



**HAL**  
open science

# Knowledge-aided STAP in heterogeneous clutter using a hierarchical bayesian algorithm

Stéphanie Bidon, Olivier Besson, Jean-Yves Tourneret

## ► To cite this version:

Stéphanie Bidon, Olivier Besson, Jean-Yves Tourneret. Knowledge-aided STAP in heterogeneous clutter using a hierarchical bayesian algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 2011, 47 (3), pp.1863 -1879. 10.1109/TAES.2011.5937270 . hal-03610420

**HAL Id: hal-03610420**

**<https://hal.science/hal-03610420v1>**

Submitted on 16 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author -deposited version published in: <http://oatao.univ-toulouse.fr/>  
Eprints ID: 4909

To link to this article: DOI: 10.1109/TAES.2011.5937270

URL: <http://dx.doi.org/10.1109/TAES.2011.5937270>

**To cite this version:** BIDON Stéphanie, BESSON Olivier, TOURNERET Jean-Yves  
Knowledge-aided STAP in heterogeneous clutter using a hierarchical bayesian algorithm.  
*IEEE Transactions on Aerospace and Electronic Systems*, 2011, vol. 47, n° 3, pp. 1863 -  
1879. ISSN 0018-9251

Any correspondence concerning this service should be sent to the repository administrator:  
[staff-oatao@inp-toulouse.fr](mailto:staff-oatao@inp-toulouse.fr)

# Knowledge-Aided STAP in Heterogeneous Clutter using a Hierarchical Bayesian Algorithm

STÉPHANIE BIDON, Member, IEEE  
OLIVIER BESSON, Senior Member, IEEE  
University of Toulouse, ISAE

JEAN-YVES TOURNERET, Senior Member, IEEE  
University of Toulouse, IRIT/ENSEEIH  
France

**The problem of estimating the covariance matrix of a primary vector from heterogeneous samples and some prior knowledge is addressed, under the framework of knowledge-aided space-time adaptive processing (KA-STAP). More precisely, a Gaussian scenario is considered where the covariance matrix of the secondary data may differ from the one of interest. Additionally, some knowledge on the primary data is supposed to be available and summarized in a prior matrix. Two KA-estimation schemes are presented in a Bayesian framework whereby the minimum mean square error (MMSE) estimates are derived. The first scheme is an extension of a previous work and takes into account the nonhomogeneity via an original relation. In search of simplicity and to reduce the computational load, a second estimation scheme, less complex, is proposed and omits the fact that the environment may be heterogeneous. Along the estimation process, not only the covariance matrix is estimated but also some parameters representing the degree of a priori and/or the degree of heterogeneity. Performance of the two approaches are then compared using STAP synthetic data. STAP filter shapes are analyzed and also compared with a colored loading technique.**

## I. INTRODUCTION

Detecting targets in highly heterogeneous environments is a challenging task for future airborne radars. Signal processing has to enable the extraction of targets embedded in noise consisting mostly of thermal noise, possibly jammers, and ground clutter whose characteristics may vary rapidly over azimuth and range. Space-time adaptive processing (STAP) is recognized today as the best candidate to perform this task [1–3]. STAP performance depends mostly on the knowledge of the noise covariance matrix. As the noise cannot entirely be known a priori, its covariance matrix is usually estimated from secondary range cells. In a homogeneous environment, where the range gates share the same covariance matrix, the sample covariance matrix (SCM) maximizes the signal-to-noise-plus-interference ratio (SINR) [4] and is also the maximum likelihood estimator [5]. Unfortunately, in real-world scenarios, the homogeneity assumption is often not satisfied. Thus the SCM may be a poor estimate of the true noise covariance matrix.

Heterogeneity can be caused by many phenomena [6, 7] including nonhomogeneous ground reflectivity (e.g., amplitude and/or spectral variation, clutter edges, discretized), secondary targets in the training interval or in the cell under test, and range-dependence of the clutter frequencies in the angle-Doppler domain (often referred to as nonstationary clutter). A variety of models have been proposed to measure the impact of several types of heterogeneity. Most of them rely on the general clutter model (GCM) covariance matrix developed by the radar community [1]. Thus, amplitude and spectral variations but also target-like signals in the secondary data can severely compromise STAP performance [8]. Highly nonstationary clutter can lead as well to very poor detection performance around clutter angle-Doppler loci [2].

Many techniques have been considered to counteract the deleterious effect of heterogeneity. Low sample support algorithms intend to reduce the presence of heterogeneous samples in the training interval. Among them one can mention reduced rank or reduced dimension algorithms [1, 2, 9], diagonal loading [10], and estimation schemes based on structured interferences [11, 12]. On the other hand, a careful selection of the secondary data allows one to discard heterogeneous samples according to a certain criterion based, e.g., on power considerations [13] or on more complex metrics such as nonhomogeneity detectors (NHD) [14]. Another way to deal with heterogeneity is to incorporate it directly via a model into the detection scheme. Some detectors for instance allow one to take into account local power fluctuations [15, 16]. However, one of the most promising ways to enhance the detection performance in heterogeneous environments may be the use of a priori knowledge.

Such algorithms are referred to as knowledge-aided STAP (KA-STAP) and are often merged with the former cited strategies.

This last decade KA-processing has received a growing interest and seems to be a key concept for the next generation of adaptive radar systems [17, 18]. In this sense, the Defense Advanced Research Projects Agency (DARPA) has initiated the KA sensor signal processing and expert reasoning (KASSPER) workshop [19]. KA-STAP aims at using prior knowledge provided by external databases (e.g., digital ground model, Global Positioning System (GPS), previous scanning data) to assist and improve detection. The a priori information can be used in two ways [20, 21]: either indirectly, e.g., to select representative training data [22, 23], or directly to form the filter. Regarding this last concern, it is generally assumed that the prior information can be summarized into a known matrix. This matrix is often built upon the GCM, where parameter values have been set up at the system design stage or by measurements. One delicate issue is then to adequately use this a priori matrix to form the STAP filter.

Among the algorithms that allow one to incorporate the a priori matrix into the detection scheme, one can mention the fast maximum likelihood with assumed clutter covariance [24], the KA parametric estimation [25] and colored loading (CL). CL is a commonly used technique whereby the covariance matrix estimate is formed as a weighted sum of the SCM and the a priori covariance matrix [26–33]. CL has an appealing form because the STAP filter can be implemented in two steps: a prewhitening step based on the prior matrix, followed by adaptive filtering [34]. Interestingly, the technique happens to be the solution to different problem formulations. In [27], [28], [34], CL is presented as the solution of the usual linearly constrained minimum variance space-time beamformer with an additional quadratic constraint. In [35]–[37], a Bayesian approach is undertaken instead and leads also to the CL technique. Though CL seems to be an efficient method to incorporate a priori information, the choice of the weighting factors remains a delicate issue and is discussed in [21], [31], [32]. An adaptive approach seems obviously more adequate. In [21], the weight of the prior matrix is derived so as to maximally whiten the observed interference data. In [31], a maximum-likelihood approach is invoked under the restrictive assumption that the weighting factors are range independent. In [32], this restrictive assumption is alleviated, and the weights are derived so as to minimize the mean-squared error (MSE) of the CL estimator.

Recently, so as to integrate the prior matrix in the estimation scheme, we have focussed our attention on a new Bayesian data model [33, 38]. The Bayesian approach turns out to be appropriate both to take

into account the nonhomogenous assumption (in a certain way) and to incorporate the prior matrix. The proposed model is tuned by two scalars called hyperparameters that were shown to represent the degree of heterogeneity of the environment and the degree of a priori knowledge. In [38], the minimum mean-squared estimator (MMSE) of the noise covariance matrix is derived, assuming these two scalars are exactly known. Obviously, for practical applications, these quantities are unknown and have to be chosen properly. To do so, we propose here to extend the work of [38] by introducing a hierarchical level to the data model. More precisely, we consider both the degree of heterogeneity and the degree of a priori knowledge as random variables with known priors and estimate them jointly with the covariance matrix. This new estimation strategy is therefore a robustified version of the algorithm presented in [38]. After describing the robustified algorithm thoroughly, we study its performance for a STAP scenario. Particularly, we intend to observe the STAP filter shape and the values of the estimated hyperparameters to highlight how the prior information is incorporated and how the heterogeneity is dealt with. Note that we do not focus our attention on the generation of the prior matrix (though a major topic). Additionally, we propose another robustified algorithm in the case where the heterogeneity would not be taken into account and show that this omission can lead to severe loss. This could happen for instance if, after an NHD, some heterogeneous samples still remained in the training interval.

The remainder of the paper is organized as follows. Section II describes the Bayesian data model and the robustified estimation procedure originally introduced in [39]. Section III presents a similar algorithm, but this algorithm alleviates the heterogeneous assumption so as to obtain less complex estimators. Section IV provides numerical results obtained from synthetic STAP data. Conclusions and perspectives are finally drawn in Section V.

## II. KNOWLEDGE-AIDED ESTIMATION IN A HETEROGENEOUS ENVIRONMENT

This section describes the data model for a KA-heterogenous environment with its MMSE estimation procedure initially introduced in [39]. The model, defined under a Bayesian framework, entails two original relations, one describing the heterogeneity of the environment and the other describing how the prior information is related to the primary data. These two relations are partly chosen to ensure mathematical tractability. Thus, so as to robustify the model, we turn to a common strategy encountered in the Bayesian philosophy, the hierarchical Bayes modeling [40]. More precisely, an additional hierarchical level is added to the data

model whereby the hyperparameters—representing the degrees of heterogeneity and a priori of the model—are considered as random variables with noninformative priors.

#### A. Data Model

We intend to estimate the covariance matrix  $\mathbf{M}_p$  of a  $\zeta$ -length complex data vector  $\mathbf{z}$  with both  $K$  secondary data  $\mathbf{Z} = [\mathbf{z}_1 \cdots \mathbf{z}_K]$  and some a priori information collected into a prior matrix  $\bar{\mathbf{M}}_p$ . As a first attempt to model heterogeneity, the  $\mathbf{z}_k$ s are supposed to be independent and Gaussian distributed with the same covariance matrix  $\mathbf{M}_s$ , which may differ from  $\mathbf{M}_p$ . This distribution is denoted by

$$\mathbf{Z} | \mathbf{M}_s \sim \tilde{\mathcal{N}}_{\zeta, K}(\mathbf{0}, \mathbf{M}_s). \quad (1)$$

The probability density function (pdf) of  $\mathbf{Z} | \mathbf{M}_s$  is thus given by

$$f(\mathbf{Z} | \mathbf{M}_s) = \pi^{-\zeta K} |\mathbf{M}_s|^{-K} \text{etr}\{-\mathbf{M} - s^{-1}\mathbf{S}\} \quad (2)$$

where  $\mathbf{S} = \mathbf{Z}\mathbf{Z}^H$ ,  $|\cdot|$  and  $\text{etr}\{\cdot\}$  stand for the determinant and the exponential of the trace of a matrix, respectively. To complete the data model, a Bayesian framework is then advocated so that both the primary and secondary covariance matrices are considered as random variables with known priors. Choosing these two priors is rather delicate. As presented hereafter, we have chosen Wishart and inverse Wishart distributions. These priors have already been invoked under statistical considerations in similar contexts [41, 42]. In a first step, the two priors are designed so that the primary matrix  $\mathbf{M}_p$  departs possibly from the secondary covariance  $\mathbf{M}_s$  and/or from the prior matrix  $\bar{\mathbf{M}}_p$ . The pdfs are also chosen among conjugate priors to make the model suitable for further derivations. In a second step, a hierarchical level is added to robustify these conjugate prior distributions, i.e., making the model less sensitive to the choice of these two priors.

1) *Heterogeneity Model*: Unlike the physical-based model [1], the heterogeneity is described here by the pdf of the secondary covariance matrix  $\mathbf{M}_s$  which has to be thoroughly designed. A complex inverse Wishart distribution—with  $\nu$  degrees of freedom and associated covariance matrix  $(\nu - \zeta)\mathbf{M}_p$ —is chosen as an adequate candidate and is denoted by

$$\mathbf{M}_s | \mathbf{M}_p, \nu \sim \tilde{\mathcal{W}}_{\zeta}^{-1}((\nu - \zeta)\mathbf{M}_p, \nu) \quad (3)$$

with pdf

$$f(\mathbf{M}_s | \mathbf{M}_p, \nu) = \frac{|(\nu - \zeta)\mathbf{M}_p|^\nu}{\tilde{\Gamma}_{\zeta}(\nu) |\mathbf{M}_s|^{(\nu + \zeta)}} \text{etr}\{-(\nu - \zeta)\mathbf{M}_s \mathbf{M}_p\} \quad (4)$$

where

$$\tilde{\Gamma}_{\zeta}(\nu) = \pi^{\zeta(\zeta - 1)/2} \prod_{k=1}^{\zeta} \Gamma(\nu - \zeta + k) \quad (5)$$

and where  $\Gamma(\cdot)$  is the gamma function. To be correctly defined,  $\nu$  has to be greater than or equal to  $\zeta$ , and the parameter matrix has to be Hermitian positive definite (i.e.,  $\nu > \zeta$ ).

The distribution (3) implies that on average the environment is homogenous, i.e.,  $\mathcal{E}\{\mathbf{M}_s | \mathbf{M}_p, \nu\} = \mathbf{M}_p$ , but ensures that  $\mathbf{M}_s$  always differs from  $\mathbf{M}_p$ . Looking now at the second-order moment, it is clear that the hyperparameter  $\nu$  monitors the distance between  $\mathbf{M}_p$  and  $\mathbf{M}_s$  via the relation

$$\mathcal{E}\{(\mathbf{M}_s - \mathbf{M}_p)^2 | \mathbf{M}_p, \nu\} = \frac{\mathbf{M}_p^2 + (\nu - \zeta)\text{Tr}\{\mathbf{M}_p\}\mathbf{M}_p}{(\nu - \zeta + 1)(\nu - \zeta - 1)}. \quad (6)$$

The larger  $\nu$  is, the more homogeneous the environment. Thus, in the following, the hyperparameter  $\nu$  is referred to as the degree of heterogeneity of the environment. Note also that the former expression (6) is defined only if  $\nu > \zeta + 1$ .

Of course we do not pretend here that the model (3) can embrace every kind of heterogeneity, especially because some limitations can be expected due to the fact that  $\mathcal{E}\{\mathbf{M}_s | \mathbf{M}_p, \nu\} = \mathbf{M}_p$  [38]. However, as shown in the numerical Section IV, model (3) is actually able to take into account, in an interesting way, the fact that  $\mathbf{M}_s \neq \mathbf{M}_p$ .

2) *KA Model*: The KA part of the model is described in a similar way to the heterogeneous model. Note that some authors [41] have previously proposed a Bayesian KA model involving a primary covariance matrix  $\mathbf{M}_p$  and a random prior matrix  $\bar{\mathbf{M}}_p$ . Indeed, given that  $\bar{\mathbf{M}}_p$  could be based on prior Gaussian observations, they have assumed that  $\bar{\mathbf{M}}_p | \mathbf{M}_p$  has a Wishart distribution with mean  $\mathbf{M}_p$ . Here, we have assumed that the primary covariance matrix  $\mathbf{M}_p$  is a random variable distributed around an ideal and known matrix  $\bar{\mathbf{M}}_p$  that is built upon relevant models and databases. Randomness allows the primary matrix  $\mathbf{M}_p$  to absorb some real effects or model inaccuracies not foreseen in the prior matrix  $\bar{\mathbf{M}}_p$  (e.g., calibration errors, near-field effects, inaccuracies of GPS and/or cultural databases, etc.). More precisely, the covariance matrix  $\mathbf{M}_p$  is supposed to be drawn from a complex Wishart distribution denoted by

$$\mathbf{M}_p | \bar{\mathbf{M}}_p, \mu \sim \tilde{\mathcal{W}}_{\zeta}(\mu^{-1}\bar{\mathbf{M}}_p, \mu) \quad (7)$$

with pdf

$$f(\mathbf{M}_p | \bar{\mathbf{M}}_p, \mu) = \frac{|\mathbf{M}_p|^{\mu - \zeta}}{\tilde{\Gamma}_{\zeta}(\mu) |\mu^{-1}\bar{\mathbf{M}}_p|^{\mu}} \text{etr}\{-\mu\mathbf{M}_p \bar{\mathbf{M}}_p^{-1}\}. \quad (8)$$

Note here that the prior matrix  $\bar{\mathbf{M}}_p$  is supposed to be Hermitian positive definite and stands for the whole noise (not only the ground clutter). Also, the hyperparameter  $\mu$  has to be greater than or equal to  $\zeta$ .

The distribution (7) implies that on average the primary covariance matrix is equal to the a priori

matrix, i.e.,  $\mathcal{E}\{\mathbf{M}_p | \bar{\mathbf{M}}_p, \mu\} = \bar{\mathbf{M}}_p$ , but ensures that  $\mathbf{M}_p$  differs from  $\bar{\mathbf{M}}_p$ . Besides, observing the second order moment, one notices that the hyperparameter  $\mu$  monitors the distance between  $\mathbf{M}_p$  and  $\bar{\mathbf{M}}_p$  via the relation

$$\mathcal{E}\{(\mathbf{M}_p - \bar{\mathbf{M}}_p)^2 | \bar{\mathbf{M}}_p, \mu\} = \frac{\text{Tr}\{\bar{\mathbf{M}}_p\}}{\mu} \bar{\mathbf{M}}_p. \quad (9)$$

The larger  $\mu$  is, the more accurate the prior information. Hence the hyperparameter  $\mu$  is referred to as the degree of a priori knowledge.

3) *Hyperparameters*: Thus far, we have established a KA-heterogenous model whereby two priors, (4) and (8), have been chosen among a family of conjugate priors to ensure mathematical tractability. In [38] and [33], we proposed an estimation procedure assuming that both the hyperparameters  $\nu$  and  $\mu$  introduced by (4) and (8) were known. For practical applications, these two quantities are unknown and have to be chosen carefully, as estimation performance depends on their values. Therefore, we develop here a robustified version of the algorithm of [38], assuming that  $\nu$  and  $\mu$  are random variables with noninformative priors. More precisely, we propose to consider them independent with uniform discrete pdfs, i.e.,

$$\nu | \nu_m, \nu_M \sim \mathcal{U}_{\{\nu_m, \dots, \nu_M\}} \quad (10a)$$

$$\mu | \mu_m, \mu_M \sim \mathcal{U}_{\{\mu_m, \dots, \mu_M\}} \quad (10b)$$

where  $\nu_m, \nu_M, \mu_m$ , and  $\mu_M$  are integers setting the bounds of the estimation.

We discuss here the values of the bounds  $\nu_m, \nu_M, \mu_m$ , and  $\mu_M$ . Assuming that no information is available either on the degree of heterogeneity or on the degree of a priori knowledge, one has to take care that (10a) and (10b) are noninformative priors. In other words, intervals  $\{\nu_m, \dots, \nu_M\}$  and  $\{\mu_m, \dots, \mu_M\}$  must have a good dynamic range to represent every possible degree of heterogeneity and a priori, respectively. As explained before, the values of the integers  $\nu$  and  $\mu$  have to verify  $\nu > \zeta$  and  $\mu \geq \zeta$ , respectively, and so do the lower-bounds, i.e.,  $\nu_m \geq \zeta + 1$  and  $\mu_m \geq \zeta$ . Discussing the values of  $\nu_M$  and  $\mu_M$  is a little bit more delicate. Indeed, the values of  $\nu$  and  $\mu$  are not a priori upper bounded. Yet, looking at (6) and (9), the distance between the two matrices at stake is approximately (or exactly) inversely proportional to the hyperparameter. Thus, beyond a certain value of  $\nu$  (respectively  $\mu$ ), the heterogeneity of the environment (respectively the accuracy of the prior matrix) does not change much. An opposite behavior is observed for small  $\nu$  and  $\mu$ . Hence it is crucial to adjust the minimal bounds  $\nu_m$  and  $\mu_m$  as low as possible, i.e.,  $\nu_m = \zeta + 1$  and  $\mu_m = \zeta$ . Furthermore, the heterogeneity of the environment and the accuracy of the prior knowledge are not misrepresented if one compels

$\nu$  and  $\mu$  to lie in appropriate upper-bounded intervals. We see later that the numerical choice of  $\nu_M$  and  $\mu_M$  also strongly drives the required computational resources.

For the sake of convenience, we omit in the rest of the paper constant and known quantities in the conditional terms.

## B. MMSE Estimation

1) *Monte Carlo Method*: We present here the joint MMSE estimation of the matrix  $\mathbf{M}_p$  and the hyperparameters  $\nu$  and  $\mu$  according to the model previously exposed. The notation  $\theta$  is used to designate indifferently one of these variables. By definition, the MMSE estimate of  $\theta$  is the average of the posterior distribution<sup>1</sup>

$$\hat{\theta}^{\text{mmse}} = \mathcal{E}\{\theta | \mathbf{Z}\} = \int \theta f(\theta | \mathbf{Z}) d\theta. \quad (11)$$

As explained in [39], the derivation of the integral (11) is feasible for neither  $\mathbf{M}_p$ , nor for  $\nu$  or  $\mu$ . Moreover, resorting to deterministic methods (such as numerical integration) to approximate (11) has to be prohibited when the problem dimension is greater than 4 [43], which is always the case in STAP applications. As a consequence, we propose the use of a Markov chain Monte Carlo (MCMC) method to generate samples distributed according to an appropriate target distribution and to use the samples to approximate  $\mathcal{E}\{\theta | \mathbf{Z}\}$ . More precisely, a hybrid Gibbs sampler [44] can be used to generate samples  $(\mathbf{M}_p^{(i)}, \mathbf{M}_s^{(i)}, \nu^{(i)}, \mu^{(i)})$  distributed according to the joint posterior distribution  $f(\mathbf{M}_p, \mathbf{M}_s, \nu, \mu | \mathbf{Z})$ . As described in Fig. 1, each sample  $\theta^{(i)}$  is generated according to its conditional distribution. After a burn-in period—say, equal to  $N_{bi}$  samples—the iterative process generates samples  $\theta^{(i)}$  distributed according to their posterior distribution  $f(\theta | \mathbf{Z})$  [44]. Thus, the MMSE estimates of  $\theta$  can be approximated by the empirical mean

$$\hat{\theta}^{\text{mmse}} \triangleq \frac{1}{N_r} \sum_{i=1}^{N_r} \theta^{(i+N_{bi})} \quad (12)$$

where  $N_r$  is a number of samples ensuring that the estimate will be correctly approximated. The values of the burn-in period  $N_{bi}$  and the number of samples of interest  $N_r$  are discussed later in Section IV. Note that the convergence of the procedure is obtained with whatever initial value  $\theta^{(0)}$  is supplied [44]. For instance, in Fig. 1,  $\nu^{(0)}$  and  $\mu^{(0)}$  are chosen randomly according to (10a) and (10b), and  $\mathbf{M}_p^{(0)}$  is chosen as the SCM given by

$$\hat{\mathbf{M}}_p^{\text{scm}} = \frac{1}{K} \mathbf{S}.$$

<sup>1</sup>Let us remind readers here that, so as to lighten the expressions, constant and known parameters (e.g.,  $\bar{\mathbf{M}}_p, \nu_m, \nu_M, \mu_m$ , and  $\mu_M$ ) are not mentioned in the conditional terms.

**Input:**  $\mathbf{Z}, \bar{\mathbf{M}}_p, \nu_m, \nu_M, \mu_m, \mu_M$

**Output:**  $\hat{\mathbf{M}}_p^{\text{mmse}}, \hat{\nu}^{\text{mmse}}, \hat{\mu}^{\text{mmse}}$

```

// Initialization
 $\mathbf{M}_s^{(0)} \leftarrow \hat{\mathbf{M}}_p^{\text{scm}}$ 
generate  $\nu^{(0)} \sim f(\nu)$ 
generate  $\mu^{(0)} \sim f(\mu)$ 
// Iterations
for  $i = 1$  to  $N_r + N_{bi}$  generate
   $\mathbf{M}_p^{(i)} \sim f(\mathbf{M}_p | \mathbf{M}_s^{(i-1)}, \nu^{(i-1)}, \mu^{(i-1)}, \mathbf{Z})$  /*step 1*/
   $\mathbf{M}_s^{(i)} \sim f(\mathbf{M}_s | \mathbf{M}_p^{(i)}, \nu^{(i-1)}, \mu^{(i-1)}, \mathbf{Z})$  /*step 2*/
   $\nu^{(i)} \sim f(\nu | \mathbf{M}_p^{(i)}, \mathbf{M}_s^{(i)}, \mu^{(i-1)}, \mathbf{Z})$  /*step 3*/
   $\mu^{(i)} \sim f(\mu | \mathbf{M}_p^{(i)}, \mathbf{M}_s^{(i)}, \nu^{(i)}, \mathbf{Z})$  /*step 4*/
end for
// Empirical mean
 $\hat{\mathbf{M}}_p^{\text{mmse}} = \frac{1}{N_r} \sum_{i=1}^{N_r} \mathbf{M}_p^{(i+N_{bi})}$ 
 $\hat{\nu}^{\text{mmse}} = \frac{1}{N_r} \sum_{i=1}^{N_r} \nu^{(i+N_{bi})}$ 
 $\hat{\mu}^{\text{mmse}} = \frac{1}{N_r} \sum_{i=1}^{N_r} \mu^{(i+N_{bi})}$ 

```

Fig. 1. Hybrid Gibbs sampler.

Note also that the MMSE of the secondary covariance matrix  $\mathbf{M}_s$  can be obtained as a byproduct of the proposed sampling strategy.

To derive the conditional distribution of each chain parameter, we first express the joint posterior distribution of  $\mathbf{M}_p, \mathbf{M}_s, \nu, \mu | \mathbf{Z}$ . Using the Bayes theorem with (2), (4), (8), and (10), one obtains

$$\begin{aligned}
& f(\mathbf{M}_p, \mathbf{M}_s, \nu, \mu | \mathbf{Z}) \\
& \propto |\mathbf{M}_s|^{-(\nu+\zeta+K)} |\mathbf{M}_p|^{\nu+\mu-\zeta} |\bar{\mathbf{M}}_p|^{-\mu} \\
& \quad \times \text{etr}\{-\mathbf{M} - s^{-1}[\mathbf{S} + (\nu - \zeta)\mathbf{M}_p] - \mu\bar{\mathbf{M}}_p^{-1}\mathbf{M}_p\} \\
& \quad \times \frac{(\nu - \zeta)^{\zeta\nu}}{\tilde{\Gamma}_\zeta(\nu)} \frac{\mu^{\zeta\mu}}{\tilde{\Gamma}_\zeta(\mu)} \mathbb{I}_{\{\nu_m, \nu_M\}}(\nu) \mathbb{I}_{\{\mu_m, \mu_M\}}(\mu) \quad (13)
\end{aligned}$$

where  $\mathbb{I}_{\{a,b\}}$  is the indicator function defined on the set of integers  $\{a, \dots, b\}$  and  $\propto$  means proportional to.

2) *Gibbs Moves*: We use the expression (13) to obtain the conditional distributions of  $\mathbf{M}_p | \mathbf{M}_s, \nu, \mu, \mathbf{Z}$  and  $\mathbf{M}_s | \mathbf{M}_p, \nu, \mu, \mathbf{Z}$ . Considering,  $\mathbf{M}_s, \nu$ , and  $\mu$  as given quantities in (13), it follows that

$$\begin{aligned}
& f(\mathbf{M}_p | \mathbf{M}_s, \nu, \mu, \mathbf{Z}) \\
& \propto |\mathbf{M}_p|^{\nu+\mu-\zeta} \text{etr}\{-[\mu\bar{\mathbf{M}}_p^{-1} + (\nu - \zeta)\mathbf{M} - s^{-1}]\mathbf{M}_p\}. \quad (14)
\end{aligned}$$

Similarly, one can show that

$$\begin{aligned}
& f(\mathbf{M}_s | \mathbf{M}_p, \nu, \mu, \mathbf{Z}) \\
& \propto |\mathbf{M}_s|^{-(\nu+K+\zeta)} \text{etr}\{-\mathbf{M} - s^{-1}[\mathbf{S} + (\nu - \zeta)\mathbf{M}_p]\}. \quad (15)
\end{aligned}$$

Therefore, the conditional distributions of  $\mathbf{M}_p | \mathbf{M}_s, \nu, \mu, \mathbf{Z}$  and  $\mathbf{M}_s | \mathbf{M}_p, \nu, \mu, \mathbf{Z}$  are recognized as complex

Wishart and inverse complex Wishart distributions, respectively,

$$\mathbf{M}_p | \mathbf{M}_s, \nu, \mu, \mathbf{Z} \sim \tilde{\mathcal{W}}_\zeta([\mu\bar{\mathbf{M}}_p^{-1} + (\nu - \zeta)\mathbf{M} - s^{-1}]^{-1}, \nu + \mu) \quad (16a)$$

$$\mathbf{M}_s | \mathbf{M}_p, \nu, \mu, \mathbf{Z} \sim \tilde{\mathcal{W}}_\zeta^{-1}(\mathbf{S} + (\nu - \zeta)\mathbf{M}_p, \nu + K). \quad (16b)$$

Complex Wishart and inverse complex Wishart distributions can be easily sampled to generate the matrices  $\mathbf{M}_p^{(i)}$  and  $\mathbf{M}_s^{(i)}$ . These two steps are referred to as Gibbs moves.

3) *Metropolis-Hastings Moves*: Let us consider now the generation of samples  $\nu^{(i)}$  and  $\mu^{(i)}$ . Using (13), the conditional distributions of  $\nu | \mathbf{M}_p, \mathbf{M}_s, \mu, \mathbf{Z}$  and  $\mu | \mathbf{M}_p, \mathbf{M}_s, \nu, \mathbf{Z}$  can be expressed as

$$\begin{aligned}
& f(\nu | \mathbf{M}_p, \mathbf{M}_s, \mu, \mathbf{Z}) \\
& \propto \frac{(\nu - \zeta)^{\zeta\nu}}{\tilde{\Gamma}_\zeta(\nu)} |\mathbf{M}_p \mathbf{M} - s^{-1}|^\nu \\
& \quad \times \text{etr}\{-(\nu - \zeta)\mathbf{M}_p \mathbf{M} - s^{-1}\} \mathbb{I}_{\{\nu_m, \nu_M\}}(\nu) \quad (17a)
\end{aligned}$$

$$\begin{aligned}
& f(\mu | \mathbf{M}_p, \mathbf{M}_s, \nu, \mathbf{Z}) \\
& \propto \frac{\mu^{\zeta\mu}}{\tilde{\Gamma}_\zeta(\mu)} |\mathbf{M}_p \bar{\mathbf{M}}_p^{-1}|^\mu \text{etr}\{-\mu\mathbf{M}_p \bar{\mathbf{M}}_p^{-1}\} \mathbb{I}_{\{\mu_m, \mu_M\}}(\mu). \quad (17b)
\end{aligned}$$

Unfortunately the pdfs (17) do not belong to any familiar class of distributions. Therefore, we resort to a Metropolis-Hastings (MH) move [44]. The MH algorithm is an iterative procedure that generates samples asymptotically distributed according to a given target distribution. Its principle is explained hereafter.

Let us focus, for instance, on the target distribution  $g$  (third step in Fig. 1), given by

$$g(\nu) = f(\nu | \mathbf{M}_p^{(i)}, \mathbf{M}_s^{(i)}, \mu^{(i-1)}, \mathbf{Z}).$$

The MH algorithm associated with this target distribution is described in Fig. 2 for  $N_{\text{MH}}$  iterations. At the  $j$ th iteration, a candidate  $c^{(j-1)}$  is drawn from a proposal distribution  $q(\cdot | \nu)$ . The candidate  $c^{(j-1)}$  is then accepted as the next sample  $\nu^{(j)}$  according to the acceptance probability

$$\zeta(\nu^{(j-1)}, c^{(j-1)}) = \min \left\{ 1, \frac{g(c^{(j-1)}) q(\nu^{(j-1)} | c^{(j-1)})}{g(\nu^{(j-1)}) q(c^{(j-1)} | \nu^{(j-1)})} \right\}. \quad (18)$$

The proposal distribution  $q(\cdot | \nu)$  has to be chosen properly. First, from an implementation point of view,  $q(\cdot | \nu)$  has to be easy to sample. Additionally, any proposal distribution  $q(\cdot | \nu)$  can be chosen as long as its support contains the support of the target distribution [45]. Then, two routes are usually taken. The first approach is referred to as a random-walk MH algorithm. Candidates are of the form  $c^{(j)} = \nu^{(j)}$

**Input:**  $M_p^{(i)}, M_s^{(i)}, \mathbf{Z}, \nu_m, \nu_M$

**Output:**  $\nu$  distributed according to  $f(\nu | M_p^{(i)}, M_s^{(i)}, \mu^{(i-1)}, \mathbf{Z})$

// Initialization of  $\nu^{(0)}$

// Iterations

**for**  $j = 1$  to  $N_{\text{MH}}$  **do**

// Generate a candidate

$$c^{(j-1)} \sim q(c^{(j-1)} | \nu^{(j-1)})$$

// Accept or reject the candidate

$$\nu^{(j)} = \begin{cases} c^{(j-1)} & \text{with probability } \zeta(\nu^{(j-1)}, c^{(j-1)}) \\ \nu^{(j-1)} & \text{with probability } 1 - \zeta(\nu^{(j-1)}, c^{(j-1)}) \end{cases}$$

**end for**

// Output  $\nu = \nu^{(N_{\text{MH}})}$

Fig. 2. MH algorithm (implementation of step 3).

$+\varepsilon$  where  $\varepsilon$  is a random variable with an appropriate distribution [44]. This technique is well suited for narrow target distributions. As depicted in Fig. 3, the target distribution (17) is quite narrow for small values of  $\nu$  but becomes much wider for larger values of  $\nu$ . We thus turn to the second approach based upon an independent proposal mechanism whereby the candidate does not depend on the instantaneous chain sample [44, p. 276]. More precisely, to fulfill the former requirements and in search of simplicity, we choose a uniform proposal distribution as in (10a), i.e.,

$$q(c^{(j-1)} | \nu^{(j-1)}) = q(c^{(j-1)}) \propto \mathbb{I}_{\{\nu_m, \nu_M\}}(c^{(j-1)}). \quad (19)$$

This way, the proposal distribution is able to explore the support of the target distribution equally likely. This is desirable, as it is assumed that no information is available about the true value of  $\nu$ . The MH acceptance probability (18) boils down to

$$\zeta(\nu^{(j-1)}, c^{(j-1)}) = \min \left\{ 1, \frac{g(c^{(j-1)})}{g(\nu^{(j-1)})} \right\} \quad (20)$$

meaning that the candidate is always accepted if it contributes to increasing the target distribution. Note that the convergence time of the MH algorithm will increase according to the length of the interval  $\{\nu_m, \dots, \nu_M\}$ , as the algorithm must explore the whole support of the target distribution. Recalling that we have earlier set  $\nu_m = \zeta + 1$ , this means that, from a computational point of view,  $\nu_M$  must be kept as low as possible. Similar reasonings can be made for the fourth step of the sampler leading to  $q(\mu' | \mu) \propto \mathbb{I}_{\{\mu_m, \mu_M\}}(\mu')$ .

Finally, note an interesting property of the hybrid Gibbs sampler depicted in Fig. 1. As explained before, the MH algorithm is an asymptotic procedure. Hence, several accept/reject procedures usually have to be performed before the samples become actually distributed according to the target distribution, i.e.,  $N_{\text{MH}} \gg 1$ . Conversely, the convergence of the hybrid Gibbs sampler is achieved even when steps 3 (and

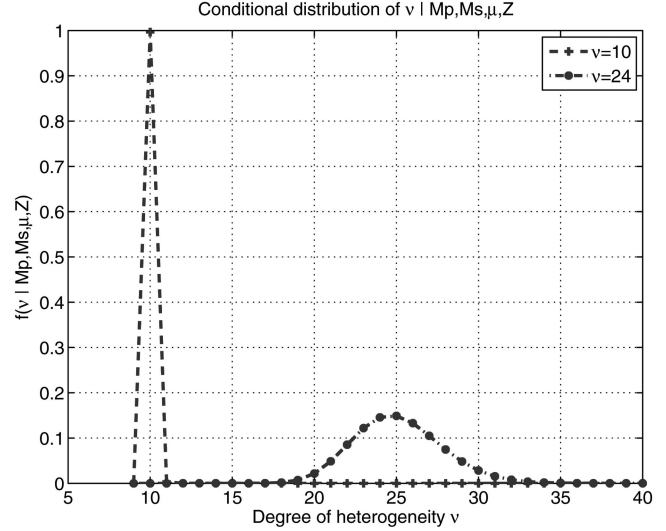


Fig. 3. Theoretical posterior conditional distribution of  $\nu | \mathbf{M}_p, \mathbf{M}_s, \mu, \mathbf{Z}$ . KA-heterogeneous model.  $K = 24$ ,  $\mu_{\min} = 9$ ,  $\mu_{\max} = 40$ .

4) are reduced to a single accept/reject step [43], i.e.,  $N_{\text{MH}}=1$ .

### III. KNOWLEDGE-AIDED ESTIMATION IN A HOMOGENOUS ENVIRONMENT

In the previous section, a KA-estimation procedure for use in a heterogenous environment has been described. One could argue against its complexity or the fact that the heterogeneity part of the model is not, strictly speaking, physical based. We intend to investigate in this section the relevance of the heterogeneity relation (3). Is it worth introducing it in the estimation scheme? To answer that question, we develop a similar estimation procedure based on a model that alleviates the heterogenous assumption. More precisely, we assume that  $\mathbf{M}_p = \mathbf{M}_s$ , and we still consider that some prior information is available and summarized into the matrix  $\bar{\mathbf{M}}_p$ . After adjusting the model of Section II to a homogeneous case, we derive the MMSE estimators of both  $\mathbf{M}_p$  and  $\mu$  that are obtained enjoyably in closed form. Note that the KA-homogeneous model presented in this section turns out to be a robustified version of the one described in [36], [37], where the parameter  $\mu$  was assumed to be known. Comparison of the KA-heterogeneous and KA-homogeneous models is performed in Section IV.

#### A. Data Model

Under the assumption of a homogeneous environment, the secondary data  $\mathbf{Z}$  share the same covariance matrix  $\mathbf{M}_p$  as the primary data  $\mathbf{z}$ , i.e.,

$$\mathbf{Z} | \mathbf{M}_p \sim \tilde{\mathcal{N}}_{\zeta, K}(\mathbf{0}, \mathbf{M}_p). \quad (21)$$

To ensure mathematical tractability, the matrix  $\mathbf{M}_p$  is supposed to be drawn from an inverse Wishart



distribution (a similar reasoning is made in [36], [37]). Indeed, this way, the distribution of  $\mathbf{M}_p | \bar{\mathbf{M}}_p, \mu$  is a conjugate prior for the likelihood  $f(\mathbf{Z} | \mathbf{M}_p)$ . To establish a certain fairness for further comparison with the KA-heterogenous model introduced earlier, the parameters of this distribution are adjusted to ensure  $\mathcal{E}\{\mathbf{M}_p | \bar{\mathbf{M}}_p, \mu\} = \bar{\mathbf{M}}_p$ , i.e.,

$$\mathbf{M}_p | \bar{\mathbf{M}}_p, \mu \sim \tilde{\mathcal{W}}_{\zeta}^{-1}((\mu - \zeta)\bar{\mathbf{M}}_p, \mu). \quad (22)$$

The pdf of  $\mathbf{M}_p | \bar{\mathbf{M}}_p, \mu$  is thus given by

$$f(\mathbf{M}_p | \bar{\mathbf{M}}_p, \mu) = \frac{|\mu - \zeta| \bar{\mathbf{M}}_p |^{\mu}}{\tilde{\Gamma}_{\zeta}(\mu) |\mathbf{M}_p|^{\mu + \zeta}} \text{etr}\{-(\mu - \zeta)\mathbf{M}_p^{-1} \bar{\mathbf{M}}_p\}. \quad (23)$$

Note also that the hyperparameter  $\mu$  can still be interpreted as the degree of a priori knowledge because we have for  $\mu > \zeta + 1$ ,

$$\mathcal{E}\{(\mathbf{M}_p - \bar{\mathbf{M}}_p)^2 | \bar{\mathbf{M}}_p, \mu\} = \frac{\bar{\mathbf{M}}_p^2 + (\mu - \zeta)\text{Tr}\{\bar{\mathbf{M}}_p\}\bar{\mathbf{M}}_p}{(\mu - \zeta + 1)(\mu - \zeta - 1)}. \quad (24)$$

The larger  $\mu$  is, the more accurate the prior information contained in the matrix  $\bar{\mathbf{M}}_p$ .

To complete and robustify the KA-homogenous model, a hierarchical level is introduced via a noninformative prior on  $\mu$ . More precisely, we keep the former distribution (10b), whereby the hyperparameter is modeled as a uniform discrete random variable on an appropriate interval  $\{\mu_m, \dots, \mu_M\}$ , i.e.,

$$\mu | \mu_m, \mu_M \sim \mathcal{U}_{\{\mu_m, \dots, \mu_M\}}.$$

Let us remind readers here that, by doing so, the prior of  $\mathbf{M}_p$ , i.e.,  $f(\mathbf{M}_p) = \int_{\mu} f(\mathbf{M}_p | \mu) f(\mu) d\mu$ , is less restrictive than a prior  $f(\mathbf{M}_p | \mu_0)$  where  $\mu$  would have been set to an arbitrary constant  $\mu_0$ . Note that the domain of the definition of  $\mu$  is now slightly different, i.e.,  $\mu > \zeta$ , and therefore  $\mu_m > \zeta$ .

## B. MMSE Estimation

According to the KA-homogeneous data model (21), (22), and (10b), we derive now the MMSE estimators of the degree of a priori knowledge  $\mu$  and the matrix  $\mathbf{M}_p$ . Using the Bayes theorem, the joint posterior distribution of  $\mu, \mathbf{M}_p | \mathbf{Z}$  can be expressed as follows:

$$f(\mu, \mathbf{M}_p | \mathbf{Z}) = \frac{1}{f(\mathbf{Z})} \frac{1}{\pi^{\zeta K}} \frac{1}{\mu_M - \mu_m + 1} \frac{|\mu - \zeta| \bar{\mathbf{M}}_p |^{\mu}}{\tilde{\Gamma}_{\zeta}(\mu) |\mathbf{M}_p|^{K + \mu + \zeta}} \times \text{etr}\{-\mathbf{M}_p^{-1}[\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p]\} \mathbb{I}_{\{\mu_m, \mu_M\}}(\mu). \quad (25)$$

We start by deriving the MMSE estimator of  $\mu$  because the posterior distribution  $f(\mu | \mathbf{Z})$  is useful in expressing the MMSE estimator of  $\mathbf{M}_p$ .

1) *MMSE Estimator of  $\mu$* : Using (25), the posterior distribution of  $\mu | \mathbf{Z}$  can be expressed as follows:

$$f(\mu | \mathbf{Z}) = \int f(\mu, \mathbf{M}_p | \mathbf{Z}) d\mathbf{M}_p \propto \frac{|\mu - \zeta| \bar{\mathbf{M}}_p |^{\mu}}{\tilde{\Gamma}_{\zeta}(\mu)} \mathbb{I}_{\{\mu_m, \mu_M\}}(\mu) \int \frac{1}{|\mathbf{M}_p|^{K + \mu + \zeta}} \times \text{etr}\{-\mathbf{M}_p^{-1}[\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p]\} d\mathbf{M}_p \quad (26)$$

$$\propto \frac{|\mu - \zeta| \bar{\mathbf{M}}_p |^{\mu}}{|\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p|^{K + \mu}} \frac{\tilde{\Gamma}_{\zeta}(K + \mu)}{\tilde{\Gamma}_{\zeta}(\mu)} \mathbb{I}_{\{\mu_m, \mu_M\}}(\mu) \quad (27)$$

where we have recognized in (26) the integral of an inverse Wishart distribution with parameter matrix  $\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p$  and  $K + \mu$  degrees of freedom. Finally, the posterior pdf of  $\mu | \mathbf{Z}$  is given by

$$f(\mu | \mathbf{Z}) = \bar{c} \frac{|\mu - \zeta| \bar{\mathbf{M}}_p |^{\mu}}{|\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p|^{K + \mu}} \frac{\tilde{\Gamma}_{\zeta}(K + \mu)}{\tilde{\Gamma}_{\zeta}(\mu)} \mathbb{I}_{\{\mu_m, \mu_M\}}(\mu) \quad (28)$$

where  $\bar{c}$  is a normalization constant such as

$$\bar{c}^{-1} = \sum_{\mu=\mu_m}^{\mu_M} \frac{|\mu - \zeta| \bar{\mathbf{M}}_p |^{\mu}}{|\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p|^{K + \mu}} \frac{\tilde{\Gamma}_{\zeta}(K + \mu)}{\tilde{\Gamma}_{\zeta}(\mu)}. \quad (29)$$

The MMSE of  $\mu$  is then obtained in closed form as the mean of the posterior distribution (28)

$$\hat{\mu}^{\text{mmse}} = \frac{\sum_{\mu=\mu_m}^{\mu_M} \mu h(\mu)}{\sum_{\mu=\mu_m}^{\mu_M} h(\mu)} \quad (30)$$

where

$$h(\mu) = \frac{|\mu - \zeta| \bar{\mathbf{M}}_p |^{\mu}}{|\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p|^{K + \mu}} \frac{\tilde{\Gamma}_{\zeta}(K + \mu)}{\tilde{\Gamma}_{\zeta}(\mu)}. \quad (31)$$

Observing (30), we can note that if  $\mu_M = \mu_m$  then  $\mu$  is a deterministic quantity and, as expected,  $\hat{\mu}^{\text{mmse}} = \mu_M = \mu_m$ .

**REMARK 1 (Implementation issue)** To avoid numerical problems, we recommend computing  $\hat{\mu}^{\text{mmse}}$  as follows:

$$\hat{\mu}^{\text{mmse}} = \frac{\sum_{\mu=\mu_m}^{\mu_M} \mu \tilde{h}(\mu)}{\sum_{\mu=\mu_m}^{\mu_M} \tilde{h}(\mu)} \quad (32)$$

where  $\tilde{h}(\mu) = \exp\{\log(h(\mu)) - \log(h(\tilde{\mu}))\}$  and  $\tilde{\mu} = \arg \max_{\mu} \log h(\mu)$ .

2) *MMSE Estimator of  $\mathbf{M}_p$* : Using (25), the conditional posterior distribution of  $\mathbf{M}_p | \mathbf{Z}, \mu$  can be expressed as

$$f(\mathbf{M}_p | \mathbf{Z}, \mu) \propto \frac{\text{etr}\{-\mathbf{M}_p^{-1}[\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p]\}}{|\mathbf{M}_p|^{K + \mu + \zeta}} \mathbb{I}_{\{\mu_m, \mu_M\}}(\mu). \quad (33)$$

We thus recognize that  $\mathbf{M}_p | \mathbf{Z}, \mu$  is distributed according to the following inverse Wishart distribution:

$$\mathbf{M}_p | \mathbf{Z}, \mu \sim \tilde{\mathcal{W}}_{\zeta}^{-1}(\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p, K + \mu). \quad (34)$$

Using the hierarchical structure of the data model, the posterior pdf of  $\mathbf{M}_p | \mathbf{Z}$  can be written as the finite sum

$$f(\mathbf{M}_p | \mathbf{Z}) = \sum_{\mu=\mu_m}^{\mu_M} f(\mathbf{M}_p | \mathbf{Z}, \mu) f(\mu | \mathbf{Z}). \quad (35)$$

Averaging the pdf (35), we obtain the MMSE estimator of  $\mathbf{M}_p$ :

$$\begin{aligned} \hat{\mathbf{M}}_p^{\text{mmse}} &= \int \mathbf{M}_p \sum_{\mu=\mu_m}^{\mu_M} f(\mathbf{M}_p | \mathbf{Z}, \mu) f(\mu | \mathbf{Z}) d\mathbf{M}_p \\ &= \sum_{\mu=\mu_m}^{\mu_M} \mathcal{E}\{\mathbf{M}_p | \mathbf{Z}, \mu\} f(\mu | \mathbf{Z}). \end{aligned} \quad (36)$$

$\mathcal{E}\{\mathbf{M}_p | \mathbf{Z}, \mu\}$  is the mean of the inverse Wishart distribution (34), i.e.,  $\mathcal{E}\{\mathbf{M}_p | \mathbf{Z}, \mu\} = [\mathbf{S} + (\mu - \zeta)\bar{\mathbf{M}}_p] / (K + \mu - \zeta)$ . The MMSE estimator of  $\mathbf{M}_p$  can finally be obtained in closed form by

$$\hat{\mathbf{M}}_p^{\text{mmse}} = w\hat{\mathbf{M}}_p^{\text{scm}} + \bar{w}\bar{\mathbf{M}}_p \quad (37)$$

where

$$w = \sum_{\mu=\mu_m}^{\mu_M} \frac{K}{K + \mu - \zeta} f(\mu | \mathbf{Z}) \quad (38a)$$

$$\bar{w} = \sum_{\mu=\mu_m}^{\mu_M} \frac{\mu - \zeta}{K + \mu - \zeta} f(\mu | \mathbf{Z}). \quad (38b)$$

To give some insight to the former expression (37), several remarks can be made.

1) The MMSE estimate of  $\mathbf{M}_p$  is equivalent to an adaptive CL technique where the weights of each matrix depend on the data through the posterior distribution  $f(\mu | \mathbf{Z})$ . More precisely, the weights of the SCM and the a priori matrix can be expressed, respectively, as

$$w = \mathcal{E} \left\{ \frac{K}{K + \mu - \zeta} \mid \mathbf{Z} \right\} \quad (39a)$$

$$\bar{w} = \mathcal{E} \left\{ \frac{\mu - \zeta}{K + \mu - \zeta} \mid \mathbf{Z} \right\}. \quad (39b)$$

2) It can be easily verified that  $w + \bar{w} = 1$ . Note that this equality has been already encountered in the literature [20, 32].

3) If the training interval  $K$  is finite, then the weight of the a priori covariance matrix is strictly positive, i.e.,  $\bar{w} > 0$  (Indeed, for  $\mu \in \{\mu_m, \dots, \mu_M\}$ , we have  $\mu > \zeta$  and  $f(\mu | \mathbf{Z}) > 0$ ). It means that  $\mathbf{M}_p$  always contributes in the MMSE estimator of  $\mathbf{M}_p$ .

4) Finally, let us consider the special case  $\mu_M = \mu_m$ , which allows us to know  $\mu$  exactly.

- a) We recover the CL estimate of [37].
- b) If the degree of a priori knowledge is as low as possible, i.e.,  $\mu_M = \mu_m = \zeta + 1$ , then  $\hat{\mathbf{M}}_p^{\text{scm}}$  contributes  $K$  times more than  $\bar{\mathbf{M}}_p$  in the estimator (37).
- c) If the degree of a priori knowledge becomes very large, i.e.,  $\mu_M = \mu_m \rightarrow +\infty$ , then the MMSE estimator of  $\bar{\mathbf{M}}_p$  is equal to the a priori covariance matrix  $\bar{\mathbf{M}}_p$ .

## IV. NUMERICAL SIMULATIONS

This section studies the performance of the MMSE estimators derived in the two last sections. We intend to show that it is essential to take into account the heterogeneity even if some prior knowledge is available. The section is divided into two parts. First, the estimators are studied independently when the data are generated according to the assumed model. Then, in the second part, their performance is analyzed using STAP synthetic data. A comparison with the CL approach proposed in [32] is also conducted. Note that our study is limited to an estimation problem.

### A. Synthetic Data

The space dimension is set to  $\zeta = 8$ , and the prior matrix is a Toeplitz matrix such that  $\bar{\mathbf{M}}_p(k, \ell) = 0.9^{|k-\ell|}$ .

1) *KA-Heterogeneous Model*: Let us consider the KA-heterogeneous model of Section II, where the data are generated according to (1), (3), and (7). We discuss first the values of some parameters and then study the estimation performance.

Adequate values for  $\nu_m$  and  $\mu_m$  have been already discussed in Section II, where  $\nu_m = \zeta + 1$  and  $\mu_m = \zeta$ . Choosing adequate values for  $\nu_M$  and  $\mu_M$  results from a trade-off between the computational load and the desired accuracy on the degrees of heterogeneity and a priori. We recall here that the value of  $\nu_M$  has to stand for a homogenous environment (i.e., small distance between  $\mathbf{M}_s$  and  $\mathbf{M}_p$ ) and that the value of  $\mu_M$  has to represent an accurate prior information (i.e., small distance between  $\mathbf{M}_p$  and  $\bar{\mathbf{M}}_p$ ). According to (6) and (9), the rule of thumb  $\nu_M = \mu_M = 5\zeta$  turns out to be a good compromise while giving satisfying results in different scenarios. It is selected in the remainder of the paper.

We discuss now the convergence assessment of the algorithm depicted in Fig. 1 which aims at setting two parameters:

$N_{bi}$ : the burn-in period required so that the outputs  $\theta^{(i)}$  of the sampler are generated according to their posterior distribution  $f(\theta | \mathbf{Z})$ ,

$N_r$ : the number of samples that ensures an accurate approximation of the MMSE estimate.

To determine these two parameters, we have adopted the methodology described in [40], where several

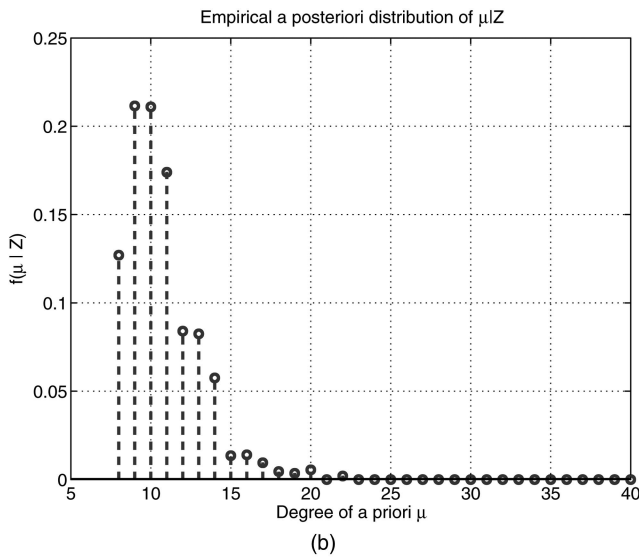
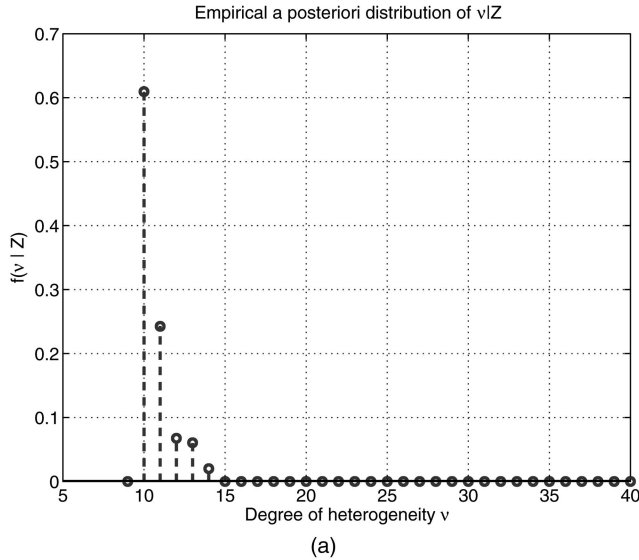


Fig. 4. Empirical posterior distribution of  $\nu | \mathbf{Z}$  and  $\mu | \mathbf{Z}$ . KA-heterogeneous model.  $\nu = 10$ ,  $\mu = 10$ ,  $K = 24$ ,  $\mu_{\min} = 9$ ,  $\mu_{\max} = 40$ .

Markov chains have to be run independently with assumed values for  $N_{b_i}$  and  $N_r$ . The appropriate values of  $N_{b_i}$  and  $N_r$  have to be chosen in order to provide an adequate potential scale reduction factor. In our case, the convergence was assessed for  $N_{b_i} = 100$  and  $N_r = 2000$  (for more details on the procedure, the reader is invited to consult the analysis conducted in [39]). We stress that the values of  $N_{b_i}$  and  $N_r$  are high and impact the computational load accordingly. But as we see later for a STAP scenario, the method, though computationally intensive, brings undeniable advantages.

Once the algorithm parameters are set to adequate values, the performance of the estimators (12) can be studied. Fig. 4 displays the empirical posterior distributions of  $\nu | \mathbf{Z}$  and  $\mu | \mathbf{Z}$  for a highly heterogeneous scenario ( $\nu = 10$ ) with inaccurate a priori ( $\mu = 10$ ). The estimated posteriors are

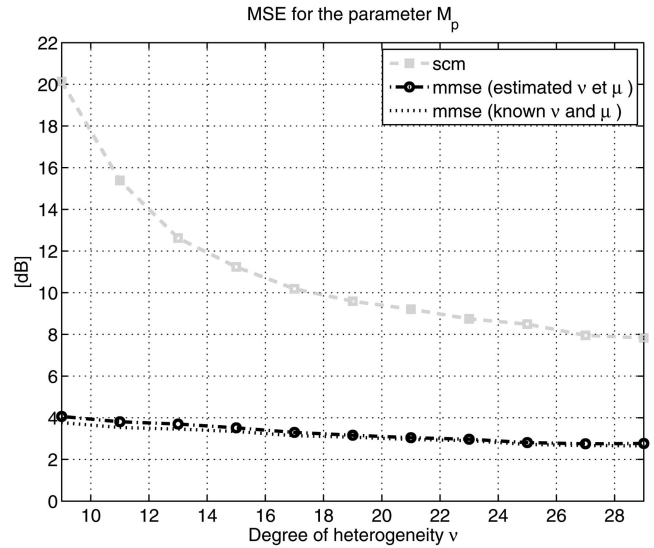


Fig. 5. MSE for parameter  $\mathbf{M}_p$ . KA-heterogeneous model.  $\mu = 20$ ,  $K = 24$ ,  $\mu_{\min} = 9$ ,  $\mu_{\max} = 40$ .

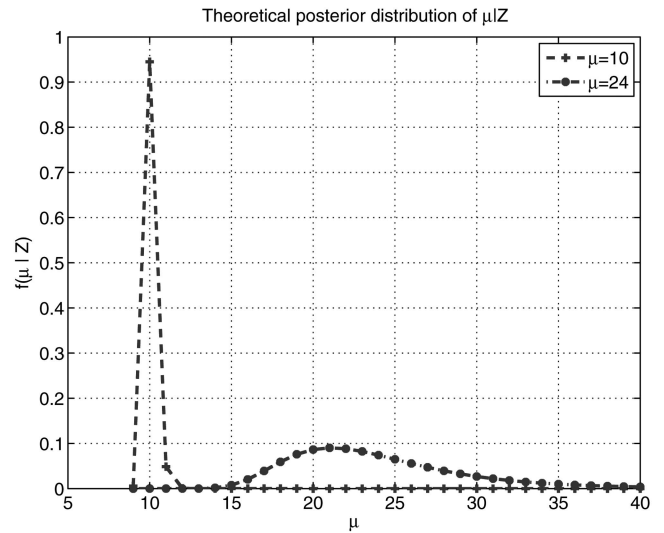


Fig. 6. Theoretical posterior distribution of  $\mu | \mathbf{Z}$ . KA-homogenous model.  $K = 24$ ,  $\mu_{\min} = 9$ ,  $\mu_{\max} = 40$ .

consistent with the true values of  $\nu$  and  $\mu$ . Fig. 5 shows the MSE for the parameter  $\mathbf{M}_p$  as a function of  $\nu$ . Three estimators of  $\mathbf{M}_p$  are compared: the MMSE estimator (12), the MMSE estimator when  $\nu$  and  $\mu$  are known [38], and also the SCM. The two MMSE estimators outperform the SCM significantly, especially when the environment is highly heterogeneous (small  $\nu$ ). We also observe that the robustified algorithm of Section II incurs small losses compared with its first version [38] that assumed known values of  $\nu$  and  $\mu$ .

2) *KA-Homogeneous Model*: This section studies the estimators (30) and (37) derived under the assumption of a KA-homogenous model. Data are generated according to (21) and (22). For comparison purposes, we set the bounds for the estimation of  $\mu$  to  $\mu_m = \zeta + 1$  and  $\mu_M = 5\zeta$ . The posterior distribution

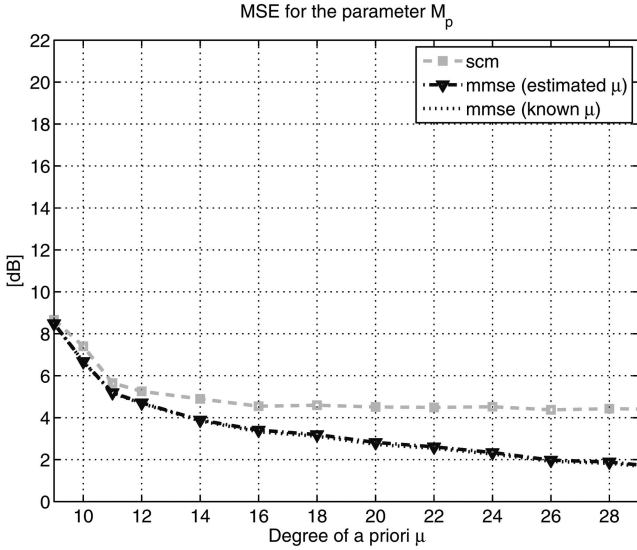


Fig. 7. MSE for parameter  $\mathbf{M}_p$ , KA-homogeneous model.  $K = 24$ ,  $\mu_{\min} = 9$ ,  $\mu_{\max} = 40$ .

$f(\mu | \mathbf{Z})$  is depicted in Fig. 6. As can be seen, the pdf is very sharp for small  $\mu$  but gets wider for larger  $\mu$ . Fig. 7 displays the MSE for the parameter  $\mathbf{M}_p$  as a function of the degree of a priori knowledge  $\mu$ . Three estimators are compared: the MMSE estimator (37), the MMSE estimator (37) with known  $\mu$  (i.e., the method of [36]), and the SCM. The robust estimator (37) performs almost as well as the MMSE estimator with known  $\mu$ . When the degree of a priori is small ( $\mu \approx \zeta + 1$ ), the MMSE estimator performs similarly to the SCM. Indeed, for small  $\mu$ ,  $\bar{\mathbf{M}}_p$  does not provide additional information on the primary covariance matrix  $\mathbf{M}_p$ , and the traditional and less complex SCM can be used instead. Yet when  $\mu$  increases, the MSE of the MMSE estimator decreases, while that of the SCM remains constant. Thus, the MMSE estimator (37) takes advantage of both the homogeneous data and the prior information.

## B. Synthetic STAP Data

1) *Scenario Parameters*: In this section, we consider an airborne radar sending a burst of  $M$  chirps at the pulse repetition frequency (PRF)  $f_r$ . The carrier frequency is denoted by  $f_0$ . The antenna is a uniform linear array (ULA) with  $N$  half-wavelength inter-spaced elements. We focus our attention on a scenario exempt from jammers. The target-free space-time snapshot  $\mathbf{z}_j$  at the  $j$ th range gate is generated as a centered complex Gaussian vector with covariance matrix  $\mathbf{M}_j$ , i.e.,

$$\mathbf{z}_j \sim \tilde{\mathcal{N}}_{\zeta}(\mathbf{0}, \mathbf{M}_j).$$

Thermal noise and ground clutter are assumed to be independent; thereby

$$\mathbf{M}_j = \mathbf{M}_{c,j} + \mathbf{M}_n \quad (40)$$

where

$\mathbf{M}_n$  is the thermal noise covariance matrix.

Assuming a mutual independence between channels and under the ULA assumption,  $\mathbf{M}_n = \sigma^2 \mathbf{I}$  with  $\sigma^2$  the noise power per element.

$\mathbf{M}_{c,j}$  is the clutter covariance matrix at the  $j$ th gate generated according to the GCM described in [1]. Succinctly, it is defined as the integral of independent clutter sources evenly distributed in azimuth

$$\mathbf{M}_{c,j} = \sigma^2 \sum_{r=1}^{N_a} \sum_{p=1}^{N_p} \xi_{p,r;j} \{ \mathbf{\Gamma}_{p,r;j} \otimes \mathbf{I}_N \} \odot \{ \mathbf{a}(p,r;j) \mathbf{a}(p,r;j)^H \} \quad (41)$$

where

$\otimes$  is the Kronecker matrix operator,

$\odot$  is the Hadamard matrix operator,

$N_a$  is the number of range ambiguities,

$N_p$  is the number of ground patches,

$\xi_{p,r;j}$  is the clutter-to-noise ratio associated with the  $(p,r)$ th patch at the  $j$ th range gate and is obtained from the radar equation as in [1, p. 23],

$\mathbf{\Gamma}_{p,r;j}$  is the intrinsic clutter motion (ICM) covariance matrix for the  $(p,r)$ th patch at the  $j$ th range gate,

$\mathbf{a}(p,r;j)$  is the space-time steering vector associated with the  $(p,r)$ th patch at the  $j$ th range gate.

The ICM stands for the pulse-to-pulse fluctuation of the clutter amplitude. In the following, the ICM is assumed to be independent of the clutter patch number, and it is modeled with a Gaussian spectrum [46], i.e.,

$$\begin{aligned} \mathbf{\Gamma}_{p,r;j} &= \text{Toeplitz}\{\gamma_c(0), \dots, \gamma_c(M-1)\} \quad \text{with} \\ \gamma_c(m) &= \exp\left\{-\frac{1}{2} \left(4\pi\sigma_v \frac{f_0}{c f_r} m\right)^2\right\} \end{aligned} \quad (42)$$

where  $c$  is the propagation velocity and  $\sigma_v$  is the velocity standard deviation. The space-time steering vector  $\mathbf{a}(p,r;j)$  can be expressed as the following Kronecker product:

$$\mathbf{a}(p,r;j) = \mathbf{a}_t(p,r;j) \otimes \mathbf{a}_s(p,r;j)$$

where  $\mathbf{a}_t(p,r;j)$  and  $\mathbf{a}_s(p,r;j)$  are, respectively, the  $M$ -length temporal steering vector and the  $N$ -length spatial steering vector associated with the  $(p,r)$ th patch at the  $j$ th range gate, i.e., for  $m = 0, \dots, M-1$  and  $n = 0, \dots, N-1$ ,

$$\begin{aligned} [\mathbf{a}_t(p,r;j)]_m &= \exp\left(j2\pi m \frac{2v}{c} \frac{f_0}{f_r} \cos(\theta_{p,r;j}) \sin(\phi_{p,r;j} + \phi_a)\right) \\ [\mathbf{a}_s(p,r;j)]_n &= \exp\left(j2\pi n \frac{d}{\lambda_0} \cos(\theta_{p,r;j}) \sin(\phi_{p,r;j})\right) \end{aligned}$$

where

$d$  is the interelement distance,

$\lambda_0 = c/f_0$  is the wavelength,

$v$  is the radar platform velocity,

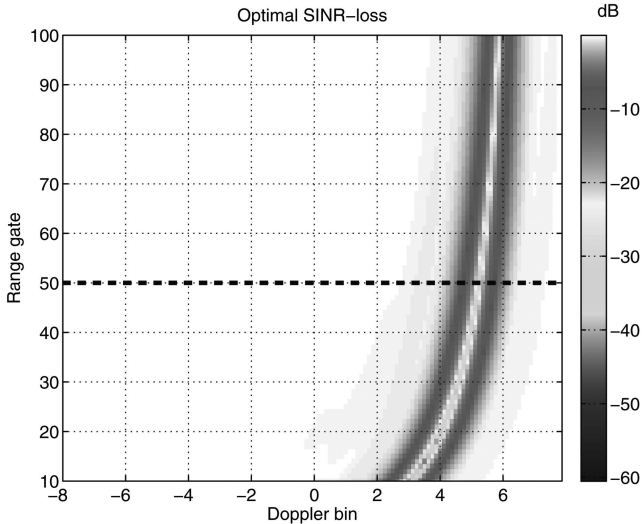


Fig. 8. Range-Doppler map: optimal SINR-loss.

TABLE I  
Scenario Parameters

Parameter	Value
velocity	$v = 100$ m/s
crab	$\phi_a = 90^\circ$
carrier	$f_0 = 450$ MHz
PRF	$f_r = 750$ Hz
pulses	$M = 16$
channel	$N = 8$
interelement distance	$d = \lambda_0/2$
element pattern	cosine
backlobe level	$b_e = -30$ dB
CNR <sup>1</sup> (range #1)	CNR = 40 dB
ICM	$\sigma_v = 0.25$ m/s

Note: <sup>1</sup>Clutter-to-noise ratio.

$(\phi_{p,r;j}, \theta_{p,r;j})$  is the azimuth and depression angle of the  $(p,r)$ th clutter patch at the  $j$ th range gate,

$\phi_a$  is the misalignment angle between the velocity vector and the lengthwise direction of the ULA. Hereafter, we assume a scenario without range ambiguities, i.e.,  $N_a = 1$ . A forward-looking antenna (i.e.,  $\phi_a = 90^\circ$ ) is considered to obtain a nonstationary clutter [1, 2]. Such geometrical configuration ensures, indeed, that the spatio-Doppler frequencies occupied by the clutter vary over range as depicted in Fig. 8 (the SINR loss is defined later). We focus the analysis on the range gate 50 where the clutter is highly nonstationary. Scenario parameters useful to generate  $\mathbf{M}_{c,j}$  according to (41) are given in Table I.

2) *Processing Parameters*: We apply the estimation procedures within a reduced-dimension algorithm with a post-Doppler structure. Indeed, it allows one to apply the estimation procedure locally over  $M' \ll M$  Doppler bins. Moreover, it significantly decreases the computational load. More precisely, an extended factored algorithm (EFA) has been chosen with  $M' = 3$  Doppler bins and without zero-padding

[1, 47]. Thus, for a given range gate, the estimation procedure is applied  $M$  times (one time per bin) with  $\zeta = NM' = 24$ .

Only one prior matrix  $\bar{\mathbf{M}}_p$  of dimension  $NM \times NM$  is generated and then projected onto the appropriate reduced Doppler-element domain. The a priori covariance matrix is built in a similar way to (40), i.e.,

$$\bar{\mathbf{M}}_p = \bar{\mathbf{M}}_n + \bar{\mathbf{M}}_{c,50} \quad (43)$$

where  $\bar{\mathbf{M}}_n$  and  $\bar{\mathbf{M}}_{c,50}$  are the noise and clutter components, respectively. The noise component  $\bar{\mathbf{M}}_n$  is supposed to be known exactly, i.e.,  $\bar{\mathbf{M}}_n = \mathbf{M}_n$ . The clutter component  $\bar{\mathbf{M}}_{c,50}$  is built upon the GCM model according to (41) but departs from the true matrix  $\mathbf{M}_{c,50}$ . Indeed, three parameters of the GCM have been changed as follows:

- 1) the ICM has been overestimated with the value  $\sigma_v = 0.5$  m/s,
- 2) a misalignment of  $\Delta\phi_a = 2^\circ$  has been introduced between the velocity vector and the longitudinal axis of the fuselage,
- 3) the backlobe level of the element pattern has been increased to  $b_e = -5$  dB.

The first two modifications represent errors that are common in real scenarios. Although the third modification is less realistic, its impact is localized in the Doppler domain and hence, as shown later, it allows us to illustrate the performance of our algorithms in a case where the environment is homogeneous and the a priori highly inaccurate. Finally, the number of secondary data is set to  $K = 2\zeta = 48$ . The bounds for hyperparameter estimation are set to

$\nu_m = \zeta + 1 = 25$ ,  $\mu_m = \zeta = 24$ ,  $\nu_M = \mu_M = 5\zeta = 120$  for the KA-heterogeneous model,

$\mu_m = \zeta + 1 = 25$ ,  $\mu_M = 5\zeta = 120$  for the KA-homogeneous model.

3) *Results*: We observe in this section the STAP filter shape for different estimates  $\hat{\mathbf{M}}_p$  of the noise covariance matrix. To do so, we display the SINR loss defined as [2]

$$\mathcal{L} = \frac{|\mathbf{w}^H \mathbf{a}|^2}{\mathbf{w}^H \hat{\mathbf{M}}_p \mathbf{w}} \frac{1}{\mathbf{a}^H \hat{\mathbf{M}}_n^{-1} \mathbf{a}} \quad (44)$$

where  $\mathbf{a}$  is the steering vector and  $\mathbf{w} \propto \hat{\mathbf{M}}_p^{-1} \mathbf{a}$  is the STAP filter. Results are averaged using  $M_c = 100$  Monte Carlo runs. Additionally for each Doppler bin, we observe the average values of the hyperparameters  $\nu$  and  $\mu$  as well as the average values of the weighting factors  $w$  and  $\bar{w}$ , given by (38). Results are displayed in Fig. 9.

Before analyzing the behavior of the MMSE estimators, we recall first that the optimal filter in an EFA structure is obtained by the clairvoyant case, i.e.,  $\hat{\mathbf{M}}_p = \mathbf{M}_p$ . Note also the expected behavior of the SCM filter that performs badly around the clutter

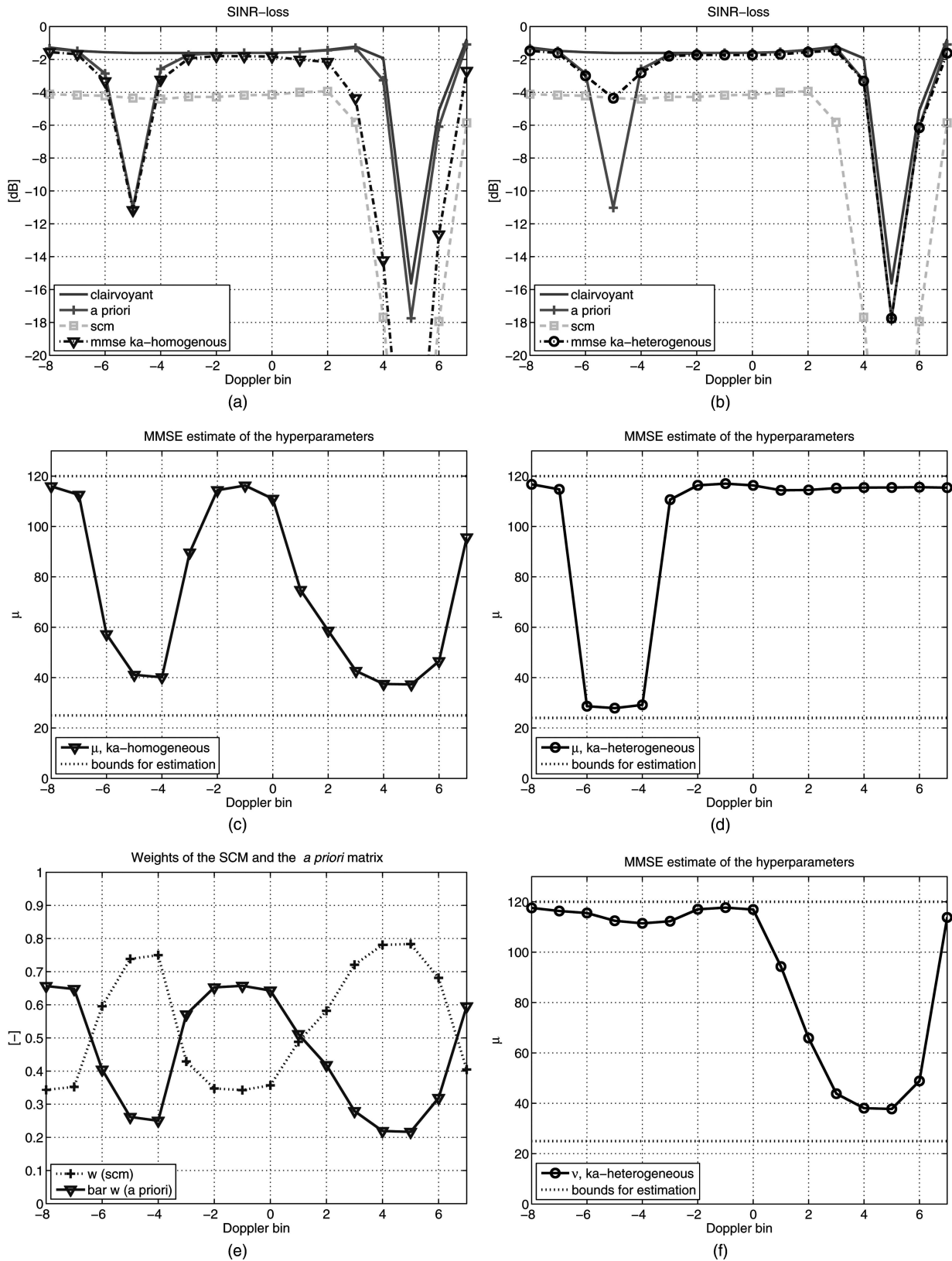


Fig. 9. KA-estimation assuming homogenous environment: SINR loss (a), estimated  $\mu$  (c), weights  $w$  and  $\bar{w}$  (e). KA-estimation assuming heterogeneous environment: SINR loss (b), estimated  $\mu$  (d), estimated  $\nu$  (f).

notch (fifth Doppler bin), as the clutter is highly nonstationary. Elsewhere, the SCM filter incurs a 3dB loss (compared with the optimal filtering in an

EFA structure) as predicted by the Reed, Mallett, and Brennan rule [5]. Finally, we emphasize that the forward-looking scenario and the a priori matrix have

TABLE II  
Distinct Domains of Environment and a priori

Domain	Bins	Environment	a priori
$\mathcal{D}_1$	$\{-8, -7\} \cap \{-3, \dots, 2\}$	homogeneous	accurate
$\mathcal{D}_2$	$\{-6, \dots, -4\}$	homogeneous	inaccurate
$\mathcal{D}_3$	$\{3, \dots, 6\}$	heterogeneous	accurate

been chosen so that three different areas can be clearly identified, as summarized in Table II:

1) Observing Fig. 8 and according to the training interval, it is clear that the clutter is highly nonstationary from bin 3 to 6. Elsewhere the secondary data are homogenous with the primary one.

2) Observing now the filter shape of the prior matrix in Fig. 9, one can see that the a priori information is rather accurate except from bin  $-6$  to  $-4$ . A slight ICM overestimation (as suggested in [21]) and a small crab error do not significantly affect the mainlobe clutter notch other than a slight widening and deepening. Note also that the backlobe level, which has been increased to build  $\bar{\mathbf{M}}_p$ , impacts only bin  $-6$  to  $-4$ .

Of course, some specific algorithms (e.g., the extended sample matrix inversion [48]) are efficient at counteracting the nonstationarity of the clutter considered here (especially when the nonstationarity is so continuous). However, the simplicity of the scenario allows us to characterize at once the behavior of the studied estimators for three distinct cases as described in Table II. The contrast is not so distinct with more realistic clutter. In the following, we comment on Figs. 9 and 10, which depict the performance of the MMSE estimators of Sections II and III and also, for comparison purposes, the CL estimator presented in [32].

Observing Figs. 9(a), (c), and (e), the following comments can be made on the MMSE estimators based on a KA-homogeneous model.

1) The filter built with the estimator (37) does not perform well in the nonstationary domain  $\mathcal{D}_3$  or in  $\mathcal{D}_2$  where the a priori information is not accurate. Close to optimal performance is observed in domain  $\mathcal{D}_1$  where the environment is actually homogeneous and the a priori matrix is accurate.

2) The estimated degree of a priori  $\mu$  (30) is large in the domain  $\mathcal{D}_1$  but has small values elsewhere in  $\mathcal{D}_2$  and  $\mathcal{D}_3$ . Thus, the estimator fails at recognizing the domain  $\mathcal{D}_3$ . Note that  $\mu$  is closely related to the weight  $\bar{w}$  of the a priori matrix.

This behavior can be explained as follows.

1) In domain  $\mathcal{D}_3$ , the data do not respect the homogeneous assumption used to derive (30) and (37). This leads to underestimation of  $\mu$  and  $\bar{w}$ . Thus

the estimator (37) does not give enough weight to the a priori matrix.

2) In domain  $\mathcal{D}_2$ , the estimation of  $\mu$  seems adequate because the a priori information is not accurate. However, the filter based on (37) does not perform much better than the a priori filter. Indeed the weight  $\bar{w}$ , though small ( $\bar{w} \ll 1$ ), is not equal to zero, and thus the a priori matrix cannot be totally rejected in the estimate. In other words, the number of secondary data  $K$  is not large enough so that the SCM outweighs the prior matrix. This behavior is very undesirable and contrasts the results obtained for the MSE of  $\bar{\mathbf{M}}_p$  (37) when the data are generated according to the KA-homogeneous model.

The performance of the MMSE estimators based on the assumption of a KA-heterogeneous model are depicted in Figs. 9(b), (d), and (f). The following points can be underlined.

1) The proposed filter always provides the best performance among the three adaptive filters under study. It behaves like the a priori filter in  $\mathcal{D}_1$  and  $\mathcal{D}_3$  when the later outperforms the SCM filter. Otherwise, the filter behaves like the SCM filter.

2) The values of the estimated  $\mu$  are large in  $\mathcal{D}_1$  and  $\mathcal{D}_3$ , i.e.,  $\hat{\mu}^{\text{mmse}} \approx \mu_M$ , when the a priori information is accurate. In domain  $\mathcal{D}_2$ , where the a priori is not precise, the estimated values of  $\nu$  are low, i.e.,  $\hat{\nu}^{\text{mmse}} \approx \nu_m$ . Thus the estimate of  $\mu$  identifies correctly and precisely whether the a priori information is reliable or not.

3) The estimated values of the degree of heterogeneity  $\nu$  are more contrasted. In domain  $\mathcal{D}_2$ , the values of  $\nu$  are small ( $\hat{\nu}^{\text{mmse}} \approx \nu_m$ ). Thus  $\mathcal{D}_2$  is correctly identified as a nonhomogeneous environment. In domains  $\mathcal{D}_1$  and  $\mathcal{D}_3$ , the values of  $\nu$  are large but endure some variations, though the environment is equally homogeneous over the area. Especially around the fourth Doppler bin, the value of  $\nu$  decreases where the degree of a priori is small.

Thus the estimate of  $\mu$  is a very important quantity. It defines the shape of the filter: if  $\mu$  is large the filter behaves like the a priori filter; if  $\mu$  is small the filter behaves like the SCM filter. It also correctly identifies if the a priori matrix brings an accurate information and incorporates or discards it accordingly in the covariance matrix estimate. As expected the heterogeneous model (3) is perfectible. Indeed the filter does not outperform the SCM filter if the prior matrix is not accurate. However, the model (3) gives some kind of degree of freedom between the prior knowledge and the secondary data that allows one an appropriate incorporation of the prior matrix. Moreover, the estimate of  $\nu$  provides additional information. It identifies correctly the heterogeneity

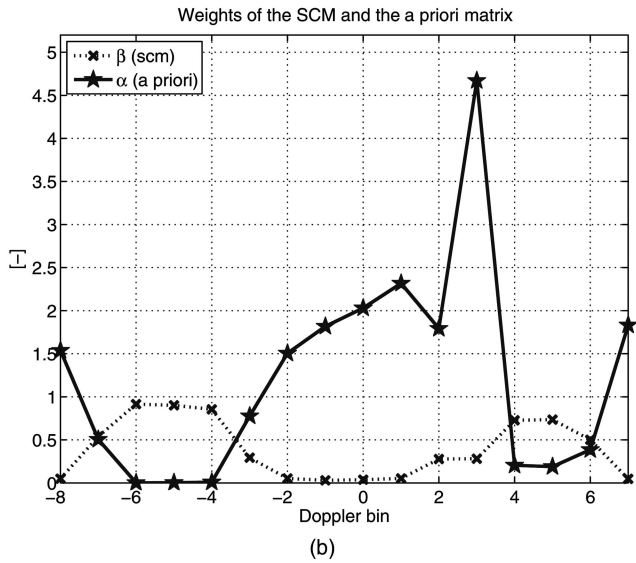
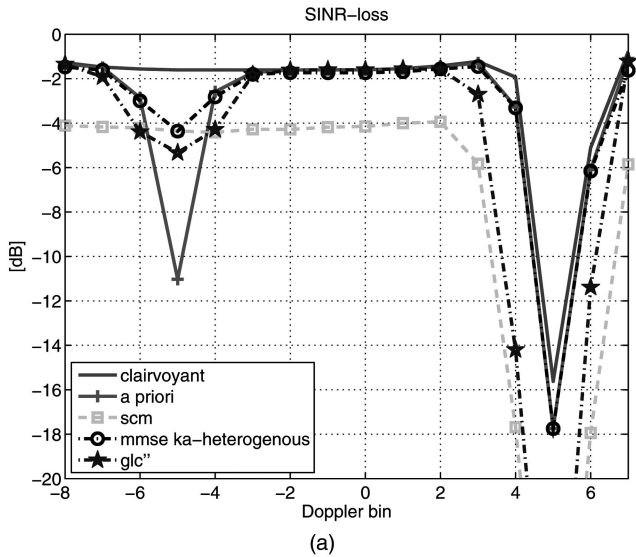


Fig. 10. GLC-CL estimation: SINR loss (a), weights  $\alpha$  and  $\beta$  (b).

as long as the prior is accurate enough (otherwise the algorithm seems to suffer from a lack of reference).

The behaviors of the two estimation procedures are summarized in Table III. Of course the Gibbs sampling strategy is highly computationally intensive. Moreover, the interpretation of its estimators is made harder because they are not obtained in closed form (12). However, unlike the estimation based on a KA-homogenous model, the algorithm is able to identify precisely the accuracy of the prior matrix whether the environment is heterogeneous or not. Furthermore, the matrix estimate derived under the heterogeneous assumption can entirely reject the a priori matrix, which is a desirable property, especially when this matrix does not bring trustworthy information. In light of these results, it is clear that using the heterogeneity relation (3) of the KA-heterogeneous model is actually relevant.

TABLE III  
Summary of the Performance of the MMSE Estimates on STAP Synthetic Data

Data Type	Environment Identification $\nu$		a priori Identification $\mu$		a priori Incorporation	
	KA-H <sup>1</sup>	KA-NH <sup>2</sup>	KA-H	KA-NH	KA-H	KA-NH
$\mathcal{D}_1$	$\emptyset$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
$\mathcal{D}_2$	$\emptyset$	$\approx$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$
$\mathcal{D}_3$	$\emptyset$	$\checkmark$	$\times$	$\checkmark$	$\times$	$\checkmark$

Note: <sup>1,2</sup>KA-homogenous model, KA-heterogeneous model.  $\emptyset$ ,  $\approx$ ,  $\checkmark$ ,  $\times$  none, approximate, correct, not correct.

Finally, before closing this section, the two Bayesian approaches are compared with the deterministic CL technique presented recently in [32]. More precisely, we have considered the general linear combination (GLC) technique. The covariance matrix estimate is supposed to be a linear combination of the SCM and the prior matrix

$$\hat{\mathbf{M}}_p^{\text{glc}} = \beta \hat{\mathbf{M}}_p^{\text{scm}} + \alpha \bar{\mathbf{M}}_p \quad (45)$$

with  $\alpha > 0$  and  $\beta > 0$ . The weighting factors  $\alpha$  and  $\beta$  are then derived so as to minimize the MSE between  $\hat{\mathbf{M}}_p^{\text{glc}}$  and the clairvoyant covariance matrix  $\mathbf{M}_p$ . More precisely, as  $\mathbf{M}_p$  is not known a priori, an estimate of the MSE is considered instead. The latest is obtained, assuming that the SCM is unbiased. Fig. 10 displays the STAP filter shape and the GLC'' weighting factors.

1) In domain  $\mathcal{D}_1$  and  $\mathcal{D}_3$ , the GLC filter behaves like the filter based on the KA-homogenous model. Near-optimal performance is obtained in  $\mathcal{D}_1$  where the environment is homogeneous and the prior matrix is accurate. In  $\mathcal{D}_3$ , the heterogeneity prevents an appropriate estimation of the weighting factors.

2) Interestingly, in domain  $\mathcal{D}_2$ , the GLC method is able to reject almost entirely the prior information that is unprecise.

Thus the deterministic CL estimate (45) performs better than the Bayesian CL estimate (37). However, the Bayesian CL technique is able to identify the degree of a priori when the environment is homogeneous. This information is not provided as clearly by the GLC technique. In any event, the Bayesian estimators, based on the assumption of a KA-heterogenous model, provide better filtering and also important information regarding the degrees of a priori and heterogeneity.

## V. CONCLUSIONS

In KA-STAP, a priori information is used to improve the detection performance in highly heterogeneous environments. In this context, we have presented two estimation schemes designed to incorporate a prior matrix in the noise covariance matrix estimate. Both schemes rely on Bayesian



data models. The first model also entails an original relation of heterogeneity that describes how the secondary covariance matrix differs from the primary one. The second model assumes a homogenous environment. The KA part and (possibly) the heterogeneity relation of the models involve hyperparameters that represent the degrees of a priori and heterogeneity, respectively. The MMSE estimators of both the hyperparameters and the covariance matrix are derived. For the KA-heterogeneous model, the estimation is performed via a Gibbs sampling strategy that is highly computationally intensive. For the KA-homogenous model, the MMSE estimators are obtained in closed form, and the algorithm turns out to belong to the well-known CL technique. Performance analysis on STAP synthetic data shows that it is essential to take into account the heterogeneity in the data model. It brings a degree of freedom between the secondary data and the prior matrix that allows one to identify and incorporate the prior information in an appropriate way. Otherwise the proposed estimators are not able to reject inaccurate information, yet achievable with a less complex deterministic CL method. Additionally, the estimation scheme based on the first model yields precise information on the degree of heterogeneity if the prior information is accurate. Finally, one can think of some refinements for the future. As expected, the heterogenous model might be improved. Indeed when the a priori information is not reliable, the KA-STAP filter does not perform better than the SCM filter. Also the computational load induced by the MCMC method is currently a drawback for real-time processing.

#### REFERENCES

- [1] Ward, J.  
Space-time adaptive processing for airborne radar.  
Lincoln Laboratory, MIT, Lexington, MA, Technical Report 1015, Dec. 1994.
- [2] Klemm, R.  
*Principles of Space-Time Adaptive Processing*.  
London: IEE Press, 2002.
- [3] Guerci, J. R.  
*Space-Time Adaptive Processing for Radar*.  
Norwood, MA: Artech House, 2003.
- [4] Brennan, L. E. and Reed, I. S.  
Theory of adaptive radar.  
*IEEE Transactions on Aerospace and Electronic Systems*, **9**, 2 (Mar. 1973), 237–252.
- [5] Reed, I. S., Mallett, J. D., and Brennan, L. E.  
Rapid convergence rate in adaptive arrays.  
*IEEE Transactions on Aerospace and Electronic Systems*, **AES-10**, 6 (Nov. 1974), 853–863.
- [6] Melvin, W. L.  
STAP in heterogeneous clutter environments.  
In R. Klemm, (Ed.), *Applications of Space-Time Adaptive Processing*, 2004, ch. 10, 305–358.
- [7] McDonald, K. F. and Blum, R. S.  
Robust techniques in space-time adaptive processing.  
In R. Klemm, (Ed.), *Applications of Space-Time Adaptive Processing*, 2004, ch. 13, 413–461.
- [8] Melvin, W. L.  
Space-time adaptive radar performance in heterogeneous clutter.  
*IEEE Transactions on Aerospace and Electronic Systems*, **36**, 2 (Apr. 2000), 621–633.
- [9] Guerci, J. R., Goldstein, J. S., and Reed, I. S.  
Optimal and adaptive reduced-rank STAP.  
*IEEE Transactions on Aerospace and Electronic Systems*, **36**, 2 (Apr. 2000), 647–663.
- [10] Carlson, B. D.  
Covariance matrix estimation errors and diagonal loading in adaptive arrays.  
*IEEE Transactions on Aerospace and Electronic Systems*, **24**, 4 (July 1988), 397–401.
- [11] Steiner, M. and Gerlach, K.  
Fast converging adaptive processor for a structured covariance matrix.  
*IEEE Transactions on Aerospace and Electronic Systems*, **36**, 4 (Oct. 2000), 1115–1126.
- [12] Fuhrmann, D. R.  
Application of Toeplitz estimation to adaptive beamforming and detection.  
*IEEE Transactions on Signal Processing*, **39**, 10 (Oct. 1991), 2194–2198.
- [13] Rabideau, D. J. and Steinhardt, A. O.  
Improved adaptive clutter cancellation through data-adaptive training.  
*IEEE Transactions on Aerospace and Electronic Systems*, **35**, 3 (July 1999), 879–891.
- [14] Chen, P., Melvin, W. L., and Wicks, M. C.  
Screening among multivariate normal data.  
*Journal of Multivariate Analysis*, **69**, 1 (Apr. 1999), 10–29.
- [15] Scharf, L. L. and McWhorter, L. T.  
Adaptive matched subspace detectors and adaptive coherence estimators.  
In *Proceedings of the 30th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1996, 1114–1117.
- [16] Pascal, F., et al.  
Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis.  
*IEEE Transactions on Signal Processing*, **56**, 1 (Jan. 2008), 34–48.
- [17] Gini, F. and Rangaswamy, M.  
*Knowledge-Based Radar Detection, Tracking and Classification*.  
Malden, MA: Wiley InterScience, 2008.
- [18] Gini, F., Goldstein, J. S. and Zoltowski, M. D. (Eds.)  
Knowledge-based systems for adaptive radar: Detection, tracking and classification.  
*IEEE Signal Processing Magazine*, **23**, 1 (Jan. 2006), special section.
- [19] Proceedings KASSPER Workshops, 2002–2005  
[Online], formerly available:  
<http://www.darpa.mil/sto/space/kassper.html> (accessed Mar. 2008).
- [20] Guerci, J. R. and Baranoski, E. J.  
Knowledge-aided adaptive radar at DARPA.  
*IEEE Signal Processing Magazine*, **23**, 1 (Jan. 2006), 41–50.
- [21] Melvin, W. L. and Guerci, J. R.  
Knowledge-aided signal processing: A new paradigm for radar and other advanced sensors.  
*IEEE Transactions on Aerospace and Electronic Systems*, **42**, 3 (July 2006), 983–996.
- [22] Capraro, C. T., et al.  
Implementing digital terrain data in knowledge-aided space-time adaptive processing.  
*IEEE Transactions on Aerospace and Electronic Systems*, **42**, 3 (July 2006).

- [23] Conte, E., et al.  
Design and analysis of a knowledge-aided radar detector for Doppler processing.  
*IEEE Transactions on Aerospace and Electronic Systems*, **42**, 3 (July 2006), 1058–1079.
- [24] Gerlach, K. and Picciolo, M. L.  
Airborne/spacebased radar STAP using a structured covariance matrix.  
*IEEE Transactions on Aerospace and Electronic Systems*, **39**, 1 (Jan. 2003), 269–281.
- [25] Melvin, W. L. and Showman, G. A.  
An approach to knowledge-aided covariance estimation.  
*IEEE Transactions on Aerospace and Electronic Systems*, **42**, 3 (July 2006), 1021–1042.
- [26] Hiemstra, J. D.  
Colored diagonal loading.  
*In Proceedings of the IEEE Radar Conference*, Chantilly, VA, Apr. 22–25, 2002, 386–390.
- [27] Teixeira, C., Bergin, J., and Techau, P.  
Reduced degree-of-freedom STAP with knowledge-aided data pre-whitening.  
*In Proceedings of KASSPER Workshops 2003*, Las Vegas, NV, Apr. 2003.
- [28] Bergin, J. S., et al.  
STAP with knowledge-aided data pre-whitening.  
*In Proceedings of the IEEE Radar Conference*, Philadelphia, PA, Apr. 26–29, 2004, 289–294.
- [29] Page, D., et al.  
Improving knowledge-aided STAP performance using past CPI data.  
*In Proceedings of the IEEE Radar Conference*, Apr. 26–29, 2004, 295–300.
- [30] Blunt, S. D., Gerlach, K., and Rangaswamy, M.  
STAP using knowledge-aided covariance matrix estimation and the FRACTA algorithm.  
*IEEE Transactions on Aerospace and Electronic Systems*, **42**, 3 (July 2006), 1043–1057.
- [31] Zhu, X., et al.  
Knowledge-aided space-time adaptive processing.  
*In Proceedings of the 41st Asilomar Conference*, Pacific Grove, CA, Nov. 4–7, 2007, 1830–1834.
- [32] Stoica, P., et al.  
On using a priori knowledge in space-time adaptive processing.  
*IEEE Transactions on Signal Processing*, **56**, 6 (June 2008), 2598–2602.
- [33] Besson, O., Bidon, S., and Tournet, J-Y.  
Covariance matrix estimation with heterogeneous samples.  
*IEEE Transactions Signal Processing*, **56**, 3 (Mar. 2008), 909–920.
- [34] Bergin, J., et al.  
Space-time beamforming with knowledge-aided constraints.  
*In Proceedings of the Adaptive Sensor Array Processing Workshop*, Mar. 11–13, 2003.
- [35] Besson, O., Tournet, J-Y., and Bidon, S.  
Knowledge-aided Bayesian detection in heterogeneous environments.  
*IEEE Signal Processing Letters*, **14**, 5 (May 2007), 355–358.
- [36] De Maio, A. and Farina, A.  
Adaptive radar detection: A Bayesian approach.  
*In Proceedings of the 2006 International Radar Symposium*, Krakow, Poland, May 24–26, 2006.
- [37] De Maio, A., Farina, A., and Foglia, G.  
Adaptive radar detection: A Bayesian approach.  
*In Proceedings of the IEEE Radar Conference*, Waltham, MA, Apr. 17–20, 2007, 624–629.
- [38] Bidon, S., Besson, O., and Tournet, J-Y.  
A Bayesian approach to adaptive detection in nonhomogeneous environments.  
*IEEE Transactions on Signal Processing*, **56**, 1 (Jan. 2008), 205–217.
- [39] Bidon, S., Besson, O., and Tournet, J-Y.  
Characterization of clutter heterogeneity and estimation of its covariance matrix.  
*In Proceedings of the 2008 IEEE Radar Conference*, Rome, Italy, May 26–30, 2008, 357–362.
- [40] Gelman, A., et al.  
*Bayesian Data Analysis* (2nd ed.).  
New York: Chapman & Hall, 2004.
- [41] Anderson, T. W.  
*An Introduction to Multivariate Statistical Analysis* (2nd ed.).  
Hoboken, NJ: Wiley, 1984.
- [42] Earp, S. L. and Nolte, L. W.  
Multichannel adaptive array processing for optimal detection.  
*In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, San Diego, CA, Mar. 1984, 267–270.
- [43] Robert, C. P.  
*The Bayesian Choice, From Decision-Theoretic Foundations to Computational Implementation*.  
New York: Springer, 2007.
- [44] Robert, C. P. and Casella, G.  
*Monte Carlo Statistical Methods* (Springer Texts in Statistics).  
New York: Springer, 2004.
- [45] Robert, C. P.  
*Discretization and MCMC Convergence Assessment*.  
Berlin: Springer-Verlag, 1998.
- [46] Barton, D. K.  
*Radar Clutter*.  
Norwood, MA: Artech House, 1975.
- [47] DiPietro, R. C.  
Extended factored space-time processing for airborne radar systems.  
*In Proceedings of the 26th Asilomar Conference on Signals, Systems, and Computers*, vol. 1, Pacific Grove, CA, Oct. 26–28, 1992, 425–430.
- [48] Hayward, S. D.  
Adaptive beamforming for rapidly moving arrays.  
*In Proceedings of the International Conference on Radar*, Beijing, China, Oct. 8–10, 1996, 480–483.



**Stéphanie Bidon** (M'08) received the engineer and master's 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.



**Olivier Besson** (SM'04) received the Ph.D. degree in signal processing in 1992 and the Habilitation à Diriger des Recherches in 1998, both from INP Toulouse.

He is currently affiliated with the Department of Electronics, Optronics, and Signal of ISAE (Institut Supérieur de l'Aéronautique et de l'Espace). He has held visiting positions at Brigham Young University, Provo, UT, and Università del Salento, Lecce, Italy. He is the co-recipient of the Eurasp Best Paper Award (1998) and the IEE Clarke-Griffiths Premium Award (1998).

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.



**Jean-Yves Tourneret** (SM'08) received the ingénieur degree in electrical engineering from École Nationale Supérieure d'Électronique, d'Électrotechnique, d'Informatique et d'Hydraulique of Toulouse (ENSEEIH) in 1989 and the Ph.D. degree from the National Polytechnic Institute of Toulouse in 1992.

He is currently a professor at the University of Toulouse (ENSEEIH), France, and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal processing with a particular interest in Markov chain Monte Carlo methods.

Dr. Tourneret was the program chair of the European Conference on Signal Processing (EUSIPCO), which was held in Toulouse, France in 2002. He was also member of the organizing committee for the International Conference ICASSP'06 which was held in Toulouse (France) in 2006. He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001–2007, 2010). He is currently serving as an associate editor for the *IEEE Transactions on Signal Processing*.