Bayesian Multi-Class Covariance Matrix Filtering for Adaptive Environment Learning

Paolo Braca*, Augusto Aubry†, Leonardo M. Millefiori*, Antonio De Maio† and Stefano Marano‡
*NATO STO Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy
†Università degli Studi di Napoli "Federico II", DIETI, Napoli, Italy
‡Università degli Studi di Salerno, DIEM, Fisciano, Italy

***Abstract*—Covariance matrix estimation is a crucial task in adaptive signal processing applied to several surveillance systems, including radar and sonar. In this paper we propose a dynamic environment learning strategy to track both the covariance matrix and its class; the class represents a set of structured covariance matrices. We assume that the posterior distribution of the covariance given the class, is basically a mixture of inverse Wishart, while the class posterior distribution evolves according to a Markov chain. The proposed multi-class inverse Wishart mixture filter is shown to outperform the class-clairvoyant maximum likelihood estimator in terms of covariance estimate accuracy, as well as the Bayesian information criterion rule in terms of classification performance.**

***Index Terms*—Model classification, adaptive filter, covariance estimation, adaptive signal processing, Bayesian information criterion, multi-class inverse Wishart mixture filter**

I. Introduction

The estimation of the interference covariance matrix is a fundamental issue in adaptive signal processing and naturally arises in several contexts including target detection, direction of arrival evaluation, secondary data selection, and spectral analysis. To predict the interference covariance, conventional adaptive strategies (such as the sample matrix inversion (SMI) filter [1] and the Kelly's receiver [2]) rely on the sample covariance matrix of a secondary data set collected from range gates spatially close to the one under test. These algorithms ensure satisfactory performance when secondary vectors exhibit the same spectral properties of the interference in the cell under test, are statistically independent of each other, and the size of the training set is larger than twice the useful signal dimension. The above requirements represent important limitations, since the number of data where the disturbance is homogeneous is usually quite limited and, more importantly, a poor training data selection can imply severe performance degradation [3].

A possible strategy to circumvent the lack of a sufficient number of homogeneous secondary data is to exploit a-priori information on the scene illuminated by the radar and reduce the unknown parameters at the estimation stage with appropriate structural models on the covariance matrix [4]–[7].

Adaptive signal processing algorithms based on the mentioned covariance estimators may suffer performance degradation in the presence of model mismatches, e.g., due to changes in the operative conditions arising from meteorological changes or appearance/disappearance of interferences. A first attempt to overcome this drawback has been pursued in [8], where an adaptive classification of the interference covariance matrix structure is addressed resorting to the theory of model order selection (MOS) [9]. By doing so, the actual interference covariance matrix model can be adaptively predicted and mismatch loss avoided.

In this paper, we still focus on adaptive environment classification and cognizance. Unlike [8], we jointly exploit multiple observations (scans) of the scene and a hybrid covariance matrix-class type of tracking is performed so as to define a dynamic environment learning strategy. To this end, as in [10], [11], we resort to Bayesian methods, except that a sequential approach on multiple observations is taken here. This work is inspired by [12] for the approximation of posterior distributions with a mixture, and by [13], [14] for the Inverse Wishart (IW) modeling of the ellipsoid of the contacts' spread around a target's position in the extended target tracking (ETT) problem. We propose to track a hybrid state at each time k, composed by a discrete random variable C_k representing the class (or model) and a positive definite matrix $\boldsymbol{R}_k$ that is related via one-to-one mapping to the actual covariance matrix given the class. Hybrid states are used also when the primary objective is to track multiple targets, but other parameters, such as the process noise, need to be estimated. A notable example is the interactive multiple model (IMM) filter [15], where usually the process noise switches among discrete predefined values. More recent developments are documented in [16], [17], where the approach is applied also to other parameters, such as the target detection probability. The posterior distribution of $\boldsymbol{R}_k$, conditioned on C_k, is approximated by a mixture of IW distributions, while a posterior distribution on C_k is also provided representing the probability that the class C_k has generated the data. The proposed approach is named multi-class inverse Wishart mixture (MC-IWM) filter, and its derivation is only sketched out in this paper, leaving a full description to an extended paper, which is in preparation.

II. Problem Formulation

Consider for time $k \geq 1$ to sequentially observe the data $\boldsymbol{Z}_k = [\boldsymbol{z}_{1,k}, \cdots, \boldsymbol{z}_{N,k}]$. Let the columns of $\boldsymbol{Z}_k$ be N i.i.d. zero-mean Gaussian random vectors of size m, with positive definite

P. Braca and L. M. Millefiori are supported by NATO ACT under project SAC000808 – "Data Knowledge Operational Effectiveness."

ISBN 978-90-827970-1-5 © EURASIP 2018

covariance matrix $\boldsymbol{M}_k$. The case of complex data distributed according to a (zero-mean circularly symmetric) multivariate Gaussian with positive definite covariance is similar and will be reported in the extended version of this paper.

Assume that in each time instant k such covariance $\boldsymbol{M}_k$ belongs to a predefined class, indicated with $C_k \in \mathcal{C}$, being $\mathcal{C} \equiv \{1, 2, \ldots, N_C\}$, and depends on a positive definite matrix $\boldsymbol{R}_k$; the combination of $\boldsymbol{R}_k$ and C_k uniquely defines the actual covariance matrix $\boldsymbol{M}_k$. Note that the dimensionalities of $\boldsymbol{R}_k$ and $\boldsymbol{M}_k$ could not be the same. For instance, the class of white noise is given by $\boldsymbol{M}_k = R_k \boldsymbol{I}_{m\times m}$ where $R_k > 0$ is a one-dimensional variable and $\boldsymbol{I}_{m\times m}$ is the identity matrix of size m.

The goal is to sequentially estimate both the class C_k and the matrix $\boldsymbol{R}_k$ based on the observed data up to k, denoted by $\boldsymbol{Z}_{1:k} = \{\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_k\}$. The current hybrid state is defined as $\boldsymbol{X}_k = \{\boldsymbol{R}_k, C_k\}$. The estimation is based on the posterior distribution of $\boldsymbol{X}_k$ given the data observed up to time k

$$\begin{aligned}\mathcal{P}_{k|k}(\boldsymbol{X}_k) &:= \mathcal{P}(\boldsymbol{X}_k \,|\boldsymbol{Z}_{1:k}) \\ &= \mathcal{P}(\boldsymbol{R}_k \,|C_k, \boldsymbol{Z}_{1:k})\, \mathcal{P}(\boldsymbol{C}_k \,|\boldsymbol{Z}_{1:k}) \\ &:= \mathcal{P}_{k|k}(\boldsymbol{R}_k \,|C_k)\, \mathcal{P}_{k|k}(C_k), \end{aligned} \tag{1}$$

where we have used the notation $\mathcal{P}_{k|j}(\boldsymbol{A}_k)$ to indicate the posterior distribution of $\boldsymbol{A}_k$ at time k given the data $\boldsymbol{Z}_{1:j}$ observed up to time j.

The time evolution of $\boldsymbol{X}_k$ follows a Markov process. Precisely, the class C_k evolves according to a Markov chain

$$\mathcal{P}(C_{k+1} = i \,|C_k = j) = \pi_{ij}, \quad i, j \in \mathcal{C}. \tag{2}$$

Conditioned on the classes C_{k+1}, C_k and on $\boldsymbol{R}_k$, the matrix $\boldsymbol{R}_{k+1}$ evolves according to a first order Markov state transition distribution $\mathcal{P}(\boldsymbol{R}_{k+1} \,|C_k, C_{k+1}, \boldsymbol{R}_k)$. Note that, if the two classes are the same $C_{k+1} = C_k = i$, then it would be reasonable to have a transition with constant mean, i.e., $\mathbb{E}[\boldsymbol{R}_{k+1} \,|C_k = i, C_{k+1} = i, \boldsymbol{R}_k] = \boldsymbol{R}_k$. Then, inspired by [13], [14] and considering that we deal with positive definite matrices, we assume that the spread around the mean is ruled by a Wishart distribution, yielding

$$\mathcal{P}(\boldsymbol{R}_{k+1} \,|C_k = i, C_{k+1} = i, \boldsymbol{R}_k) = \mathcal{W}\left(\boldsymbol{R}_{k+1}; \frac{\boldsymbol{R}_k}{\nu}, \nu\right), \tag{3}$$

where we have indicated with $\mathcal{W}(\boldsymbol{R}; \boldsymbol{A}, \nu)$ a Wishart distribution with matrix parameter $\boldsymbol{A}$ and ν degrees of freedom. Finally, if $C_k \neq C_{k+1}$, (3) is still a Wishart distribution, but with different parameters; the details are omitted for brevity.

III. Multi-class inverse Wishart mixture filter

With the assumptions made in the previous section, the covariance matrix under each class can be estimated sequentially by approximating its posterior distributions with mixtures of IW components. This is similar to the approach taken in [12], where posterior distributions are approximated by mixtures of Gaussian components. The choice of IW components originates from the Gaussian nature of the data $\boldsymbol{Z}_k$ and from the IW distribution being the conjugate prior of $\boldsymbol{M}_k$, given the observed data $\boldsymbol{Z}_k$, e.g., see [13], [18]. This property holds also for $\boldsymbol{R}_k$ in many class definitions of interest. Alternatively, the moment matching approximation can be exploited to maintain the IW structure. The proposed approach allows to sensibly reduce the complexity with respect to a brute force particle filtering strategy. Indeed, as m increases, but even for moderately small values, the representation of objects in $\mathbb{R}^{m\times m}$ would require a prohibitively large number of particles.

Let us assume that the predicted posterior for each class at time k given $\boldsymbol{Z}_{1:k-1}$ is a mixture of inverse Wishart

$$\mathcal{P}_{k|k-1}(\boldsymbol{R}_k \,|C_k) = \sum_{n=1}^{N_W} w_{k|k-1}^{(n,C_k)}\, \mathcal{IW}\left(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)},\, \hat{\nu}_{k|k-1}^{(n,C_k)}\right), \tag{4}$$

where $\widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)}$ is the parameter of the inverse Wishart and $\hat{\nu}_{k|k-1}^{(n,C_k)}$ are the related degrees of freedom. The weight $w_{k|k-1}^{(n,C_k)}$ represents the probability of the n-th component of the mixture at time k for the class C_k after observing the data up to time $k-1$:

$$w_{k|k-1}^{(n,C_k)} = \mathcal{P}(N_k = n \,|C_k,\, \boldsymbol{Z}_{1:k-1}), \tag{5}$$

where N_k defines the switching variable among the different modes. The quantities $\widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)}$, $\hat{\nu}_{k|k-1}^{(n,C_k)}$ and $w_{k|k-1}^{(n,C_k)}$ summarize the information acquired by the data up to time $k-1$, related to both C_k and $\boldsymbol{R}_k$. When the new set of observations $\boldsymbol{Z}_k$ is gathered, the aforementioned quantities are updated as shown in Algorithm 1. Details on the derivation of Algorithm 1 are not reported for space reasons.

Assuming to have available the class probability $p_{k|k-1}(c) := \mathcal{P}_{k|k-1}(C_k = c)$ at time k given the data up to $k-1$, when $\boldsymbol{Z}_k$ is available the class probability is updated as $p_{k|k}(c) := \mathcal{P}_{k|k}(C_k = c)$ following Algorithm 1.

The prediction step, reported in Algorithm 1, restates the class probabilities $p_{k+1|k}(C_{k+1})$ and the mode probabilities $w_{k+1|k}^{(n,C_{k+1})}$ following the Markov chain (2). The IW parameters $\left(\widehat{\boldsymbol{R}}_{k+1|k}^{(n,C_{k+1})},\, \hat{\nu}_{k+1|k}^{(n,C_{k+1})}\right)$ are updated by marginalizing with respect to C_k the following distribution

$$\begin{aligned}&\mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{Z}_{1:k}) \\ &\quad= \int \mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{R}_k)\, \mathcal{P}_{k|k}(\boldsymbol{R}_k|C_k) \mathrm{d}\boldsymbol{R}_k \\ &\quad= \sum_{n=1}^{N_W} \mathcal{P}(n|C_{k+1}, C_k, \boldsymbol{Z}_{1:k}) \int \mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{R}_k) \\ &\qquad \times \mathcal{IW}(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}_{k|k}^{(n,C_k)};\, \hat{\nu}(n, C_k))\, \mathrm{d}\boldsymbol{R}_k. \end{aligned} \tag{6}$$

Forcing the solution of the integral in the last equality of (6) to be again an IW and marginalizing (6) with respect to C_k, we obtain the prediction distribution

$$\begin{aligned}&\mathcal{P}_{k+1|k}(\boldsymbol{R}_{k+1}|C_{k+1}) \\ &\quad= \sum_{n=1}^{N_W N_c} w_{k+1|k}^{(n,C_{k+1})} \mathcal{IW}(\boldsymbol{R}_{k+1};\, \widehat{\boldsymbol{R}}_{k+1|k}^{(n,C_{k+1})};\, \hat{\nu}(n, C_{k+1})). \end{aligned} \tag{7}$$

The quantities $\widehat{\boldsymbol{R}}_{k+1|k}^{(n,C_{k+1})}$ and $\hat{\nu}^{(n,C_{k+1})}$ stem from the moment matching approximation (used also in [13]) with a further

ISBN 978-90-827970-1-5 © EURASIP 2018

transformation that accounts for the different dimensionality when $C_k \neq C_{k+1}$. Such update is expressed by the function $f_P(\cdot)$ in Algorithm 1.

In the prediction step the number of components of the mixture increases to $N_W \times N_C$. To avoid an increase of the computational complexity, following [12], at each time update a pruning criterion is adopted. This consists of retaining only the first N_W components, sorted according to their weights.

A. Two-Class Example

A special case is analyzed in this subsection, where only two classes are possible. The first class is the white noise class with $\boldsymbol{M}_k = R_k \boldsymbol{I}_{m\times m}$ and $R_k > 0$ while in the second class $\boldsymbol{M}_k = \boldsymbol{R}_k \succ \boldsymbol{0}$ is a generic $m \times m$ covariance matrix. The likelihood function is given by

$$\mathcal{P}(\boldsymbol{Z}_k|R_k, C_k) = \begin{cases} (2\pi R_k)^{-\frac{Nm}{2}} \, e^{-\mathrm{Tr}\left(\frac{\boldsymbol{Z}_k\boldsymbol{Z}_k^T}{2R_k}\right)} & C_k = 1, \\ |2\pi \boldsymbol{R}_k|^{-\frac{N}{2}} \, e^{-\mathrm{Tr}\left(\frac{\boldsymbol{Z}_k\boldsymbol{Z}_k^T\boldsymbol{R}_k^{-1}}{2}\right)} & C_k = 2, \end{cases} \tag{8}$$

where $|\cdot|$ and $(\cdot)^T$ denote the determinant and transpose operators, respectively. The variables $\alpha_k^{(n,C_k)}$, reported in Algorithm 1, are given as follows

$$\alpha_k^{(n,C_k)} = \int \mathcal{P}\left(\boldsymbol{Z}_k|C_k, \boldsymbol{R}\right) \mathcal{IW}\left(\boldsymbol{R};\, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)},\, \hat{\nu}_{k|k-1}^{(n,C_k)}\right) \mathrm{d}\boldsymbol{R}.$$

Specifically, for $C_k = 1, 2$ we have, respectively

$$\alpha_k^{(n,1)} = \pi^{\frac{-Nm}{2}} \left|\widehat{R}_{k|k-1}^{(n,1)}\right|^{\frac{\hat{\nu}_{k|k-1}^{(n,1)}}{2}} \frac{\kappa_1(Nm + \hat{\nu}_{k|k-1}^{(n,1)})}{\kappa_1(\hat{\nu}_{k|k-1}^{(n,1)})} \times \left|\mathrm{Tr}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^T\right) + \widehat{R}_{k|k-1}^{(n,1)}\right|^{-\frac{Nm+\hat{\nu}_{k|k-1}^{(n,1)}}{2}}, \tag{9}$$

$$\alpha_k^{(n,2)} = \pi^{\frac{-Nm}{2}} \left|\widehat{\boldsymbol{R}}_{k|k-1}^{(n,2)}\right|^{\frac{\hat{\nu}_{k|k-1}^{(n,2)}}{2}} \frac{\kappa_m(N + \hat{\nu}_{k|k-1}^{(n,2)})}{\kappa_m(\hat{\nu}_{k|k-1}^{(n,2)})} \times \left|\boldsymbol{Z}_k\boldsymbol{Z}_k^T + \widehat{\boldsymbol{R}}_{k|k-1}^{(n,2)}\right|^{-\frac{N+\hat{\nu}_{k|k-1}^{(n,2)}}{2}}, \tag{10}$$

where $\kappa_p(\nu)$ is defined as

$$\kappa_p(\nu) = \prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(\nu + 1 - i)\right]. \tag{11}$$

Furthermore, for any mode n, it follows that the function $f_U(\cdot)$ defined in Algorithm 1 is given by

$$\widehat{R}_{k|k}^{(n,C_k)} = \begin{cases} \mathrm{Tr}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^T\right) + \widehat{R}_{k|k-1}^{(n,C_k)} & C_k = 1, \\ \boldsymbol{Z}_k\boldsymbol{Z}_k^T + \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)} & C_k = 2. \end{cases} \tag{12}$$

$$\hat{\nu}_{k|k}^{(n,C_k)} = \begin{cases} Nm + \hat{\nu}_{k|k-1}^{(n,C_k)} & C_k = 1, \\ N + \hat{\nu}_{k|k-1}^{(n,C_k)} & C_k = 2. \end{cases} \tag{13}$$

The specialization of (7) to the two-class example is not reported for brevity.

Input : $\boldsymbol{Z}_k = [\boldsymbol{z}_{1,k}, \cdots, \boldsymbol{z}_{N,k}]$, N_C, N_W
Output : $\left(\widehat{\boldsymbol{R}}_{i|k}^{(n,c)}, \hat{\nu}^{(n,c)}\right)$, $w_{i|k}^{(n,c)}$, $p_{i|k}(c)$, for $i = k, k+1$

Initialization
$k \longleftarrow 1, \quad p_{k|k-1}(c) \longleftarrow N_C^{-1}$
for $c \in \mathcal{C}$ **do**
  for $n \in \{1, \dots, N_W\}$ **do**
    $w_{k|k-1}^{(n,c)} \longleftarrow N_W^{-1}$
    $\left(\widehat{\boldsymbol{R}}_{k|k-1}^{(n,c)}, \hat{\nu}_{k|k-1}^{(n,c)}\right) \longleftarrow \left(\boldsymbol{R}_0^{(n,c)}, \nu_0^{(n,c)}\right)$
  end
end
for $k \in \{1, \dots, K\}$ **do**
  Update
  Observe new data $\boldsymbol{Z}_k$
  for $C_k = c \in \{1, \dots, N_C\}$ **do**
    for $n \in \{1, \dots, N_W\}$ **do**
      $\alpha_k^{(n,c)} \longleftarrow \int \mathcal{P}\left(\boldsymbol{Z}_k|C_k = c, \boldsymbol{R}_k = \boldsymbol{R}\right) \times \mathcal{IW}\left(\boldsymbol{R};\, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,c)},\, \hat{\nu}_{k|k-1}^{(n,c)}\right) \mathrm{d}\boldsymbol{R}$
      $w_{k|k}^{(n,c)} \longleftarrow \frac{\alpha_k^{(n,c)} w_{k|k-1}^{(n,c)}}{\sum_{n=1}^{N_W} \alpha_k^{(n,c)} w_{k|k-1}^{(n,c)}}$
      $\left(\widehat{\boldsymbol{R}}_{k|k}^{(n,c)}, \hat{\nu}_{k|k}^{(n,c)}\right) \longleftarrow f_U\left(\boldsymbol{Z}_k, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,c)}, \hat{\nu}_{k|k-1}^{(n,c)}, C_k = c\right)$
    end
    $p_{k|k}(c) \longleftarrow \frac{\left(\sum_{n'=1}^{N_W} w_{k|k-1}^{(n',c)} \alpha_k^{(n',c)}\right) p_{k|k-1}(c)}{\sum_{c'=1}^{N_C} \sum_{n'=1}^{N_W} w_{k|k-1}^{(n',c')} \alpha_k^{(n',c')} p_{k|k-1}(c')}$
  end
  Prediction
  for $C_{k+1} = c \in \{1, \dots, N_C\}$ **do**
    $n' \longleftarrow 1$
    $p_{k+1|k}(c) \longleftarrow \sum_{c'} \pi_{cc'}\, p_{k|k}(c')$
    for $C_k = \bar{c} \in \{1, \dots, N_C\}$ **do**
      for $n \in \{1, \dots, N_W\}$ **do**
        $w_{k+1|k}^{(n',c)} \longleftarrow w_{k|k}^{(n,\bar{c})} \frac{\pi_{c\bar{c}} p_{k|k}(\bar{c})}{p_{k+1|k}(c)}$
        $\left(\widehat{\boldsymbol{R}}_{k+1|k}^{(n',c)}, \hat{\nu}_{k|k}^{(n',c)}\right) \longleftarrow f_P\left(\widehat{\boldsymbol{R}}_{k|k}^{(n,\bar{c})}, \hat{\nu}_{k|k}^{(n,\bar{c})}, C_{k+1} = c, C_k = \bar{c}\right)$
        $n' \longleftarrow n' + 1$
      end
    end
  end
  Pruning
  for $C_{k+1} = c \in \{1, \dots, N_C\}$ **do**
    Sort the $N_W \times N_C$ components according to the weights $\{w_{k+1|k}^{(n',c)}\}_{n'=1}^{N_W\, N_C}$ retaining the first N_W elements
    for $n \in \{1, \dots, N_W\}$ **do**
      $w_{k+1|k}^{(n,c)} \longleftarrow \frac{w_{k+1|k}^{(n,c)}}{\sum_{n'=1}^{N_W} w_{k+1|k}^{(n',c)}}$
    end
  end
end

Algorithm 1: Multi-class inverse Wishart mixture filter.

IV. Numerical Results

In this section we present the results of computer experiments in the two-class case presented in Sec. III-A. The covariance matrices used to generate the data $\boldsymbol{Z}_k$ are defined as $\boldsymbol{M}_k = \sigma^2 \boldsymbol{I}_{m\times m}$ if $C_k = 1$, and as $\boldsymbol{M}_k = \boldsymbol{R}(\sigma,\rho)$ if $C_k = 2$. In the second class the covariance matrix component are defined as $\{\boldsymbol{R}(\sigma,\rho)\}_{ii} = \sigma^2$ on the diagonal and $\{\boldsymbol{R}(\sigma,\rho)\}_{ij} = \{\boldsymbol{R}(\sigma,\rho)\}_{ji} = \sigma^2 \rho^{|i-j|}$ off the diagonal, for $i,j = 1,\ldots,m$. It is worthwhile to remark that the classification is nested because the first class is a special case of the second one when $\rho = 0$; nevertheless, the probability that $\boldsymbol{M}_k$ is a white covariance matrix under the probability measure in the second class is zero. Moreover, the structure of the second class is generic, i.e., the cross-correlation structure $\rho^{|i-j|}$ is not taken into account in the proposed filter.

Among the several available Bayesian estimators, we select the posterior mean that is defined as:

$$\begin{aligned}
\widehat{\boldsymbol{M}}_{k|k} &= \mathbb{E}\left[\boldsymbol{M}_k|\boldsymbol{Z}_{1:k}\right] \\
&= \sum_{C_k} \int \boldsymbol{M}_k(\boldsymbol{R}_k, C_k)\mathcal{P}_{k|k}(C_k)\mathcal{P}_{k|k}(\boldsymbol{R}_k|C_k)\,\mathrm{d}\boldsymbol{R}_k \\
&= \sum_{c=1}^{N_C}\sum_{n=1}^{N_W} p_{k|k}(c)\, w_{k|k}^{(n,c)} \widehat{\boldsymbol{M}}_{k|k}(n,c),
\end{aligned} \tag{14}$$

where $\widehat{\boldsymbol{M}}_{k|k}(n,c)$ is the expected mean conditioned to the n-th mode of the IW mixture and the class c. Recalling that the mean of an IW distribution $\mathcal{IW}\left(\boldsymbol{R};\, \widehat{\boldsymbol{R}},\, \hat{\nu}\right)$ of dimensionality p is $\widehat{\boldsymbol{R}}/(\hat{\nu}-p-1)$, we have

$$\widehat{\boldsymbol{M}}_{k|k}(n,c) = \begin{cases} \left(\hat{\nu}_{k|k}^{(n,1)} - 2\right)^{-1} \widehat{R}_{k|k}^{(n,1)} \boldsymbol{I}_{m\times m}, & c = 1, \\ \left(\hat{\nu}_{k|k}^{(n,2)} - m - 1\right)^{-1} \widehat{\boldsymbol{R}}_{k|k}^{(n,2)}, & c = 2. \end{cases} \tag{15}$$

In Figures 1a to 1c we show the Frobenius norm of the mean error of the posterior mean $\widehat{\boldsymbol{M}}_{k|k}$ and the class-clairvoyant ML estimator, which, at each time scan k, has knowledge of the true class and computes the sample covariance matrix of the data $\boldsymbol{Z}_k$. Even if the class-clairvoyant ML is expected to get a head start from the knowledge of the class, the posterior mean is able to outperform it taking advantage of the information contained in the previous steps. Such improvement is not surprising, as it is present in most Bayesian strategies, e.g., in the Kalman filter. However, it corroborates the effectiveness of the IW approximation on the posterior used to derive the MC-IWM filter. In Figures 1a to 1c a time evolution of 40 steps is reported in which class 2 is in force for the first and last 10 steps, while class 1 is in force from step 11 to 30. The simulations are evaluated for three values of $\rho = 0.2, 0.5, 0.7$ and $m = 4, 8, 16$ with $\sigma^2 = 1$. The estimation error is more affected by m rather than ρ. Indeed, the larger is m, the larger the estimation error is. This is especially true under class 2, as one would intuitively expect after an increase of the degrees of freedom with the same number of observations.

The classification capability, in terms of the posterior probability $p_{k|k}(c)$, is affected mainly by ρ. For ρ equals to 0.2 the classification capability exhibits poor performance in correctly detecting class 2, as Figures 1d to 1f show. Even if this classification performance is poor, it is still better if compared to the ML. This aspect can be explained considering that the posterior mean estimator under the wrong class can be still a good estimator given that the covariance cross-terms are small if ρ is small. However, increasing ρ up to 0.5 and 0.7 the classification performance improves sensibly, as the covariance matrices under the two classes become more and more dissimilar. Finally, in Figures 1g to 1i the detection capability is compared with that of the BIC rule [9], computed on $\boldsymbol{Z}_k$, versus the data size N. Specifically, we compare the averaged posterior class probability, after 10 time steps, with the BIC under the true classes. The proposed strategy shows a sensible improvement of the detection capability, especially in the regime of small values of N and for all the values of ρ considered in the analysis.

V. Conclusion

The dynamic estimation of the environment (covariance matrix) enables adaptive signal processing strategies to improve performance and robustness in several fields, e.g., surveillance, automotive, robotics, etc. A Bayesian approach, named MC-IWM filter, has been proposed to track both the covariance class and its unknown parameters. The MC-IWM filter is computationally efficient, given that it avoids to use a prohibitively large number of particles that would be otherwise required by a brute-force particle filtering strategy. The reported computer experiments show that the MC-IWM filter outperforms the class-clairvoyant maximum likelihood estimator in terms of covariance estimate accuracy, as well as the Bayesian information criterion rule in terms of classification performance. Future developments of this work will be devoted to the study of the effects of radar and sonar performance when the environment is learned using the proposed MC-IWM filter.

References

[1] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Trans. Aerosp. Electron. Syst.*, no. 6, pp. 853–863, 1974.

[2] E. J. Kelly, "An Adaptive Detection Algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, no. 2, pp. 115–127, 1986.

[3] B. Himed and W. L. Melvin, "Analyzing Space-Time Adaptive Processors Using Measured Data," in *Asilomar Conf. on Signals, Systems and Computers*, 1997, pp. 930–935.

[4] M. Steiner and K. Gerlach, "Fast Converging Adaptive Processor or a Structured Covariance Matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, 2000.

[5] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Maximum Likelihood Estimation of a Structured Covariance Matrix with a Condition Number Constraint," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3004–3021, 2012.

[6] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage Algorithms for MMSE Covariance Estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, 2010.

[7] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 40, no. 5, pp. 1314–1323, 1999.

[8] V. Carotenuto, A. De Maio, D. Orlando, and P. Stoica, "Model Order Selection Rules For Covariance Structure Classification," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5305–5317, 2017.

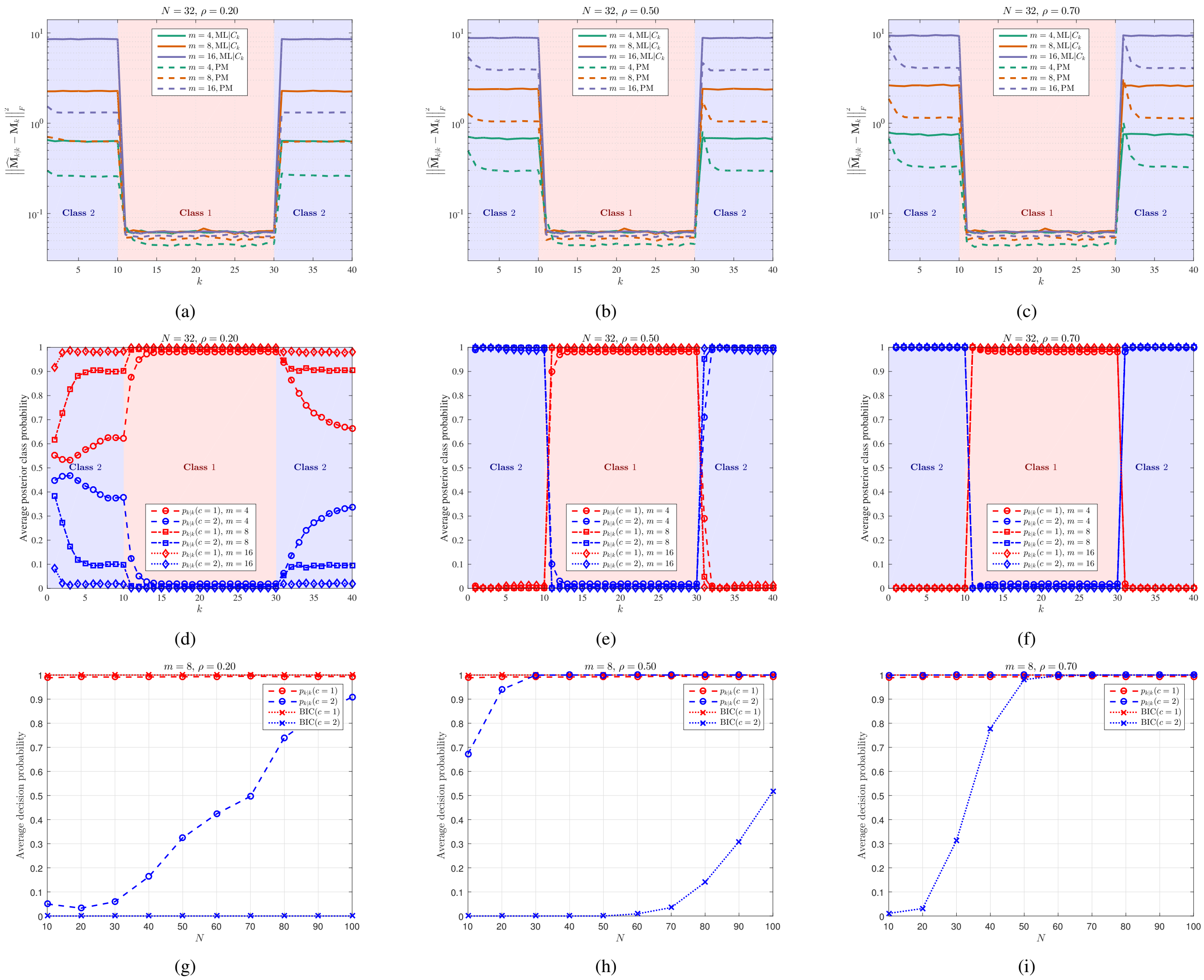

Fig. 1: Panels (a) to (c) show the estimation error over time of the clairvoyant ML (solid lines) and of the posterior mean (PM) of the MC-IWM filter (dashed lines). Panels (d) to (f) show the classification performance of the MC-IWM filter; red color refers to class 1, blue to class 2, and the background color identifies the true class over time. Panels (g) to (i) show a comparison of the classification performance of the proposed approach (dashed lines) with the BIC (dotted lines), versus N. All the figures are averaged over $N_{\mathrm{MC}} = 3000$ Monte Carlo trials and $\sigma^2 = 1$.

[9] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, 2004.

[10] O. Besson, J. Y. Tourneret, and S. Bidon, "Knowledge-aided Bayesian detection in heterogeneous environments," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 355–358, 2007.

[11] P. Wang, H. Li, and B. Himed, "A Bayesian parametric test for multichannel adaptive signal detection in nonhomogeneous environments," *IEEE Signal Process. Lett.*, vol. 17, no. 4, pp. 351–354, 2010.

[12] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2602–2612, 2003.

[13] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 3, pp. 1042–1059, 2008.

[14] K. Granström, A. Natale, P. Braca, G. Ludeno, and F. Serafino, "Gamma Gaussian inverse Wishart probability hypothesis density for extended target tracking using X-band marine radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6617–6631, 2015.

[15] Y. Bar-Shalom, P. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT: YBS Publishing, 2011.

[16] G. Soldi and P. Braca, "Online estimation of unknown parameters in multisensor-multitarget tracking: A belief propagation approach," in *Intern. Conf. on Inform. Fusion (FUSION)*, 2018.

[17] G. Papa, P. Braca, S. Horn, S. Marano, V. Matta, and P. Willett, "Multisensor adaptive bayesian tracking under time-varying target detection probability," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2193–2209, 2016.

[18] A. D. Maio, A. Farina, and G. Foglia, "Knowledge-aided Bayesian radar detectors & their application to live data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, 2010.