{\boldsymbol{M}_p^{(i)}\right\}_{i=N_{bi}+1}^{N_{bi}+N_r}$. An additional advantage of this approach is that it enables us to obtain the MMSE estimates of both $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$; the latter can be obtained from the matrices $\left\{\boldsymbol{M}_s^{(i)}\right\}_{i=N_{bi}+1}^{N_{bi}+N_r}$ generated by the Gibbs sampler. This may be of interest, e.g., to characterize the secondary data and the degree of heterogeneity between the primary and the secondary data.

Remark 1: The MMSE estimate of $\boldsymbol{M}_p$ introduced above uses the secondary data only. In principle, $\boldsymbol{M}_p$ could be estimated using both $\boldsymbol{z}$ and $\boldsymbol{Z}$. However, this approach has some drawbacks, as briefly explained below. Under the stated hypotheses, it is straightforward to show that the posterior density of $(\boldsymbol{M}_p,\boldsymbol{M}_s)$ given $(\boldsymbol{z},\boldsymbol{Z})$ is given by

$$\begin{aligned} f(&\boldsymbol{M}_p,\boldsymbol{M}_s|\boldsymbol{z},\boldsymbol{Z}) \\ &\propto f(\boldsymbol{z},\boldsymbol{Z}|\boldsymbol{M}_p,\boldsymbol{M}_s)f(\boldsymbol{M}_s|\boldsymbol{M}_p)f(\boldsymbol{M}_p) \\ &\propto |\boldsymbol{M}_s|^{-(\nu+m+K)}|\boldsymbol{M}_p|^{\nu+\mu-m-1} \\ &\quad\times \operatorname{etr}\left\{-\boldsymbol{M}_s^{-1}\left[\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p\right]\right\} \\ &\quad\times \operatorname{etr}\left\{-\mu\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\right\}\operatorname{etr}\left\{-\boldsymbol{M}_p^{-1}(\boldsymbol{z}-\bar{\boldsymbol{z}})(\boldsymbol{z}-\bar{\boldsymbol{z}})^H\right\}. \end{aligned} \tag{26}$$

It ensues that $f(\boldsymbol{M}_s|\boldsymbol{M}_p,\boldsymbol{z},\boldsymbol{Z})$ is still given by (25), and thus $\boldsymbol{z}$ does not bring any additional information regarding $\boldsymbol{M}_s$. In contrast, we now have

$$\begin{aligned} f(\boldsymbol{M}_p|\boldsymbol{M}_s,\boldsymbol{z},\boldsymbol{Z}) \propto &|\boldsymbol{M}_p|^{\nu+\mu-m-1} \\ &\times \operatorname{etr}\left\{-\boldsymbol{M}_p\left[(\nu-m)\boldsymbol{M}_s^{-1}+\mu\bar{\boldsymbol{M}}_p^{-1}\right]\right\} \\ &\times \operatorname{etr}\left\{-\boldsymbol{M}_p^{-1}(\boldsymbol{z}-\bar{\boldsymbol{z}})(\boldsymbol{z}-\bar{\boldsymbol{z}})^H\right\}. \end{aligned} \tag{27}$$

Therefore, $\boldsymbol{M}_p|\boldsymbol{M}_s,\boldsymbol{z},\boldsymbol{Z}$ has not a simple Wishart distribution. Consequently, the generation of matrices drawn from (27) is more problematic when compared to the case where we use secondary data only. Moreover, this p.d.f. depends on $\bar{\boldsymbol{z}}$ which is unknown. As a consequence, it is easier to use $\boldsymbol{Z}$ only to estimate $\boldsymbol{M}_p$, even though $\boldsymbol{z}$ contains information about $\boldsymbol{M}_p$.

D. MAP Estimation

Since obtaining the MMSE estimate is rather complicated, we now turn to the MAP estimator, which can be obtained by maximizing $f(\boldsymbol{M}_p|\boldsymbol{Z})$. Using (19), it follows that:

$$\begin{aligned} \ln f(\boldsymbol{M}_p|\boldsymbol{Z}) = &\text{const.} + (\nu+\mu-m)\ln|\boldsymbol{M}_p| \\ &-(\nu+K)\ln|\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p| - \operatorname{Tr}\left\{\mu\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\right\}. \end{aligned} \tag{28}$$

Differentiating the previous equation and equating the result to zero yields

$$\begin{aligned} &\mu(\nu-m)\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\boldsymbol{M}_p-(\nu+\mu-m)\boldsymbol{S} \\ &\qquad -\boldsymbol{M}_p\left[(\nu-m)(\mu-m-K)\boldsymbol{I}-\mu\bar{\boldsymbol{M}}_p^{-1}\boldsymbol{S}\right]=\boldsymbol{0}. \end{aligned} \tag{29}$$

The previous equation is recognized as a quadratic matrix equation. In the Appendix, we show that there exists a unique solution to (29), which is given by

$$\widehat{\boldsymbol{M}}_p^{\text{map}} = \bar{\boldsymbol{M}}_p^{1/2}\boldsymbol{U}\operatorname{diag}(\lambda_k)\boldsymbol{U}^H\bar{\boldsymbol{M}}_p^{1/2} \tag{30}$$

where $\bar{\boldsymbol{M}}_p^{1/2}$ stands for the Hermitian square-root of $\bar{\boldsymbol{M}}_p$, $\operatorname{diag}(\lambda_k)$ is a diagonal matrix with diagonal entries λ_k, $\boldsymbol{U}$ is the matrix of the eigenvectors of

$$\tilde{\boldsymbol{S}} = \bar{\boldsymbol{M}}_p^{-1/2}\boldsymbol{S}\bar{\boldsymbol{M}}_p^{-1/2} = \boldsymbol{U}\operatorname{diag}(\ell_k)\boldsymbol{U}^H$$

and

$$\begin{aligned} \lambda_k = &\left(\frac{\mu-m-K}{2\mu}-\frac{\ell_k}{2(\nu-m)}\right) \\ &+\sqrt{\left(\frac{\mu-m-K}{2\mu}-\frac{\ell_k}{2(\nu-m)}\right)^2+\frac{\nu+\mu-m}{\mu(\nu-m)}\ell_k}. \end{aligned} \tag{31}$$

It is interesting to comment on the form of $\widehat{\boldsymbol{M}}_p^{\text{map}}$. First, a (quasi) prewhitening step —with $\bar{\boldsymbol{M}}_p^{-1/2}$ as the whitening matrix—is applied. Next, instead of using the quasi-whitened sample covariance matrix, the latter is slightly modified. More precisely, its eigenvectors are retained while its eigenvalues are modified. It is interesting to note that the technique which consists of modifying the eigenvalues of the sample covariance matrix has already been proposed in the statistical literature, within the framework of robust covariance matrix estimation under Stein's loss, see e.g., [54] and [55]. Hence, the MAP estimate in (30) belongs to a known class of covariance matrix estimators. Note also that, in contrast to the MMSE estimator, the MAP estimator can be obtained in closed-form, which is very appealing from a computational point of view. It remains to evaluate whether its performance will be comparable to that of the MMSE estimator.

IV. Detection

This section focuses on the problem of deciding between hypotheses H_0 and H_1 in (7), under the assumptions stated in (17).

First, an approximate GLRT for the problem at hand is derived. Next two AMF-like detectors are proposed which are obtained after deriving the GLRT for known $\boldsymbol{M}_p$, and replacing $\boldsymbol{M}_p$ by its MMSE estimate or its MAP estimate.

A. Approximate GLRT

The GLRT for the detection problem 7 is given by [1]

$$\frac{\max_b f_1(\boldsymbol{z},\boldsymbol{Z}|b)}{f_0(\boldsymbol{z},\boldsymbol{Z})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{32}$$

where, for notational convenience, we have used subscripts $_0$ and $_1$ to denote the distributions under H_0 and H_1, respectively. Using the hierarchical structure between $(\boldsymbol{z},\boldsymbol{Z})$, $\boldsymbol{M}_s$ and $\boldsymbol{M}_p$, the density $f_1(\boldsymbol{z},\boldsymbol{Z}|b)$ is given by

$$\begin{aligned} f_1(\boldsymbol{z},\boldsymbol{Z}|b) &= \int\int f_1(\boldsymbol{z},\boldsymbol{Z}|\boldsymbol{M}_p,\boldsymbol{M}_s,b) f(\boldsymbol{M}_s|\boldsymbol{M}_p) \\ &\quad \times f(\boldsymbol{M}_p) d\boldsymbol{M}_p d\boldsymbol{M}_s \\ &\propto \int\int |\boldsymbol{M}_s|^{-(\nu+m+K)} |\boldsymbol{M}_p|^{\nu} \\ &\quad \times \operatorname{etr}\left\{-\boldsymbol{M}_s^{-1}\left[\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p\right]\right\} \\ &\quad \times f(\boldsymbol{z}|\boldsymbol{M}_p) f(\boldsymbol{M}_p) d\boldsymbol{M}_p d\boldsymbol{M}_s \\ &\propto \int \frac{|\boldsymbol{M}_p|^{\nu}}{|\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p|^{\nu+K}} f(\boldsymbol{z}|\boldsymbol{M}_p) f(\boldsymbol{M}_p) d\boldsymbol{M}_p \\ &= C \int \frac{|\boldsymbol{M}_p|^{\nu-1}}{|\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p|^{\nu+K}} \\ &\quad \times \operatorname{etr}\left\{-\boldsymbol{M}_p^{-1}(\boldsymbol{z}-b\boldsymbol{s})(\boldsymbol{z}-b\boldsymbol{s})^H\right\} f(\boldsymbol{M}_p) d\boldsymbol{M}_p. \end{aligned} \tag{33}$$

The density $f_0(\boldsymbol{z},\boldsymbol{Z})$ can be obtained by setting $b=0$ in (33). For the sake of notational convenience, let us momentarily note $u(\boldsymbol{M}_p) = \left(|\boldsymbol{M}_p|^{\nu-1}/|\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p|^{\nu+K}\right) f(\boldsymbol{M}_p)$. In order to obtain the MLE of b, one should differentiate $f_1(\boldsymbol{z},\boldsymbol{Z}|b)$ with respect to b. Using (33), one obtains

$$\begin{aligned} \frac{\partial f_1(\boldsymbol{z},\boldsymbol{Z}|b)}{\partial b} &= C \int u(\boldsymbol{M}_p) e^{-(\boldsymbol{z}-b\boldsymbol{s})^H \boldsymbol{M}_p^{-1}(\boldsymbol{z}-b\boldsymbol{s})} \\ &\quad \times \left[\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z} - b\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}\right] d\boldsymbol{M}_p. \end{aligned} \tag{34}$$

Therefore, the MLE of b must satisfy

$$b = \frac{\int u(\boldsymbol{M}_p) e^{-(\boldsymbol{z}-b\boldsymbol{s})^H \boldsymbol{M}_p^{-1}(\boldsymbol{z}-b\boldsymbol{s})} \left(\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}\right) d\boldsymbol{M}_p}{\int u(\boldsymbol{M}_p) e^{-(\boldsymbol{z}-b\boldsymbol{s})^H \boldsymbol{M}_p^{-1}(\boldsymbol{z}-b\boldsymbol{s})} \left(\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}\right) d\boldsymbol{M}_p}. \tag{35}$$

The previous equation is implicit in b as the right-hand side depends on b. There does not exist a closed-form solution for b and one should resort to iterative techniques to find the MLE of b. As this is expected to be very complicated, we propose an approximate GLRT. It is well known [6] that

$$\min_b (\boldsymbol{z}-b\boldsymbol{s})^H \boldsymbol{M}_p^{-1}(\boldsymbol{z}-b\boldsymbol{s}) = \boldsymbol{z}^H\boldsymbol{M}_p^{-1}\boldsymbol{z} - \frac{\left|\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}\right|^2}{\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}} \tag{36}$$

and is achieved for $b = \left(\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}\right)^{-1}\left(\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}\right)$. Therefore,

$$\max_b f_1(\boldsymbol{z},\boldsymbol{Z}|b) < C \int g(\boldsymbol{M}_p) \exp\left\{\frac{\left|\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}\right|^2}{\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}}\right\} f(\boldsymbol{M}_p) d\boldsymbol{M}_p \tag{37}$$

with

$$g(\boldsymbol{M}_p) = \frac{|\boldsymbol{M}_p|^{\nu-1}}{|\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p|^{\nu+K}} \operatorname{etr}\left\{-\boldsymbol{M}_p^{-1}\boldsymbol{z}\boldsymbol{z}^H\right\}. \tag{38}$$

Our approximation consists of replacing the numerator in (32) by its upper-bound in (37). Doing so, our approximate GLRT (AGLRT) can be written as

$$\frac{\int g(\boldsymbol{M}_p) \exp\left\{\frac{\left|\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}\right|^2}{\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}}\right\} f(\boldsymbol{M}_p) d\boldsymbol{M}_p}{\int g(\boldsymbol{M}_p) f(\boldsymbol{M}_p) d\boldsymbol{M}_p} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{39}$$

Note that, if $\boldsymbol{M}_p$ was known, $f(\boldsymbol{M}_p)$ would be a Dirac function, and the AGLR would boil down to $\left(\left|\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}\right|^2/\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}\right)$, which is the GLRT for known $\boldsymbol{M}_p$. Under the Bayesian framework used here, this test statistic is weighted and averaged over the density of $\boldsymbol{M}_p$.

In practice, implementation of the AGLRT requires evaluating the integrals in (39). As is well-known it is preferable to rewrite these integrals as a function of the *a posteriori* p.d.f. $f(\boldsymbol{M}_p|\boldsymbol{Z})$ rather than a function of the *a priori* p.d.f. $f(\boldsymbol{M}_p)$, as the former bears more information. Using the relation between $f(\boldsymbol{M}_p|\boldsymbol{Z})$ and $f(\boldsymbol{M}_p)$ in (19), it is straightforward to show that the AGLRT can be rewritten as

$$\frac{\int h(\boldsymbol{M}_p) \exp\left\{\frac{\left|\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}\right|^2}{\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}}\right\} f(\boldsymbol{M}_p|\boldsymbol{Z}) d\boldsymbol{M}_p}{\int h(\boldsymbol{M}_p) f(\boldsymbol{M}_p|\boldsymbol{Z}) d\boldsymbol{M}_p} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{40}$$

where

$$h(\boldsymbol{M}_p) = |\boldsymbol{M}_p|^{-1} \operatorname{etr}\left\{-\boldsymbol{M}_p^{-1}\boldsymbol{z}\boldsymbol{z}^H\right\} \propto f(\boldsymbol{z}|\boldsymbol{M}_p, H_0). \tag{41}$$

As already explained, obtaining analytical expressions for these integrals is not feasible, and numerical deterministic methods are not appropriate. In contrast, the Gibbs sampler provides random matrices drawn from the *a posteriori* distribution $f(\boldsymbol{M}_p|\boldsymbol{Z})$. Therefore, we propose to implement the AGLRT by calculating the average values of both $h(\boldsymbol{M}_p)\exp\left\{|\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}|^2/\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}\right\}$ and $h(\boldsymbol{M}_p)$ over this set of matrices. In other words, the AGLR test statistic is computed practically as

$$\text{AGLR} \simeq \frac{\sum_{i=N_{bi}+1}^{N_{bi}+N_r} h\left(\boldsymbol{M}_p^{(i)}\right) \exp\left\{\frac{|\boldsymbol{s}^H\boldsymbol{M}_p^{(i)-1}\boldsymbol{z}|^2}{\boldsymbol{s}^H\boldsymbol{M}_p^{(i)-1}\boldsymbol{s}}\right\}}{\sum_{i=N_{bi}+1}^{N_{bi}+N_r} h\left(\boldsymbol{M}_p^{(i)}\right)}. \tag{42}$$

B. Bayesian AMFs

As an alternative to the AGLRT, AMF-like detectors can be investigated, where instead of replacing $\boldsymbol{M}_p$ by its MLE (under a frequentist framework), it is replaced by its MMSE or MAP estimate based on secondary data. In order to obtain the MMSE estimate of $\boldsymbol{M}_p$, the Gibbs sampler will be used, see (21). Once the MMSE $\widehat{\boldsymbol{M}}_p^{\text{mmse}}$ of $\boldsymbol{M}_p$ is obtained, the Bayesian-AMF detector is given by

$$\frac{\left|\boldsymbol{s}^H\left(\widehat{\boldsymbol{M}}_p^{\text{mmse}}\right)^{-1}\boldsymbol{z}\right|^2}{\boldsymbol{s}^H\left(\widehat{\boldsymbol{M}}_p^{\text{mmse}}\right)^{-1}\boldsymbol{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \zeta. \tag{43}$$

The previous detector will be referred to as the BAMF-MMSE. Examining (40), it can be observed that the BAMF-MMSE is an approximation of the AGLRT, considering that the posterior p.d.f. $f(\boldsymbol{M}_p|\boldsymbol{Z})$ is highly concentrated around $\widehat{\boldsymbol{M}}_p^{\text{mmse}}$. Indeed, under this assumption

$$\int h(\boldsymbol{M}_p)\exp\left\{\frac{\left|\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{z}\right|^2}{\boldsymbol{s}^H\boldsymbol{M}_p^{-1}\boldsymbol{s}}\right\}f(\boldsymbol{M}_p|\boldsymbol{Z})d\boldsymbol{M}_p \simeq h\left(\widehat{\boldsymbol{M}}_p^{\text{mmse}}\right)\exp\left\{\frac{\left|\boldsymbol{s}^H\left(\widehat{\boldsymbol{M}}_p^{\text{mmse}}\right)^{-1}\boldsymbol{z}\right|^2}{\boldsymbol{s}^H\left(\widehat{\boldsymbol{M}}_p^{\text{mmse}}\right)^{-1}\boldsymbol{s}}\right\}$$

and the AGLRT reduces to the BAMF-MMSE. An alternative to using the MMSE estimator is to use the MAP estimator, which results in

$$\frac{\left|\boldsymbol{s}^H\left(\widehat{\boldsymbol{M}}_p^{\text{map}}\right)^{-1}\boldsymbol{z}\right|^2}{\boldsymbol{s}^H\left(\widehat{\boldsymbol{M}}_p^{\text{map}}\right)^{-1}\boldsymbol{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \xi. \tag{44}$$

We will refer to (44) as the BAMF-MAP detector.

V. Numerical Examples

In this section we first study the convergence of the Gibbs sampler. Then we illustrate the performances of the AGLRT and the Bayesian AMF detectors, and compare them with that of the AMF, which is designed under the assumption that the environment is homogeneous. In all simulations, we consider a simplified scenario. We assume an array with $m = 8$ elements and the signature of the signal of interest is $\boldsymbol{s} = [1 \quad 1 \quad \ldots \quad 1]^T$. Moreover, the nominal primary data covariance matrix is $\bar{\boldsymbol{M}}_p(k,\ell) = \beta^{|k-\ell|}$ with $\beta = 0.9$, and the number of training samples is $K = 2m = 16$.

A. Convergence of the Gibbs Sampler

It is known that the Gibbs sampler provides random matrices that are asymptotically distributed according to the target distribution. However, a critical issue is to determine the numbers of iterations N_{bi} and N_r (for burn-in and computation, respectively) that are sufficient to have an accurate estimate of $\boldsymbol{M}_p$ with (21). A first *ad hoc* method consists of assessing convergence by evaluating the MSE through the iterations. More precisely, in order to find N_r, large values of N_{bi} and N_r are first chosen, which provides a reference estimate. Next, the MSE between the estimate obtained from N_r iterations and this reference is computed. The number of iterations N_r is selected as the value above which the MSE is deemed to be sufficiently small. Such a procedure was applied in our case and the result is reported in Fig. 1. As can be observed, the MSE decreases smoothly along the iterations. In order to save computational time, and since it provides accurate estimates of $\boldsymbol{M}_p$, the value $N_r = 100$ is retained and will be used in the sequel. A similar procedure is applied to obtain N_{bi}. Fig. 2 shows the evolution of the MSE as a function of N_{bi} when $N_r = 100$. It indicates that a very short burn-in period is enough to ensure a good estimation

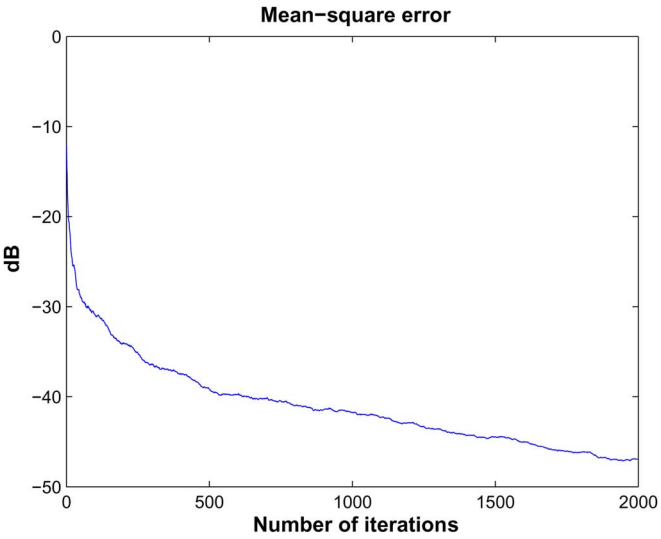


Fig. 1. Influence of the number of iterations on the convergence of the Gibbs sampler. $\nu = m+1, \mu = m$, and $K = 2m$.

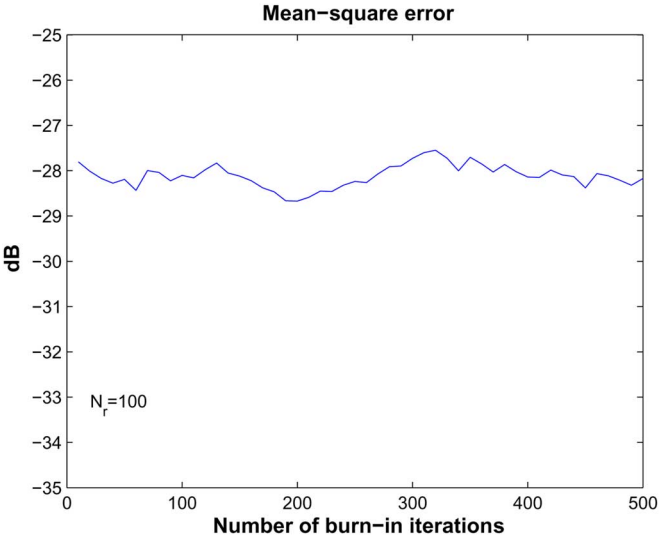


Fig. 2. Influence of the number of burn-in iterations on the convergence of the Gibbs sampler. $N_r = 100, \nu = m+1, \mu = m$, and $K = 2m$.

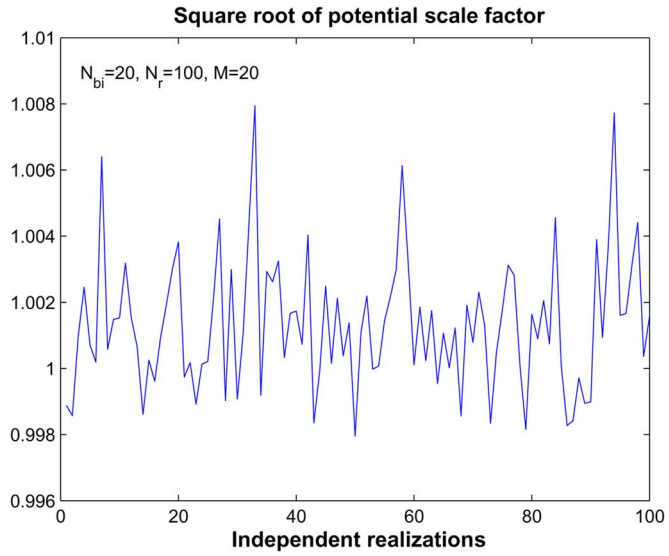


Fig. 3. Potential scale factor. $N_{bi} = 20, N_r = 100$, and $M = 20$. $\nu = m+1$, $\mu = m$, and $K = 2m$.

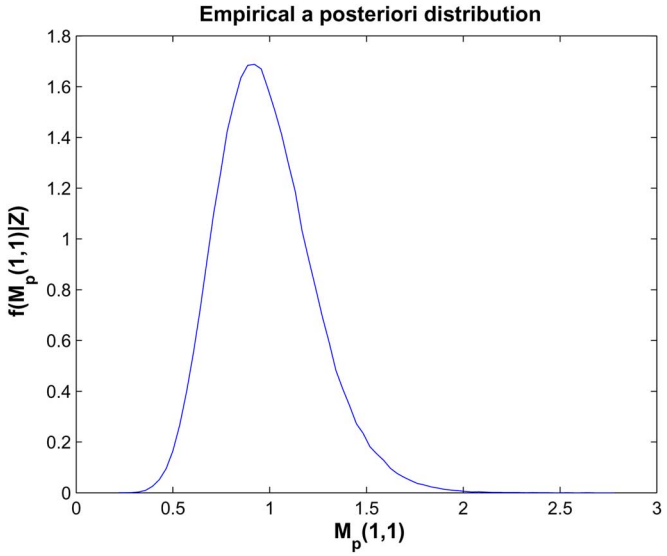


Fig. 4. Empirical *a posteriori* distribution of the $(1, 1)$ element of the matrices $\boldsymbol{M}_p$ generated by the Gibbs sampler. $\nu = m + 1, \mu = m$, and $K = 2m$.

of $\boldsymbol{M}_p$. In the sequel, we set $N_{bi} = 20$. Once these parameters have been chosen, a rigorous way to assess convergence is to use the between-within variance criterion. The principle is to run M parallel chains of length (N_{bi}, N_r), with different initial values. Let $\boldsymbol{M}_p^{(i,j)}$ be the matrix obtained at the ith iteration of the jth chain and let us note

$$\widehat{\boldsymbol{M}}_p^{(\cdot,j)} = N_r^{-1} \sum_{i=N_{bi}+1}^{N_{bi}+N_r} \boldsymbol{M}_p^{(i,j)}$$

$$\breve{\boldsymbol{M}}_p = M^{-1} \sum_{j=1}^{M} \widehat{\boldsymbol{M}}_p^{(\cdot,j)}$$

where $\widehat{\boldsymbol{M}}_p^{(\cdot,j)}$ corresponds to the MMSE for the jth chain, and $\breve{\boldsymbol{M}}_p$ is the average value over the M chains. The between-sequence and within-sequence variances for the M Markov chains corresponding to the (p, q) element of $\boldsymbol{M}_p$ (denoted as B_{pq} and W_{pq} respectively), are defined by

$$B_{pq} = \frac{N_r}{M-1} \sum_{j=1}^{M} \left(\left[\widehat{\boldsymbol{M}}_p^{(\cdot,j)} \right]_{pq} - \left[\breve{\boldsymbol{M}}_p \right]_{pq} \right)^2$$

$$W_{pq} = \frac{1}{M} \sum_{j=1}^{M} \frac{1}{N_r - 1} \sum_{i=N_{bi}+1}^{N_{bi}+N_r} \left(\left[\boldsymbol{M}_p^{(i,j)} \right]_{pq} - \left[\widehat{\boldsymbol{M}}_p^{(\cdot,j)} \right]_{pq} \right)^2.$$

The convergence of the Gibbs sampler can be monitored by the so-called potential scale factor ρ defined as [52, p. 332]

$$\rho_{pq} = \frac{N_r - 1}{N_r} + \frac{1}{N_r} \frac{B_{pq}}{W_{pq}}.$$

A value of $\sqrt{\rho_{pq}}$ less than 1.2 is recommended for convergence assessment in [52, p. 332]. Fig. 3 displays the value of $\sqrt{\rho_{11}}$ when $N_{bi} = 20$, $N_r = 100$ and $M = 20$. The experiment was run a hundred times and the values of $\sqrt{\rho_{11}}$ are plotted for these 100 independent realizations. It is clearly seen that these values of N_{bi} and N_r ensure convergence of the Gibbs sampler, which validates our selection.

As a final illustration of the Gibbs sampler properties, Fig. 4 shows the estimated posterior distribution of $[\boldsymbol{M}_p]_{11}|\boldsymbol{Z}$ computed from an histogram of the $(1, 1)$ entry of the matrices $\boldsymbol{M}_p^{(i)}$, $i = N_{bi}+1, \ldots, N_{bi}+N_r$. This posterior distribution—or more generally the estimated posterior distributions computed from $\boldsymbol{M}_p^{(i)}$—can be used to provide information on the significance of the estimates (such as confidence intervals, variances, …). It should be observed that the *a posteriori* density does not seem to be symmetric around its maximum, and hence one may expect some difference between the MAP and the MMSE estimators.

B. Detection Performance

This section compares the performance of our detectors with that of the conventional AMF. For both the AGLRT and the BAMF-MMSE, the Gibbs sampler was used with $N_{bi} = 20$ and $N_r = 100$, as validated previously. In all simulations below, the probability of false alarm is set to $P_{fa} = 10^{-3}$. The thresholds for each detector were obtained from 200,000 simulations, with different values of $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$ drawn from $f(\boldsymbol{M}_s|\boldsymbol{M}_p)$ and $f(\boldsymbol{M}_p)$ at each run. The probability of detection P_d was obtained from 100,000 Monte Carlo runs. Figs. 5–8 display P_d as a function of the signal to noise ratio (which is defined as $\text{SNR} = |b|^2 \boldsymbol{s}^H \boldsymbol{M}_p^{-1} \boldsymbol{s}$) for different values of ν and μ. Examining these figures, the following observations can be made.

- In all cases, both the AGLRT and the Bayesian AMFs outperform the AMF. For $P_d \sim 0.7$, the difference between the AGLRT and the AMF varies from 1.2 to 4.7 dB, which is significant. The BAMF-MMSE is seen to perform slightly worse than the AGLRT (the difference is about 0.6–0.7 dB) but also outperforms the conventional AMF. The BAMF-MMSE is 0.6–0.7 dB better than the BAMF-MAP. The Bayesian detectors introduced thus constitute a useful solution in this nonhomogeneous scenario. Note also that the level of performance increases with the complexity of the detector.

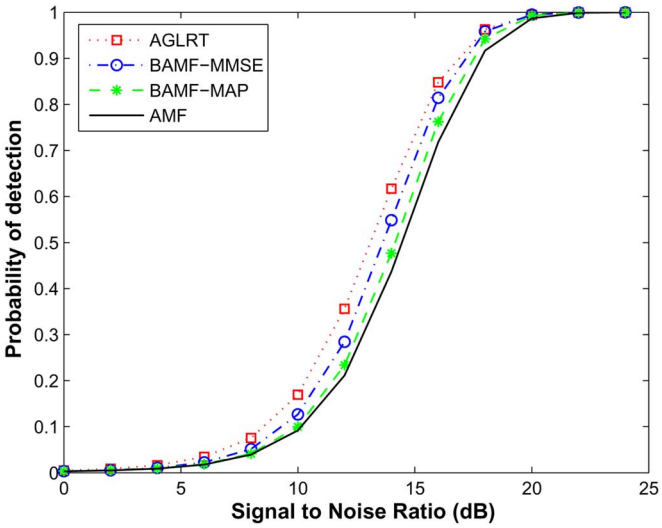


Fig. 5. Probability of detection versus SNR. $\nu = m + 1$ and $\mu = m$.

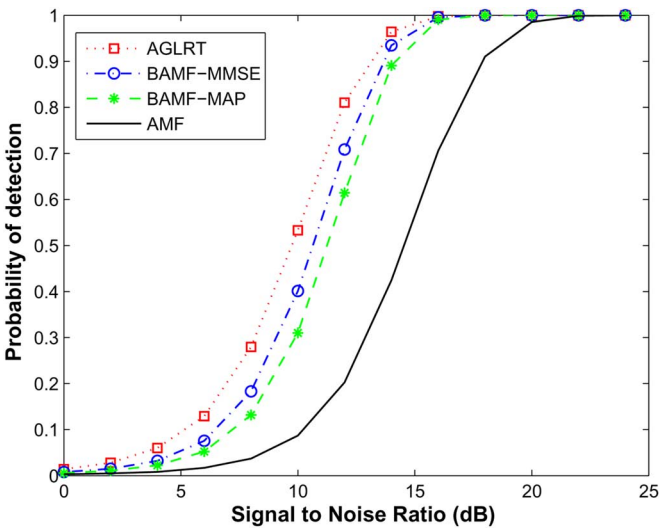


Fig. 6. Probability of detection versus SNR. $\nu = m + 1$ and $\mu = 2m$.

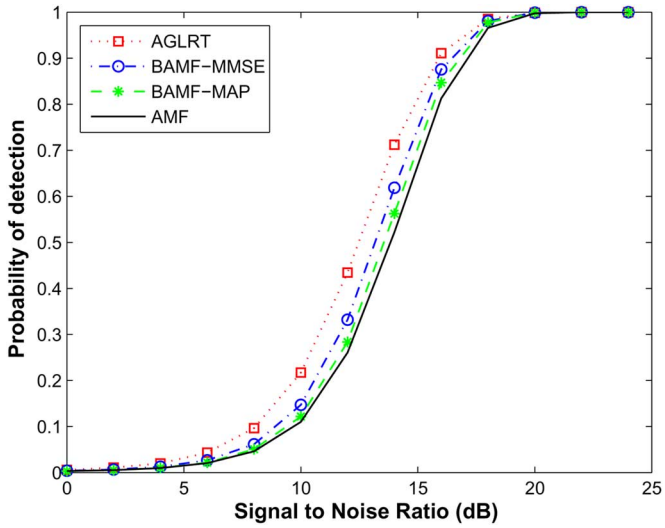


Fig. 7. Probability of detection versus SNR. $\nu = 2m$ and $\mu = m$.

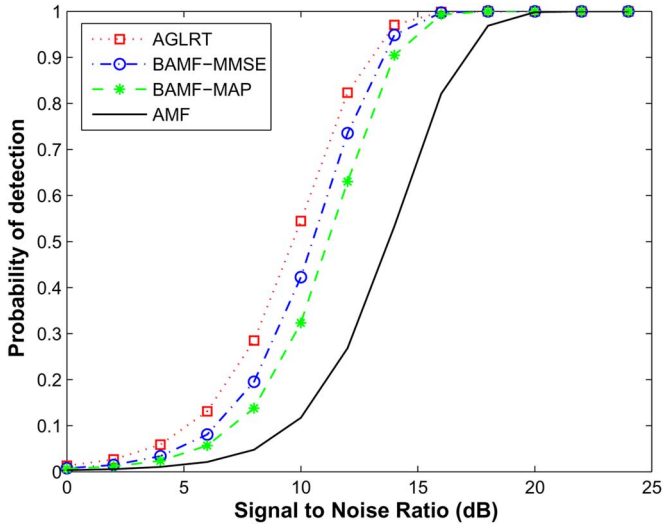


Fig. 8. Probability of detection versus SNR. $\nu = 2m$ and $\mu = 2m$.

- The performance gain of Bayesian detectors versus the conventional AMF detector is more pronounced when μ increases. This can be seen as a logical consequence of the *a priori* knowledge, namely $\bar{\boldsymbol{M}}_p$, which is more important as μ increases.
- The detection improvement resulting from the proposed Bayesian model is more significant when ν decreases, that is, as the environment is less homogeneous. Note however that the AMF is quite close to the BAMFs, whatever ν, when μ is small. Hence, in our setting, the AMF turns out to be rather robust. This can be explained by the fact that $\mathcal{E}\{\boldsymbol{M}_s|\boldsymbol{M}_p\} = \boldsymbol{M}_p$, and hence, on the average, the environment is homogeneous.
- The parameter with most influence is clearly μ. When μ goes from $\mu = 8$ to $\mu = 16$, the difference between the AGLRT and the AMF increases by about 3 dB, while when ν goes from $\nu = 16$ to $\nu = 9$, it only increases by 0.7 dB. This clearly indicates that μ is the parameter which contributes most to improvement. Hence incorporating *a priori* knowledge is a very effective means to improve performance in nonhomogeneous environments.

To summarize this section, we proposed three detection schemes. The first two, the AGLRT and the BAMF-MMSE, require a rather numerically intensive Gibbs sampling procedure while the third detector, the BAMF-MAP, is less computationally intensive, as it makes use of a closed-form expression for the MAP estimate of the primary-data covariance $\boldsymbol{M}_p$. The performance of the detectors is in inverse relation to their computational cost, with the AGLRT performing best, and the BAMF-MAP performing least among the three. That said, the BAMF-MAP still performs quite well against the standard AMF, particularly when we are given reliable prior information about the covariance, see Figs. 6 and 8.

VI. Conclusion

This paper provided a Bayesian model and knowledge-aided detectors for nonhomogeneous environments. Towards this end, a statistical relation between primary data and training samples

was proposed. More precisely, we assumed that the covariance matrices of the primary and secondary data are random, with some appropriate joint distribution. This model is rather flexible as both the importance of the *a priori* knowledge and the degree of heterogeneity can be tuned through scalar variables. Within this framework, we derived an approximate GLRT and two Bayesian versions of the AMF. Their implementation was conducted using an appropriate Gibbs sampler. The new detectors were shown to perform better than conventional detectors in nonhomogeneous scenarios. The importance of incorporating *a priori* information about the primary data covariance matrix was illustrated. Of course, some issues need to be further examined. One issue concerns the model itself. As it stands, the proposed model provides a good way of modeling information in the prior. However, modeling inhomogeneity between primary and secondary data could be improved; indeed the simulations showed that heterogeneity had less impact on the performance than prior information. Therefore, there seems to be a need to search for a different model that could handle other types and degrees of heterogeneity. A second issue concerns the computational complexity which, at this stage, still remains high. Future efforts should be devoted to more computationally efficient detection schemes. However, the proposed Bayesian detectors provide a reference to which suboptimal detectors can be compared. Finally, the robustness of these detectors should be tested in more realistic scenarios or real data. Accordingly, there is a need to derive detectors which would require less *a priori* knowledge.

APPENDIX
DERIVATION OF THE MAP ESTIMATE OF $\boldsymbol{M}_p$

In this appendix, we derive the MAP estimate of the primary data covariance matrix. As indicated previously, the MAP estimate of $\boldsymbol{M}_p$ is obtained by solving

$$\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\boldsymbol{M}_p - \boldsymbol{M}_p\left[\frac{\mu-m-K}{\mu}\boldsymbol{I} - \frac{1}{\nu-m}\bar{\boldsymbol{M}}_p^{-1}\boldsymbol{S}\right] - \frac{\nu+\mu-m}{\mu(\nu-m)}\boldsymbol{S} = \boldsymbol{0}. \quad (45)$$

For the sake of notational convenience, let us note $\alpha = (\mu - m - K/2\mu)$, $\beta = (1/2(\nu-m))$ and $\gamma = (\nu+\mu-m/\mu(\nu-m))$. Accordingly, let us define $\tilde{\boldsymbol{M}}_p = \bar{\boldsymbol{M}}_p^{-1/2}\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1/2}$ and $\tilde{\boldsymbol{S}} = \bar{\boldsymbol{M}}_p^{-1/2}\boldsymbol{S}\bar{\boldsymbol{M}}_p^{-1/2}$, where $\bar{\boldsymbol{M}}_p^{-1/2}$ is the inverse of the square root of $\bar{\boldsymbol{M}}_p$. With these new variables, (45) can be rewritten as

$$\tilde{\boldsymbol{M}}_p\tilde{\boldsymbol{M}}_p - 2\tilde{\boldsymbol{M}}_p\left[\alpha\boldsymbol{I} - \beta\tilde{\boldsymbol{S}}\right] - \gamma\tilde{\boldsymbol{S}} = \boldsymbol{0}. \quad (46)$$

Let us notice that $\tilde{\boldsymbol{M}}_p\left[\alpha\boldsymbol{I} - \beta\tilde{\boldsymbol{S}}\right] = \left[\alpha\boldsymbol{I} - \beta\tilde{\boldsymbol{S}}\right]\tilde{\boldsymbol{M}}_p$. Let

$$\tilde{\boldsymbol{S}} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{U}^H = \boldsymbol{U}\begin{pmatrix}\ell_1 & & & \\ & \ell_2 & & \\ & & \ddots & \\ & & & \ell_m\end{pmatrix}\boldsymbol{U}^H \quad (47)$$

denote the eigenvalue decomposition of $\tilde{\boldsymbol{S}}$ and let $\boldsymbol{v}$ be an eigenvector of $\tilde{\boldsymbol{M}}_p$, associated with eigenvalue λ. Then, postmultiplying (46) by $\boldsymbol{v}$, we obtain

$$\left(\lambda^2\boldsymbol{I} - 2\lambda\left[\alpha\boldsymbol{I} - \beta\tilde{\boldsymbol{S}}\right] - \gamma\tilde{\boldsymbol{S}}\right)\boldsymbol{v} = \boldsymbol{0} \quad (48a)$$

$$\Rightarrow \left|\lambda^2\boldsymbol{I} - 2\lambda\left[\alpha\boldsymbol{I} - \beta\tilde{\boldsymbol{S}}\right] - \gamma\tilde{\boldsymbol{S}}\right| = 0 \quad (48b)$$

$$\Rightarrow \left|\lambda^2\boldsymbol{I} - 2\lambda\left[\alpha\boldsymbol{I} - \beta\boldsymbol{L}\right] - \gamma\boldsymbol{L}\right| = 0 \quad (48c)$$

$$\Rightarrow \prod_{k=1}^{m}\left(\lambda^2 - 2\lambda\left[\alpha - \beta\ell_k\right] - \gamma\ell_k\right) = 0. \quad (48d)$$

However, since $\gamma\ell_k > 0$ and $(\alpha-\beta\ell_k)^2+\gamma\ell_k > 0$, the polynomial $P_k(\lambda) = \lambda^2 - 2\lambda\left[\alpha - \beta\ell_k\right] - \gamma\ell_k$ has two real roots, one being positive, the other being negative. Since the eigenvalues of $\tilde{\boldsymbol{M}}_p$ must be positive, they are necessarily in the set

$$\lambda_k = (\alpha - \beta\ell_k) + \sqrt{(\alpha-\beta\ell_k)^2 + \gamma\ell_k}, \quad k = 1, \ldots, m. \quad (49)$$

Let $\boldsymbol{v}_k$ be the eigenvector associated with λ_k. Then, using (48a), we have

$$\left(\lambda_k^2 - 2\alpha\lambda_k\right)\boldsymbol{v}_k = \left(\gamma - 2\beta\lambda_k\right)\tilde{\boldsymbol{S}}\boldsymbol{v}_k.$$

It is straightforward to show that $(\gamma - 2\beta\lambda_k)$ cannot be equal to zero: using (49), $(\gamma - 2\beta\lambda_k) = 0$ implies that $(\nu+\mu-m)(\nu+K) = 0$, which is impossible. Therefore

$$\tilde{\boldsymbol{S}}\boldsymbol{v}_k = \frac{\lambda_k^2 - 2\alpha\lambda_k}{\gamma - 2\beta\lambda_k}\boldsymbol{v}_k = \ell_k\boldsymbol{v}_k \quad (50)$$

where, to obtain the last equality, we made use of (48d). Consequently, $\boldsymbol{v}_k$ is the eigenvector of $\tilde{\boldsymbol{S}}$ associated with ℓ_k. Since $\boldsymbol{M}_p$ is positive definite, the solution to (46) is thus unique and is given by

$$\begin{aligned}\tilde{\boldsymbol{M}}_p &= \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H \\ &= \boldsymbol{U}\mathrm{diag}\Big((\alpha-\beta\ell_k) + \sqrt{(\alpha-\beta\ell_k)^2+\gamma\ell_k}\Big)\boldsymbol{U}^H. \end{aligned} \quad (51)$$

This means that $\tilde{\boldsymbol{M}}_p$ and $\tilde{\boldsymbol{S}}$ share the same eigenvectors but have different eigenvalues. Equivalently, $\tilde{\boldsymbol{M}}_p$ can be written as

$$\begin{aligned}\tilde{\boldsymbol{M}}_p &= \boldsymbol{U}\left[\alpha\boldsymbol{I} - \beta\boldsymbol{L}\right]\boldsymbol{U}^H + \boldsymbol{U}\left(\gamma\boldsymbol{L} + \left[\alpha\boldsymbol{I} - \beta\boldsymbol{L}\right]^2\right)^{1/2}\boldsymbol{U}^H \\ &= \alpha\boldsymbol{I} - \beta\tilde{\boldsymbol{S}} + \left[\gamma\tilde{\boldsymbol{S}} + \left(\alpha\boldsymbol{I} - \beta\tilde{\boldsymbol{S}}\right)\left(\alpha\boldsymbol{I} - \beta\tilde{\boldsymbol{S}}\right)^H\right]^{1/2}. \end{aligned} \quad (52)$$

Finally, the MAP estimate of $\boldsymbol{M}_p$ is given by

$$\widehat{\boldsymbol{M}}_p^{\mathrm{map}} = \bar{\boldsymbol{M}}_p^{1/2}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H\bar{\boldsymbol{M}}_p^{1/2}.$$

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the anonymous reviewers for their comments which helped them to significantly improve the paper. They are also grateful to A. Ferrari for his helpful comments about the derivation of the MAP estimator.

REFERENCES

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

[2] A. Farina, *Antenna Based Signal Processing Techniques for Radar Systems*. Norwood, MA: Artech House, 1992.

[3] H. L. V. Trees, *Optimum Array Processing*. New York: Wiley, 2002.

[4] R. Klemm, *Principles of Space-Time Adaptive Processing*, ser. IEE Radar, Sonar, Navigation and Avionics Series 12. London, U.K.: The Institution of Electrical Engineers, 2002.

[5] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA: Artech House, 2003.

[6] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 115–127, Mar. 1986.

[7] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[8] G. A. Fabrizio, A. Farina, and M. D. Turley, "Spatial adaptive subspace detection in OTH radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1407–1428, Oct. 2003.

[9] A. D. Maio, G. Foglia, E. Conte, and A. Farina, "CFAR behavior of adaptive detectors: An experimental analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 1, pp. 233–251, Jan. 2005.

[10] E. J. Kelly, "Performance of an adaptive detection algorithm; rejection of unwanted signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 2, pp. 122–133, Apr. 1989.

[11] S. Z. Kalson, "An adaptive array detector with mismatched signal rejection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 195–207, Jan. 1992.

[12] 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.

[13] C. D. Richmond, "Response of sample covariance based MVDR beamformer to imperfect look and inhomogeneities," *IEEE Signal Process. Lett.*, vol. 2, pp. 325–327, Dec. 1998.

[14] C. D. Richmond, "Performance of a class of adaptive detection algorithms in nonhomogeneous environments," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1248–1262, May 2000.

[15] C. D. Richmond, "Statistics of adaptive nulling and use of the generalized eigenrelation (GER) for modeling inhomogeneities in adaptive processing," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1263–1273, May 2000.

[16] R. S. Blum and K. F. McDonald, "Analysis of STAP algorithms for cases with mismatched steering and clutter statistics," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 301–310, Feb. 2000.

[17] K. F. McDonald and R. S. Blum, "Exact performance of STAP algorithms with mismatched steering and clutter statistics," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2750–2763, Oct. 2000.

[18] 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.

[19] F. Gini, M. V. Greco, M. Diani, and L. Verrazzani, "Performance analysis of two adaptive radar detectors against non-Gaussian real sea clutter data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1429–1439, Oct. 2000.

[20] D. J. Rabideau and A. O. Steinhardt, "Improving the performance of adaptive arrays in nonstationary environments through data-adaptive training," in *Proc. 30th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 3–6, 1996, pp. 75–79.

[21] 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.

[22] W. L. Melvin and M. C. Wicks, "Improving practical space-time adaptive radar," in *Proc. IEEE Nat. Radar Conf.*, May 13–15, 1997, pp. 48–53.

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

[24] 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.

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

[26] L. L. Scharf and T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Proc. 30th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 3–6, 1996, pp. 1114–1117.

[27] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.

[28] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: A uniformly most powerful invariant adaptive detection statistic," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 427–438, Feb. 2005.

[29] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant process," *Proc. Inst. Elect. Eng. Radar, Sonar Navigation*, vol. 134, no. 2, pp. 191–197, Apr. 1987.

[30] K. J. Sangston and K. R. Gerlach, "Coherent detection of radar targets in a non-Gaussian background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 2, pp. 330–340, Apr. 1994.

[31] F. Gini, A. Farina, and M. Greco, "Selected list of references on radar signal processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 1, pp. 329–359, Jan. 2001.

[32] J. B. Billingsley, A. Farina, F. Gini, M. V. Greco, and L. Verrazzani, "Statistical analyses of measured radar ground clutter data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 2, pp. 579–593, Apr. 1999.

[33] E. Conte, A. De Maio, and C. Galdi, "Statistical analysis of real clutter at different range resolutions," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 903–918, Jul. 2004.

[34] E. Conte, A. De Maio, and A. Farina, "Statistical tests for higher order analysis of radar clutter—Their analysis to L-band measured data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 1, pp. 205–218, Jan. 2005.

[35] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 617–625, Apr. 1995.

[36] E. Conte, M. Lops, and G. Ricci, "Adaptive matched filter detection in spherically invariant noise," *IEEE Signal Process. Lett.*, vol. 3, pp. 248–250, Aug. 1996.

[37] E. Conte, M. Lops, and G. Ricci, "Adaptive detection schemes in compound-Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 4, pp. 1058–1069, Oct. 1998.

[38] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Laboratory, MIT, Lexington, MA, 1994, Tech. Rep. 1015.

[39] "Knowledge-based systems for adaptive radar: Detection, tracking and classification," *IEEE Signal Process. Mag., Special Issue*, Jan. 2006.

[40] J. R. Guerci and E. J. Baranoski, "Knowledge-aided adaptive radar at DARPA—An overview," *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 41–50, Jan. 2006.

[41] *KASSPER Workshops, 2002–2005* [Online]. Available: http://www.darpa.mil/spo/programs/kassper.htm

[42] J. S. Bergin, C. M. Teixeira, P. M. Techau, and J. R. Guerci, "Space-time beamforming with knowledge-aided constraints," presented at the 11th Adaptive Sensor Array Process. Workshop, Lexington, MA, Mar. 11–12, 2003.

[43] C. M. Teixeira, J. S. Bergin, and P. M. Techau, "Reduced degree-of-freedom STAP with knowledge-aided data prewhitening," presented at the KASSPER Workshop, Las Vegas, NV, Apr. 15–16, 2003.

[44] J. S. Bergin, C. M. Teixeira, P. M. Techau, and J. R. Guerci, "STAP with knowledge-aided data prewhitening," in *Proc. IEEE Radar Conf.*, Philadelphia, PA, Apr. 26–29, 2004, pp. 289–294.

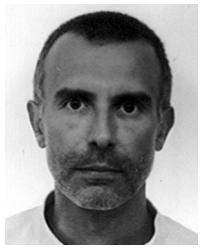
[45] K. Gerlach and M. L. Picciolo, "Airborne/spacebased radar STAP using a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 1, pp. 269–281, Jan. 2003.

[46] S. L. Earp and L. W. Nolte, "Multichannel adaptive array processing for optimal detection," in *Proc. ICASSP Conf.*, Mar. 1984, vol. 1, pp. 267–270.

[47] A. Dogandžić and B. Zhang, "Bayesian complex amplitude estimation and adaptive signal detection in low-rank interference," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1176–1182, Mar. 2007.
[48] O. Besson, J.-Y. Tourneret, and S. Bidon, "Knowledge-aided Bayesian detection in heterogeneous environments," *IEEE Signal Process. Lett.*, vol. 14, pp. 355–358, May 2007.
[49] J. A. Tague and C. I. Caldwell, "Expectations of useful complex Wishart forms," *Multidimens. Syst. Signal Process.*, vol. 5, pp. 263–279, 1994.
[50] L. Svensson, "Bayesian inference with unknown noise covariance," Ph.D. dissertation, Chalmers Univ. Technol., Göteborg, Sweden, 2004.
[51] L. Svensson and M. Lundberg, "On posterior distributions for signals in Gaussian noise with unknown covariance matrix," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3554–3571, Sep. 2005.
[52] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 1995.
[53] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. New York: Springer-Verlag, 2005.
[54] D. K. Dey and C. Srinivasan, "Estimation of a covariance matrix under Stein's loss," *Annal. Statist.*, vol. 13, no. 4, pp. 1581–1591, Dec. 1985.
[55] T. Kubokawa and M. S. Srivastava, "Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution," *Annal. Statist.*, vol. 27, no. 2, pp. 600–609, Apr. 1999.

Stéphanie Bidon (S'06) received the Engineer degree in aeronautics and the M.Sc. degree in signal processing from ENSICA, Toulouse, France, in 2004 and 2005, respectively. She is currently pursuing the Ph.D. degree with the Department of Avionics and Systems, ENSICA.

Her thesis deals with space time adaptive processing for airborne radars.



Olivier Besson (S'90–M'92–SM'04) received the Ph.D. Degree in signal processing and the "Habilitationà Diriger des Recherches," both from Institut National Polytechnique, Toulouse, France, in 1992 and 1998, respectively.

Since October 1993, he has been with the Department of Avionics and Systems, ENSICA, where he is now an Associate Professor. His research interests are in the general area of statistical signal and array processing with particular interest to robustness issues in detection/estimation problems for radar and communications. He has held visiting positions at Brigham Young University, Provo, UT, and Università del Salento, Lecce, Italy.

Dr. Besson is a past Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS. He is a member of the IEEE Sensor Array and Multichannel (SAM) Technical Committee, and served as the Co-Technical Chairman of the IEEE SAM2004 Workshop.

Jean-Yves Tourneret (M'94) received the Ingénieur degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique (ENSEEIHT), Toulouse, France. He received the Ph.D. degree from the National Polytechnic Institute, Toulouse, in 1992.

He is currently a Professor in the University of Toulouse (ENSEEIHT). He is a member of the IRIT laboratory (UMR 5505 of the CNRS), where his research activity is centered around estimation, detection and classification of non-Gaussian and nonstationary processes.

Dr. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO), which was held in Toulouse, in September 2002. He was also a member of the organizing committee for the International Conference on Acoustics, Speech, and Signal Processing which was held in Toulouse, France, in May 2006 (ICASSP'06). He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society.